April 2021

Substrate selection in endoplasmic reticulum protein quality control

Ben Adams
University of Massachusetts Amherst

Substrate selection in endoplasmic reticulum protein quality control

Ben Adams

Substrate selection in endoplasmic reticulum protein quality control

A Dissertation Presented

by

BENJAMIN M. ADAMS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2021

Molecular and Cellular Biology

Substrate selection in endoplasmic reticulum protein quality control

A Dissertation Presented

By

BENJAMIN M. ADAMS

Approved as to style and content by:

_____
Daniel N. Hebert, Chair

_____
Lila M. Gierasch, Member

_____
Scott C. Garman, Member

_____
Thomas J. Maresca, Member

_____
Thomas J Maresca
Graduate Program Director
Molecular and Cellular Biology

DEDICATION

To my wife, Erika, and to my parents and siblings.

ACKNOWLEDGMENTS

There are many people I owe thanks, so I will proceed roughly chronologically. First, I'd like to thank my advisor Dr. Dan Hebert. As with all his students, Dan has sought to help me succeed both in graduate school and in my future career. He has invested a great deal of time mentoring me and through his teaching I have progressed in many aspects of academic science, from experimental design to writing and presenting finished work. The experimental questions and systems we have developed are very exciting and I am grateful for his encouragement to take ownership of this work as that has been a key part of my scientific development. I will always be thankful for my time in his lab.

My committee members, Dr. Lila Gierasch, Dr. Scott Garman, and Dr. Thomas Maresca, have offered useful feedback and direction which has helped shape the work I've presented here. Having such diverse and informative perspectives on my work has been extremely valuable, and I'd like to thank you each for the time you have spent thinking about and commenting on my work.

The past and present members of the Hebert lab have helped make it a great place to work and develop, offering suggestions for improvement ranging from experimental procedures to presentations. Thanks to Abla, Johan, Lydia, Jill, Nathan, Kevin, and Haiping, as well as the undergraduates who worked with me in the lab, Seleem and Michela. I'd like to specifically thank Nathan and Kevin for their help with my most recent work regarding the identification of UGGT1 and UGGT2 substrates. Having lab mates who are excited to discuss data and explore possibilities was a great help.

The MCB program is filled with many great people both professionally and personally, and it was the ideal environment in which to spend grad school. I'd like to thank the LSL third floor labs for being a great place to work every day, as well as specifically thank Rilee Zeinert, Kathy Sanidad, Alex Wells, Cameron Butova, Tom Sawyer and my lab mates I've already mentioned for their friendships.

I'd also like to thank my undergraduate mentor Dr. Ben Brammell, for first instilling in me a love of research. Dr. Brammell was a vital part of my early growth and his guidance and discussions were invaluable. Having someone take a personal interest in my early scientific development was extremely important.

My parents, Mike and Kim, and siblings Chris, Nate, and Sami have encouraged me for many years and gave me the lifelong guidance and tools necessary, and couldn't be thanked enough here. Thank you for supporting my interest in discovery from a young age, and for putting up with living far away to pursue it.

Finally, thanks to my wife Erika. When we got married six years ago I don't think we imagined spending so much time apart so we could each develop our careers, but together we've found what works best for us. You've never asked for more time than I could give and have always been fully supportive of my time here and excited to know about my work. Your presence, near or far, has helped smooth the bumps of grad school and I'll always be thankful for your help.

ABSTRACT

CHAPERONE-MEDIATED PROTEIN FOLDING IN THE ENDOPLASMIC RETICULUM

FEBRUARY 2021

BENJAMIN M. ADAMS, B.A., ASBURY UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Daniel N. Hebert

Protein folding and maturation is a complex and error-prone process. Errors in this process may lead to deleterious effects ranging from non-functional single proteins to large-scale protein aggregation leading to cell death. It is essential for cellular function that protein misfolding does not occur unchecked, and therefore numerous chaperone systems exist within the cell. For the thousands of proteins which traffic through the secretory pathway, the primary site of folding and maturation is the endoplasmic reticulum (ER). Multiple chaperone pathways within the ER, generally termed ER protein quality control, must support the proper maturation process of these thousands of substrates. While some simple secretory pathway proteins may be able to fold with minimal chaperone engagement, more complex proteins may commonly misfold even under native conditions, which is especially important for multi-cellular organisms which have larger and more complex secretory pathway proteomes.

The chaperone pathways within the ER engage substrates based generally on features those substrates possess. These include hydrophobic regions, free cysteines, and N-glycans. However, which substrates are selected by each of these pathways is not well understood on a systematic level. The work presented here examines the chaperone selection process for a substrate which possess all features and demonstrates that substrate features do not dictate chaperone pathway engagement. As such, an understanding of which substrates are engaged by which pathway under endogenous conditions requires experimental determination. The N-glycan based chaperone pathway was next examined, and the substrates which heavily engage this process under endogenous conditions were described. This information allows for a previously lacking understanding of the folding and maturation process of many proteins and therefore presents possible interventions for the diseases and cellular functions associated with these proteins.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

Protein Quality Control in the Endoplasmic Reticulum

Adams B.M.[1,2]., Oster M.[1], Hebert D.N.[1,2]

published in Protein Journal, 2019

[1]Department of Biochemistry and Molecular Biology, University of Massachusetts, 240 Thatcher Road, Amherst, MA, 01003, USA
[2]Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, MA, 01003, USA

Abstract

The site of protein folding and maturation for the majority of proteins that are secreted, localized to the plasma membrane or targeted to endomembrane compartments is the endoplasmic reticulum (ER). It is essential that proteins targeted to the ER are properly folded in order to carry out their function, as well as maintain protein homeostasis, as accumulation of misfolded proteins could lead to the formation of cytotoxic aggregates. Because protein folding is an error-prone process, the ER contains protein quality control networks that act to optimize proper folding and trafficking of client proteins. If a protein is unable to reach its native state, it is targeted for ER retention and subsequent degradation. The protein quality control networks of the ER that oversee this evaluation or interrogation process that decides the fate of maturing nascent chains is comprised of three general types of families: the classical chaperones, the carbohydrate-dependent system, and the thiol-dependent system. The cooperative action of these families promotes protein quality control and protein homeostasis in the ER. This review will describe the families of the ER protein quality control network and discuss the functions of individual members.

Introduction

Gunter Blobel's seminal studies on protein targeting that led to his proposal of the *signal hypothesis theory* in the 1970s (Blobel & Sabatini, 1971; Blobel & Dobberstein, 1975b, 1975a), laid the conceptual framework for later studies on the eukaryotic secretory pathway and provided a valuable experimental system to dissect processes involving the endoplasmic reticulum (ER) (Bulleid & Freedman, 1988; Nicchitta & Blobel, 1993; Hebert et al., 1995). A third of the proteome was later found to be targeted to the ER for entry into the secretory pathway for maturation and eventually secretion or delivery to various locations within the secretory/endocytic pathways (Huh et al., 2003). With the discovery of molecular chaperones a decade later by Ellis, Hartl, Horwich, Laskey, Lorimer, Pelham and others (Laskey et al., 1978; Pelham, 1986; Cheng et al., 1989; Goloubinoff et al., 1989; Ellis, 1996), protein folding and assembly was found to be an assisted process within the cell. Several molecular chaperone families reside in the ER to help these early maturation events including the folding reaction for proteins that traverse the secretory pathway. As protein folding is an error-prone process, prolonged binding to molecular chaperones is also utilized in an interrogation or evaluation process to determine if the structural integrity of the protein product is sound so that native proteins can be passed along the secretory pathway and non-native products can be retained and repaired, or eventually sorted for degradation. This cellular interrogation step was termed *protein quality control* by Ari Helenius shortly after the discovery of molecular chaperones (Hurtley & Helenius, 1989). In this article, we will review the quality control processes that take place in the early eukaryotic secretory pathway or ER

that seek to ensure that only native substrates are passed along the secretory pathway while terminally misfolded proteins are targeted for degradation.

The ER is the site of protein maturation for secretory pathway cargo. These processes are assisted by chaperone and oxidoreductase family members that help increase the efficiency of reaching the native and active state of a protein. Many of these maturation assistance factors are also central players in the protein quality control evaluation and sorting processes. There are three general categories of protein quality control factors in the ER that will be discussed below: (1) the classical molecular chaperone systems; (2) glycan-dependent molecular chaperone systems; and (3) thiol-dependent oxidoreductases. These factors play diverse but well integrated roles in maintaining protein homeostasis in the ER that involves the passage of properly folded cargo, while defective cargo is retained and subsequently degraded.

The Classical Chaperones of the ER

Chaperones from the classical heat shock (heat shock proteins, Hsp) families generally bind to hydrophobic domains on substrates in order to promote productive folding and prevent aggregation (Horwich et al., 1990). The binding cycle of the classical chaperones is regulated by adenine nucleotides. The ER contains members from the Hsp40s, Hsp70 (BiP/GRP78) and Hsp90 (GRP94) families. Of these, BiP is the most abundant and appears to play the widest role in the ER, including assisting protein folding, protein translocation, ER retention and promotion of ER-associated degradation (ERAD) of misfolded substrates, and inducing the unfolded protein response (UPR)

signaling cascade (Hendershot, 2004; Behnke et al., 2015). Here, we will focus largely on the protein quality control functions of BiP and its associated regulators.

BiP binds promiscuously to clients frequently at a number of sites on a maturing protein (Behnke et al., 2016; Blond-Elguindi et al., 1993). These sites are generally hydrophobic and contain alternating aliphatic residues (Flynn et al., 1991; Blond-Elguindi et al., 1993). There are algorithms that predict BiP sites on client proteins by analyzing the primary amino acid sequence (Schneider et al., 2016). However, the utilization of these predicted sites by BiP requires experimental demonstration for validation. BiP interacts with substrates via its substrate binding domain (SBD) that is regulated by its nucleotide binding domain (NBD; Figure 1.1A and B) (Behnke et al., 2015). When the NBD is bound by ATP, BiP exhibits a low substrate affinity due to the lid of the SBD being in an open conformation, leading to a high on-and-off rate of the substrate. Hydrolysis of ATP, leaving BiP in an ADP-bound state, increases substrate affinity by allowing the lid of the SBD to close over the substrate. In the case of some substrates, direct interaction between the lid and the substrate can occur without the lid domain closing over the substrate, therefore allowing for significant substrate tertiary structure and suggesting that BiP has a relatively flexible substrate preference (Schlecht et al., 2011).

BiP co-factors regulate the binding and release of substrates. Proteins that contain J-domains or HPD motifs (the tripeptide His-Pro-Asp) are known to interact with Hsp70 family members to induce their ATPase activity at locations throughout the cell (Kampinga et al., 2018). The ATPase activity of BiP is stimulated by the ERdj proteins that also assist the recruitment of substrates to BiP (Otero et al., 2010). Release of

substrate occurs when ADP is exchanged for ATP, placing BiP back in a low substrate affinity conformation. Nucleotide exchange is regulated by nucleotide exchange factors (NEFs). There are two known NEFs for BiP: Sil1 and Lhs1 in *Saccharomyces cerevisiae* (Boisramé et al., 1998; Kabani et al., 2002; Tyson & Stirling, 2000), and Bap and GRP170 are mammalian homologues (Tyson & Stirling, 2000; Chung et al., 2002; Steel et al., 2004; Behnke et al., 2015). GRP170, in addition to being a NEF for BiP, also possesses a region that is homologous to Hsp70 itself (X. Chen et al., 1996). However, the role for this region is uncertain.

In yeast, BiP (Kar2p), is localized to the ER translocon by its association with the J-domain protein Sec63 (Matlack et al., 1999). This positions BiP in its substrate bound state to help with the translocation and early stages of nascent chains folding to its native state (Nicchitta & Blobel, 1993; Helenius & Hammond, 1994; Hebert et al., 1998). BiP also plays a central role in targeting misfolded proteins for degradation (Plemper et al., 1997; Skowronek et al., 1998; Brodsky et al., 1999). These alternate fates for a BiP substrate may be directed by multiple factors. Prolonged substrate binding, as opposed to transient interactions with early folding intermediates, may target a substrate for ERAD (Farinha & Amaral, 2005; Sörgjerd et al., 2006). Further specificity may be conferred by the BiP cofactor J-proteins. While the role of each J-protein is currently unclear, certain J-proteins have been shown to promote distinct fates for BiP substrates. ERdj4 promotes the degradation of a natural variant of surfactant protein C (SP-C) and represses UPR by promoting BiP interaction with IRE1 and repressing IRE1 dimerization (Dong et al., 2008; Amin-Wetzel et al., 2017). ERdj5 possesses reductase activity and has been shown to promote the degradation of multiple substrates

including SP-C and the α1-antitrypsin variant *null* Hong Kong (NHK), possibly by reduction of disulfide bonds, allowing the substrate to be threaded through the ERAD retrotranslocon (Cunnea et al., 2003; Dong et al., 2008; Ushioda et al., 2008). A cellular peptide library demonstrated that ERdj4 and ERdj5 recognize peptides rich in aromatic residues, suggesting that these J-proteins recognize misfolded residues exposing aggregation prone domains and therefore support degradation of these proteins (Behnke et al., 2016). ERdj3 has a diverse set of roles, but generally promotes protein folding (Behnke et al., 2015; Khodayari et al., 2017). However, ERdj3 can also promote degradation of specific substrates such as β-glucocerebrosidase (Tan et al., 2014). ERdj6 ($p58^{IPK}$) appears to promote protein folding and the protection against ER stress (Rutkowski et al., 2007). The J-domain co-factors of BiP control the localization of BiP, its substrate selection, and activity, thereby contributing to the diversification of its roles in ER protein quality control.

The activity of BiP can also be regulated by chemical modification. In the absence of ER stress, BiP is chemically modified by AMPylation (Ham et al., 2014; Anwesha et al., 2015; Preissler et al., 2015). AMPylation of BiP by FICD (or HYPE) has been shown to be inhibitory to BiP substrate binding as it confers a substrate affinity similar to the ATP-bound state (Preissler, Rohland, et al., 2017). AMPylation does not inhibit ERdj protein binding but allosterically inhibits ERdj mediated ATP-hydrolysis, causing BiP to remain in a low substrate affinity state. Under stress, the level of BiP increases and it is de-AMPylated by FICD, which can act to both AMPylate and de-AMPylate BiP, depending on its functional state (Figure 1.1B) (Preissler et al., 2015; Preissler, Rato, et

al., 2017). In this manner, a large pool of BiP can be quickly converted into an active state upon ER stress, thereby decreasing the time required to respond to stress.

While the ER hsp90 family member GRP94, is absent from the yeast ER, it is found in higher eukaryotes. ERAD substrates including α-1 antitrypsin NHK and γ-aminobutyric acid type A (GABAA) receptors were stabilized by the knockdown of GRP94 in cells (Christianson et al., 2009; Zhong et al., 2015; Di et al., 2016). These studies also found GRP94 associated with the lectin ERAD receptor, Os-9, and the ERAD E3 ligase HRD1. However, other data demonstrates that Os-9 preferentially interacts with hyperglycosylated GRP94, and knockdown of GRP94 does not stabilize α-1 antitrypsin NHK (Dersh et al., 2014). Together these results support a possible role for GRP94 in ERAD substrate selection and targeting, but these conclusions remain unclear.

Carbohydrate-dependent protein quality control

The majority of proteins that are targeted to the ER are N-linked glycosylated (Apweiler et al., 1999). N-linked glycosylation plays numerous roles during protein folding. These large, hydrophilic protein appendages promote protein solubility and alter folding energetics (Figure 1.2), as well as protein function (Haraguchi et al., 1995; Cai et al., 2005; Skropeta, 2009; Culyba et al., 2011; Hebert et al., 2014). N-linked glycans also act as reporters of the folded state and age of the glycoprotein. As a glycoprotein folds and matures, the glycan is modified, through both trimming and addition, and these modifications affect the glycoprotein's interaction with carbohydrate-binding proteins resident to the ER, thereby altering folding and trafficking (Hebert et al., 2005; Caramelo & Parodi, 2015; Lamriben et al., 2016). At the hub of the glycan-dependent quality

control system are the lectin chaperones calnexin and calreticulin. Calnexin and calreticulin substrate binding plays a central role in the folding and quality control of glycosylated cargo in the ER.

The N-linked glycan (Figure 1.2) is preassembled on a dolichol phosphate in the ER membrane and is appended *en bloc* by the oligosaccharyltransferase (OST) as $GlcNAc_2Man_9Gluc_3$ to the Asn in the consensus site of Asn-X-Ser/Thr (where X is not Pro) on the substrate (Breitling & Aebi, 2013; Lamriben et al., 2016). While most glycosylation occurs co-translationally via OST complexes containing the catalytic subunit STT3A, skipped acceptor sites can be glycosylated post-translationally via STT3B-containing OST complexes (Shrimal et al., 2015).

After glycosylation, the glycan is then trimmed by α-glucosidase I, generating a glycan with two glucoses. In this state, the glycoprotein can interact with the lectin malectin. Malectin is a membrane-associated ER-resident protein that specifically recognizes di-glucosylated glycans (Schallus et al., 2008). Malectin associates with the OST component ribophorin I and acts to retain misfolded glycoproteins in the ER (Qin et al., 2012; K. Takeda et al., 2014). The ER retention role of malectin is somewhat surprising as it resides in a region of the ER where early folding steps occur. A potential answer to this conundrum may be that malectin is ER-stress induced, and therefore may only play a relevant role under stress when proteins translocating into the ER are less likely to undergo productive folding (Galli et al., 2011). Alternatively, the association of malectin with an OST subunit may aid in glycosylation of downstream sites as most glycoproteins contain more than one site of modification.

Trimming of the second glucose by α-glucosidase II creates monoglucosylated glycans. In this state, the glycan can be bound by the membrane-bound lectin chaperone calnexin or its soluble paralogue calreticulin that bind to monoglucosylated glycans with micromolar affinity (Figure 1.4) (Hammond et al., 1994; Schrag et al., 2001; Gopalakrishnapai et al., 2006). Trimming of the final glucose by α-glucosidase II yields a non-glucosylated glycan that supports release of the glycoprotein from calnexin and calreticulin. In this state, natively folded glycoproteins can be further trafficked through the ER. Non-natively folded glycoproteins are recognized by the folding sensor UDP-glucose: glycoprotein glucosyltransferase 1 (UGGT1). UGGT1 reglucosylates substrates regenerating monoglucosylated glycans that then become substrates for calnexin/calreticulin rebinding, supporting retention of misfolded substrates while allowing for continued folding (Hebert et al., 1995; Sousa & Parodi, 1995a). Cycles of substrate binding to calnexin and calreticulin, trimming by α-glucosidase II, and reglucosylation by UGGT1 support the retention of misfolded glycoproteins in the ER while promoting folding through interactions with chaperones and their associated co-factors. As such, UGGT1 acts as a gatekeeper of the calnexin/calreticulin cycle, determining if glycoproteins may exit the ER or must be retained.

While calnexin and calreticulin share 39% sequence identity, each has unique properties (Hebert et al., 1997; Vassilakos et al., 1998). Both calnexin and calreticulin are composed of a lectin domain, a flexible proline-rich domain (P-domain) and a C-terminal domain (Figure 1.3A and B) (Schrag et al., 2001). Calnexin possesses a transmembrane domain near its C-terminus, while the soluble calreticulin's C-terminal domain has low affinity and high capacity calcium binding sites comprised of a series of acidic residues

that support its role as a calcium buffer (Z. Li et al., 2001). The lectin binding domain of both calnexin and calreticulin folds to a globular conformation containing a single carbohydrate binding site (Figure 1.3B). Isothermal calorimetry data has demonstrated that the lectin domain alone is capable of binding substrate (Kozlov, Pocanschi, et al., 2010). The P-domain adopts an extended, arm-like conformation with a hairpin turn. The P-domain of calnexin contains four copies of a Pro rich motif (Ellgaard et al., 2002), while the P-domain of calreticulin contains three such motifs, though both are structurally similar. The P-domain of both calnexin and calreticulin are interaction sites for the cofactors ERp57 and ERp29, protein disulfide isomerases; and cyclophilin B, a peptidyl proline isomerase (Oliver et al., 1997; Kozlov, Bastos-Aristizabal, et al., 2010; Kozlov et al., 2017a). By interacting with a diverse set of folding factors, the P-domains of calnexin and calreticulin function to bring substrates and folding factors in close proximity, supporting productive protein folding and quality control.

UGGT1 is a soluble ER-resident protein consisting of an N-terminal folding sensor domain and a C-terminal glucosyltransferase domain. Recent x-ray crystal structures from both *Thermomyces dupontii* and *Chaetomium thermophilum* have found that the N-terminal domain consists of four catalytically inactive thioredoxin like domains (Roversi et al., 2017; Satoh et al., 2017). Three of these domains are sequential while the fourth is non-sequential, comprised of long range interactions. The N-terminal folding sensor domain adopts a flexible, curved conformation with a prominent central cavity that contains hydrophobic patches. These findings suggest that UGGT1 interacts with high-mannose substrates primarily via hydrophobic interactions. UGGT1 has a homologue UGGT2 that is 55% identical to UGGT1. UGGT2 has been shown to possess

an active glucosyltransferase domain when domain swapped with UGGT1, but biological substrates have not been identified for UGGT2 (Y. Takeda et al., 2014).

Selenoprotein 15 (Sep15/Selenof) is a UGGT binding protein with redox activity (Ferguson et al., 2006). Sep15 interacts tightly with UGGT1 and 2, such that the entire pool of Sep15 is bound to UGGT1/2, though not all UGGT1/2 are bound by Sep15 (Korotkov et al., 2001). *In vitro*, Sep15 has been shown to increase the enzymatic activity of UGGT1/2 (Y. Takeda et al., 2014). Sep15 may modulate the selection of substrates by UGGT or act upon substrates modified by UGGT to help with their repair and folding (Yim et al., 2018). Selenocysteines, considered the 21st amino acid, are found in the active site of Sep15. Since selenocysteines generally possess reducing or disulfide bond breaking activity, it is possible that Sep15 acts to reduce disulfides to try to repair non-native cargo. Additional work is needed to elucidate the biological role of Sep15.

The roles of the calnexin/calreticulin cycle are diverse and important for proper ER function. These roles range from promoting productive folding, limiting deleterious folding pathways and aggregation, and retention of non-native or incompletely assembled proteins (Rajagopalan & Brenner, 1994; Cannon et al., 1996; Vassilakos, Myrna, et al., 1996; Hebert et al., 1996). Indicative of these vital roles is the embryonic lethality of knocking out multiple factors in the calnexin/calreticulin pathway including calreticulin (18 days), UGGT1 (13 days) and ERp57 (13.5 days) (Mesaeli et al., 1999; Molinari et al., 2005; Coe et al., 2010). Calnexin knockout mice display decreased rates of survival within 48 hr post-birth while survivors displayed multiple motor disorders (Denzel et al., 2002). Such lethality and disorders are indicative of the diverse roles of calnexin and

calreticulin, and the extensive set of substrates that are dependent on this carbohydrate-dependent chaperone pathway for proper folding and trafficking.

A central role of the calnexin/calreticulin pathway is to promote proper folding of glycoprotein intermediates. Numerous substrates exhibit reduced folding efficiency and decreased rates of trafficking when interactions with calnexin and calreticulin are abrogated, including hemagglutinin, neuraminidase, vesicular stomatitis virus G protein, MHC class I, antithrombin III and corin (Cannon et al., 1996; Hebert et al., 1996; Vassilakos, Myrna, et al., 1996; N. Wang et al., 2008; Chandrasekhar et al., 2016; H. Wang et al., 2018). Calnexin and calreticulin act as molecular chaperones by decreasing the rate of folding of substrates in a region-specific manner (Hebert et al., 1997; Daniels et al., 2003; Chandrasekhar et al., 2016; N. Wang et al., 2008). This is done by sterically blocking regions proximal to the glycan interacting with the lectin domain while leaving regions distal to the glycan free to fold or interact with other factors. In this manner, the folding pathway of a protein can be determined both by the amino acid sequence and the location of glycans.

When proper folding or multimer assembly of calnexin and calreticulin substrates does not occur, the calnexin/calreticulin pathway generally acts to retain such substrates through persistent reglucosylation by UGGT1 and rebinding to calnexin or calreticulin (Gao et al., 2002; Pearse et al., 2008, 2010; Tannous et al., 2015)**.** However, data from *UGGT1*$^{-/-}$ mouse embryonic fibroblasts showed that the effect of reglucosylation and interaction with the calnexin/calreticulin pathway is substrate specific (Soldà et al., 2007). There appears to exist three classes of UGGT1 substrates. The first class are proteins that require only one round of binding to calnexin or calreticulin, and therefore

do not require reglucosylation by UGGT1 to fold properly. When UGGT1 is knocked out, the trafficking of these proteins is unaltered. The second class involves proteins that are secreted faster when reglucosylation does not occur, and this might be due to the decreased interaction time with calnexin/calreticulin. Alternatively, some proteins displayed decreased rates of secretion, suggesting that reglucosylation and multiple rounds of calnexin/calreticulin binding acts to increase the rate of secretion. This may be due to the calnexin/calreticulin pathway promoting folding efficiency and sequestration of structural elements that act to retain such substrates in the ER. Overall, the calnexin/calreticulin pathway alters the folding and trafficking of a diverse set of substrates, but it does not function in a common manner for all substrates. Eventually non-productively folding substrates must be extracted from the calnexin/calreticulin cycle for targeting for degradation to maintain proper ER homeostasis (Tannous et al., 2015). Mannose trimming plays a central role both in the extraction of terminally misfolded or slow folding glycoproteins from the calnexin/calreticulin cycle and the sorting for degradation to the ERAD process. Trimming of A-branch mannose residues precludes the ability of UGGT1 to modify the glycan for calnexin and calreticulin rebinding. Whereas removal of B- and C-branch mannoses creates a degradation signal that is recognized by downstream sorting lectins that target the demannosylated substrates for ERAD. The secretory pathway possesses a number of mannosidases that trim the various mannose residues including ER mannosidase I (ER Man I also called Man1B1), the EDEM family members (EDEM1-3), Golgi α1,2-mannosidases (Golgi Man IA, IB, and IC) and endomannosidase that help to support these functions (Figure 1.5) (Olivari & Molinari, 2007; Sunryd et al., 2014; Caramelo & Parodi, 2015).

The extraction of substrates from the calnexin and calreticulin binding cycle appears to be cell type dependent. Some cells possess an endomannosidase activity capable of cleaving the A-branch mannose residues and abolishing UGGT1 modification. However, this activity is missing from Chinese Hamster Ovary (CHO) cells and possibly other cells (Karaivanova et al., 1998). While EDEM1 has been shown to be able to extract substrates from the lectin chaperone binding cycle (Molinari et al., 2003; Oda et al., 2003), it is not clear if this is through direct binding to the substrate, mannose trimming of the A-branch glycans, or possibly the utilization of both properties. Further investigation is required to understand the full mechanism of extraction of misfolded substrates from the calnexin and calreticulin binding cycle.

Treatment with mannosidase inhibitors stabilizes ERAD substrates (Su et al., 1993). Mannose trimming on B- and C-branches is linked to creating a degradation signal on ERAD glycoprotein substrates. Initially, it was thought that the transition of the N-linked glycans to $Man_8GlcNAc_2$ sorted the protein for ERAD (Jakob et al., 1998). ER Man I/Man1B1 has been implicated in removing the B-branch terminal mannose and recognizing tertiary and quaternary structure (Aikawa et al., 2012; Shenkman et al., 2018). The activity of ERManI/Man1B1 is enhanced in the presence of the oxidoreductants PDI or TXNDC11 (Shenkman et al., 2018). In *S. cerevisiae*, the ER Man I equivalent, Mns1p, removes the outermost mannose from branch B, creating an eight mannose residue glycan (termed M8B) (Jakob et al., 1998; Słomińska-Wojewódzka & Sandvig, 2015). More recent studies discovered that degradation signaling required an additional ER mannosidase in yeast, Htm1p, that removes the C-branch terminal mannose, to expose an α1,6-linked mannose residue, and it is this residue that is the

signal for ERAD (Jakob et al., 1998; Quan et al., 2010). The ER lectin yos9p recognizes and binds ERAD substrates containing exposed or terminal α1,6-linked mannoses and targets them to the Hrd1p-containing dislocon complex in the ER membrane for dislocation, ubiquitination and degradation by the proteasome (Bhamidipati et al., 2005; W. Kim et al., 2005; Szathmary et al., 2005; Denic et al., 2006).

The process in metazoans appears to be more complex as the mannosidases involved have diversified. The EDEM family of three (EDEM1-3) serves as homologues to Htm1p. While all three EDEM family members have now been shown to possess mannosidase activity in cells, as well as using purified proteins more recently (Olivari et al., 2005; Hosokawa et al., 2010; Ninagawa et al., 2014; Lamriben et al., 2018; Shenkman et al., 2018), the precise glycans they act upon and generate remains uncertain. Following the yeast paradigm, ER ManI/Man1B1 and the EDEMs work together to create Man7-5 glycans with exposed α1,6-linked mannose residues on the C-branch (Ninagawa et al., 2014). These demannosylated ERAD substrates are recognized by two downstream mannose-receptor homology (MRH) domains containing lectins that reside in the ER; OS9 and XTP3B (Hosokawa et al., 2008; Christianson et al., 2009; van der Goot et al., 2018). It is these carbohydrate-binding proteins that then target the substrate to the HRD1 dislocation/ubiquitination complex in the ER membrane. Multiple mechanisms likely protect substrates from improper trimming by mannosidases including mannosidase sub-compartmentalization and regulated concentration of mannosidases in the ER (Calì et al., 2008; Benyair et al., 2015).

The mannosidases involved in ERAD do not act as traditional glycosidases that trim glycans by transiently interacting with the carbohydrate. Some of them appear to

contain folding sensing properties directly or through their associated co-factors. Htm1p mannosidase activity is aided by an associated oxidoreductase, Pdi1p (Gauss et al., 2011; C.-Y. Liu et al., 2016). While EDEM1 appears to stably bind ERAD substrates independently of its associated factors BiP and ERdj5, its binding appears to be bi-partite in that it has a stable interaction that survives harsh treatments that is thiol dependent and likely covalent, as well as a weaker interaction with substrates that is thiol independent (Cormier et al., 2010; Lamriben et al., 2018). The associated ERdj5 appears to assist in the reduction of substrates to make them competent for dislocation but is not required for thiol dependent binding (Ushioda et al., 2008; Cormier et al., 2010; Ushioda et al., 2013; Lamriben et al., 2018). BiP is recruited to the EDEM1 complex by the J-domain of ERdj5. BiP may also contribute to ERAD substrate selection, translocation preparation and dislocon complex delivery in some way.

The role for EDEMs in protein quality control is further complicated by the fact that they also appear to bind and select substrates in a glycan-independent manner (Cormier et al., 2010; Tang et al., 2014; Lamriben et al., 2018). In the case of EDEM1, the mannosidase-like domain can also act as a lectin and aid in the delivery of ERAD substrates to downstream dislocation complex by binding to the N-linked glycans on the HRD1 complex adapter, SEL1L (Hrd3p homologue in yeast) (Cormier et al., 2010; Saeed et al., 2011). While mannose trimming clearly marks terminally misfolded substrates for turnover through the ERAD process, the versatility and possible redundancy of lectin quality control factors has complicated the full understanding of their functions and mechanisms.

Thiol-dependent protein quality control

The ER is an oxidizing environment and the site of disulfide bond formation between proximal Cys pairs on maturing proteins. Disulfide bonds are crucial for the structure and activity of many proteins that traverse the secretory pathway. Numerous oxidoreductase proteins in the ER aid in the formation, reduction, and isomerization of disulfide bonds in order to ensure the correct native disulfides are formed, which is dependent upon the redox potential of an active site disulfide (Figure 1.6). As most Cys are paired in disulfides before trafficking from the ER, free thiols can act as an indicator that a protein is non-native. The ER has quality control machinery that recognizes proteins with free thiols. While extensive studies have been performed to elucidate the mechanism of disulfide bond formation and the factors involved, here we will focus on thiol-dependent protein quality control.

Protein disulfide isomerase (PDI^f) is a large family of ER proteins with more than twenty members that play essential roles in disulfide bond formation and maintenance (Hatahet & Ruddock, 2007; Appenzeller-Herzog et al., 2008; Määttänen et al., 2010; Saeed et al., 2011; Sato et al., 2013). We will focus on a subset of PDI^f which are well characterized, though other proteins of the PDI family exist which may interact with a smaller subset of clients (TXNDC5, TXNDC15, TMX1, etc.). A protein by the same name as the family is also used to identify the most abundant member of the family and PDI can help form, reduce, and isomerize disulfide bonds. The oxidative partner of PDI is ERO1α/β, allowing PDI to remain in a redox competent state (Sevier & Kaiser, 2008). PDI is comprised of four thioredoxin like domains; a, b, b' and a'. Of these domains, a and a' are active thioredoxins, containing catalytically active Cys-Xxx-Xxx-Cys motifs

(Oka & Bulleid, 2013). PDI appears to be able to bind substrates in both a redox dependent and independent manner as the b' domain of PDI binds to hydrophobic substrates (McLaughlin & Bulleid, 1998; Pirneskoski et al., 2004; Denisov et al., 2009). PDI may be capable of scanning proteins for misfolded regions and potentially oxidize, reduce, or isomerize disulfide bonds of clients in order to promote proper folding (Okumura et al., 2015). PDI is a versatile protein as it has also been shown to promote the degradation of misfolded substrates (Forster et al., 2006).

Two members of the PDI family associate with the P-domain of calnexin and calreticulin, ERp57 and ERp29, for recruitment to glycosylated substrates. Similar to PDI, ERp57 is comprised of four thioredoxin-like domains of which two are catalytically active (Frickel et al., 2004). ERp57 acts as an oxidoreductase and is brought in close contact with glycoprotein folding intermediates or retained substrates through its interaction with the carbohydrate-binding chaperones, allowing for scanning of substrates to ensure proper disulfide bonding (Zapun et al., 1998). ERp29 is unusual in that it is a dimer and lacks a catalytic active site as it only possesses a single Cys residue. It is possible that this Cys is used for isomerization. Its role in polyomavirus infection and virus disassembly and penetration across the ER membrane is analogous to the preparation and translocation of aberrant proteins for ERAD but this connection will require further investigation (Walczak & Tsai, 2011).

ERp72 consists of five thioredoxin domain; a°, a, a', b, and b', of which a°, a, and a' possess catalytically active Cys-Xxx-Xxx-Cys motifs (Mazzarella et al., 1990). Despite the high level of structural similarity to ERp57, ERp72 does not interact with calnexin or calreticulin, likely due to differences in exposed surface charges (Kozlov et

al., 2009). However, ERp72 interacts with cyclophilin B via a polyacidic stretch of amino acids on ERp72 (Jansen et al., 2012). This interaction increases the rate of *in vitro* folding of immunoglobulin G. ERp72 retains misfolded cholera toxin and thyroglobulin as shown by RNAi and overexpression studies (Forster et al., 2006). ERp72 has been shown to stably interact with and retain thyroglobulin through a disulfide bond (Menon et al., 2007). While more work is needed to fully understand the role of ERp72, these data suggest that ERp72, for a subset of clients, promotes folding and ER retention.

ERdj5 is the largest PDI family member. It has six thioredoxin domains and a J-domain that recruits BiP. Its redox state favors a role as a reductase for preparation of non-native proteins for dislocation to the cytoplasm for proteasomal degradation (Ushioda et al., 2008). ERdj5 can reduce glycosylated substrates through its association with EDEM1. BiP can also capture nonglycosylated ERAD substrates independent from EDEM1 for passage to ERdj5 and delivery to the SEL1L/HRD1 dislocation complex (Ushioda et al., 2013). Work from the Hendershot group using a peptide library, mapped ERdj5 binding sites to aggregation-prone sequences on a protein underscoring its role in quality control (Behnke et al., 2016). The structure of ERdj5 is dynamic as measured by high-speed atomic force microscopy and this flexibility in some way assists its ability to enhance ERAD (Maegawa et al., 2017). ERdj5 plays a central reductive role in the preparation of both glycosylated and non-glycosylated substrates for degradation.

While numerous proteins facilitate proper disulfide bond formation, errors can occur that could, if progressed unchecked, lead to secretion of non-native proteins with free Cys. This is potentially problematic as reactive thiols can enhance aggregation mediated by non-native intermolecular disulfide bonds. As such, ER quality control

recognizes proteins with non-native, reactive free thiols and retains them in the ER. ERp44 is a soluble chaperone of the protein disulfide isomerase family that cycles between the ER and early Golgi (Anelli et al., 2007). It contains three thioredoxin domains; a, b, and b', of which a is catalytically active (Määttänen et al., 2010). ERp44 uses Cys29 to retain non-native or unassembled substrates containing free thiols and this has been shown for substrates including unassembled IgM, adiponectin and SUMF1 (Tiziana et al., 2003; Qiang et al., 2007; Fraldi et al., 2008; Mariappan et al., 2008). The manner in which ERp44 binds and retains proteins in the ER is proposed to be pH dependent (Anelli et al., 2015; Watanabe et al., 2017). In this model, ERp44 traffics from the ER to the Golgi, where the pH is progressively lower. The C-terminal tail of ERp44 undergoes a pH-dependent conformational change in the Golgi, exposing the Cys29 necessary for client binding while also exposing the RDEL retention/retrieval sequence at its C-terminus (L. Wang et al., 2008; Vavassori et al., 2013). ERp44 then traffics back to the ER by interacting with KDEL receptors with its substrate covalently attached. It is currently unclear how ERp44 releases substrates upon re-entry into the ER, though the intermolecular disulfide bond is likely reduced by another member of the PDI family. In this manner, clients with free thiols can be retained in the ER via cycling between the ER and early Golgi.

Open questions remain regarding the extent to which free thiols act as hallmarks of misfolded proteins. Native proteins with free thiols exit the ER for secretion or residence in the lysosome, as in the case of lysosomal cysteine proteases with active site cysteines. How these proteins escape quality control recognition is unclear. One possibility is any unpaired Cys are buried and therefore not accessible to factors involved

in recognizing proteins with free thiols. Also, the nearby environment or context of the Cys may contribute to its pKa and reactiveness, and thereby determine its effectiveness in supporting retention and subsequent degradation. Additionally, non-native proteins with free thiols may evade thiol dependent quality control by forming non-native, intermolecular disulfide bonds, thereby generating disulfide linked protein aggregates which are resistant to degradation.

Summary

Sequence specific chaperone and oxidoreductase binding appears to be the basis for quality control recognition. Hallmarks of misfolded proteins include exposed hydrophobic domains and free thiols. These features can also impact the N-linked glycan composition to further signal aberrancy to the quality control system. As the quality control process is responsible for monitoring thousands of proteins that traverse the secretory pathway, the sequences that signal non-nativeness must be degenerate. While some factors appear to recognize substrates that are folding intermediates and should be given additional time to fold and can be repaired, others must recognize terminally misfolded proteins to target them for degradation. Chaperones play promiscuous roles in all these functions with their precise role in some cases being determined by their associated co-factors, while others possess more specialized activities. The interplay between these diverse quality control systems that relies on exposed hydrophobic residues, free thiols or carbohydrate compositions helps to evaluate the large variety of cargo that travel through the ER. The interplay is underscored by the many complexes found between traditional chaperones, oxidoreductases and lectin chaperones or

glycosidases. A deeper understanding of these quality control processes and their manipulation provides an avenue for disease intervention for the large number of diseases with etiologies rooted in quality control decisions.

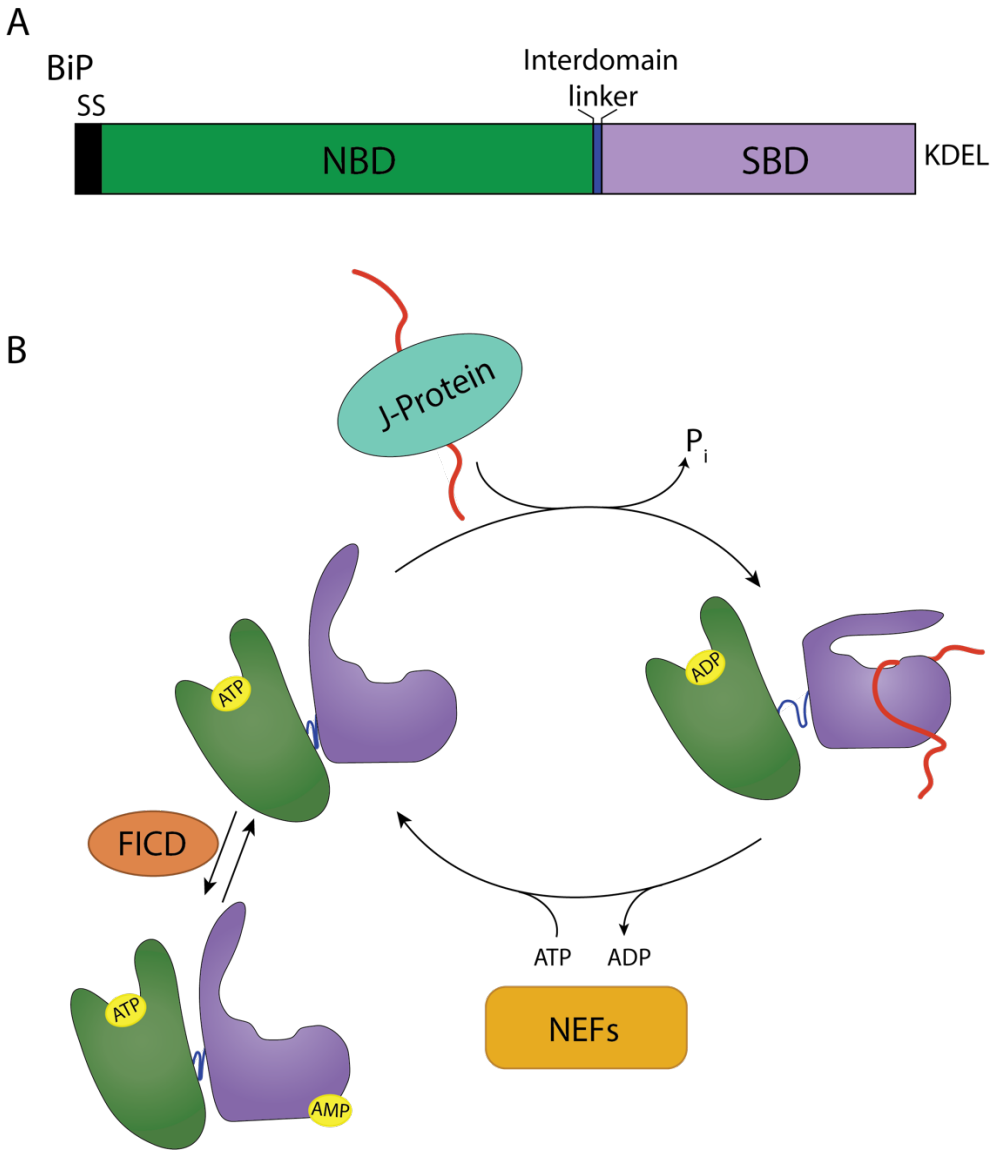

Figure 1.1 BiP domain architecture and binding cycle

(A) BiP is targeted to the ER via a signal sequence (SS) that is cleaved in the mature form of the protein. From N- to C-terminus, BiP is comprised of a nucleotide binding domain (NBD) (green), interdomain linker (blue), and substrate binding domain (SBD) (purple). It is retained in the ER via a KDEL motif. (B) The substrate binding cycle of BiP is regulated by ATP. When the NBD is bound by ATP, BiP is in a low substrate affinity state. Interaction with a substrate bound J-protein promotes ATP hydrolysis, leading to an extended conformation of the interdomain linker, SBD lid closing, and a high substrate affinity. A BiP nucleotide exchange factor (NEF) can then exchange ADP for ATP, placing BiP back in a low substrate affinity state. This process can be inhibited by AMPylation of BiP by FICD. AMPylation places BiP in a similar state to an ATP-bound state.

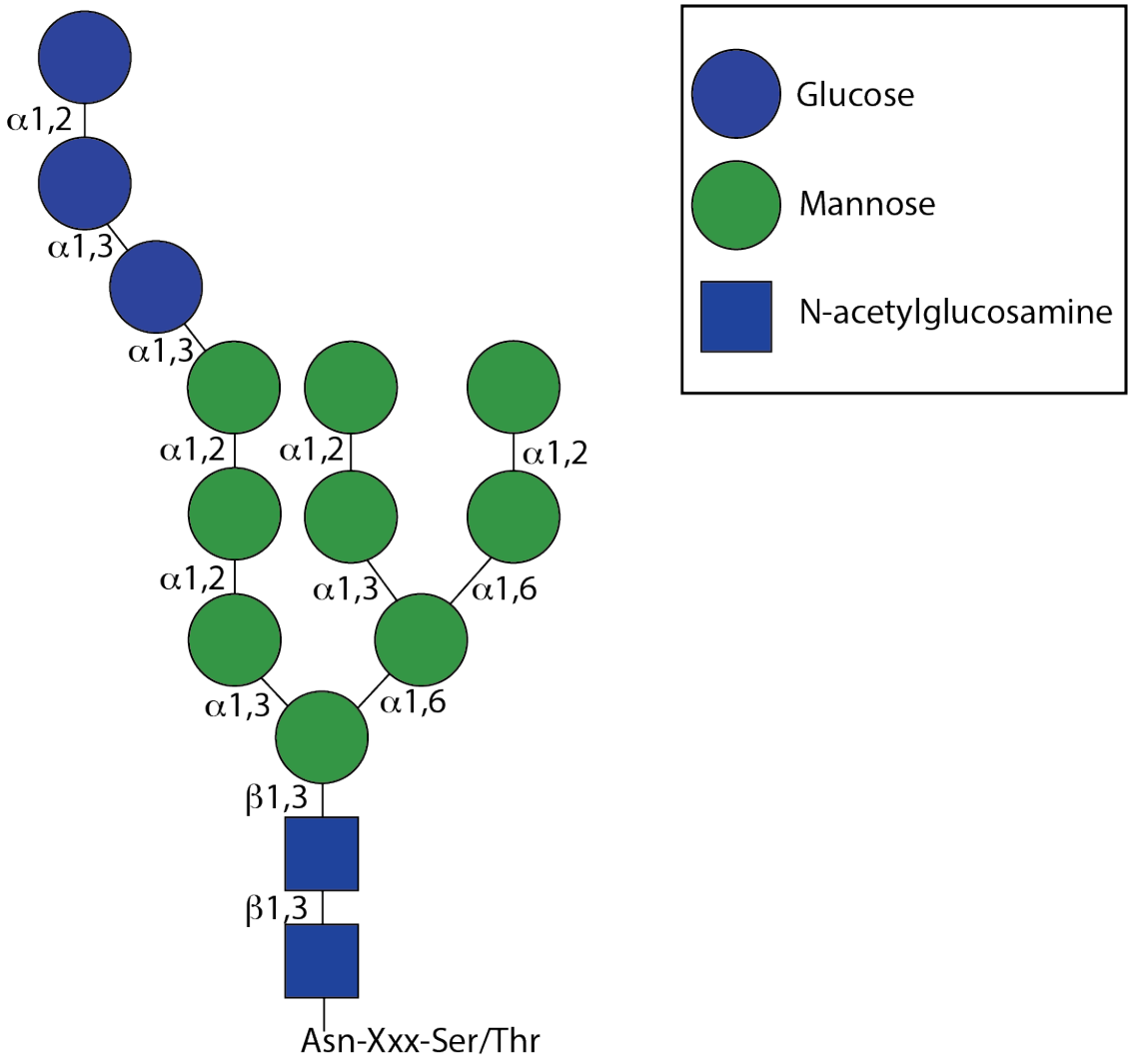

Figure 1.2 Structure of an N-linked glycan

N-linked glycans are transferred *en bloc* to an Asn residue in acceptor sites Asn-X-Ser/Thr/Cys, where X is not a proline. The precursor glycan is depicted, which can be dynamically remodeled during protein maturation. Glycosidic bonds are denoted.

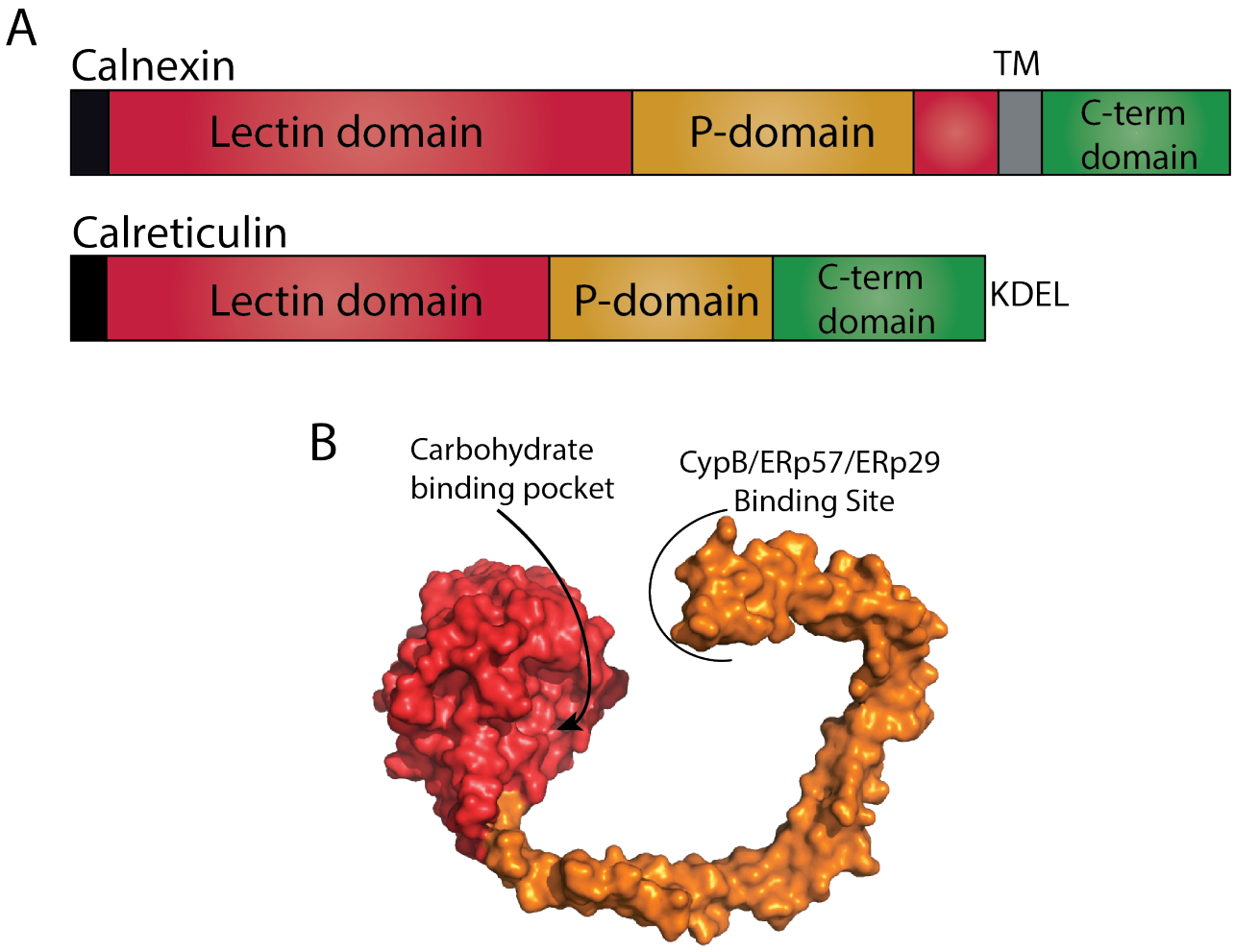

Figure 1.3 The domain architecture of calnexin and calreticulin

(A) Both calnexin and calreticulin possess an N-terminal signal sequence (black) that is cleaved in the mature protein. Calnexin possesses a lectin domain (red) that is composed of two regions separated by the P-domain (orange), a transmembrane region (TM) (grey) and a cytosolic C-terminal domain (green). Calreticulin possesses a contiguous lectin domain, a P-domain, a C-terminal domain, and a KDEL retention motif. (B) Surface representation of the crystal structure of the luminal domain of calnexin (PDB: 1JHN). The lectin domain is shown in red and the P-domain in orange. The carbohydrate binding pocket in the lectin domain and the binding site of CypB/ERp57/ERp29 on the tip of the P-domain are designated.

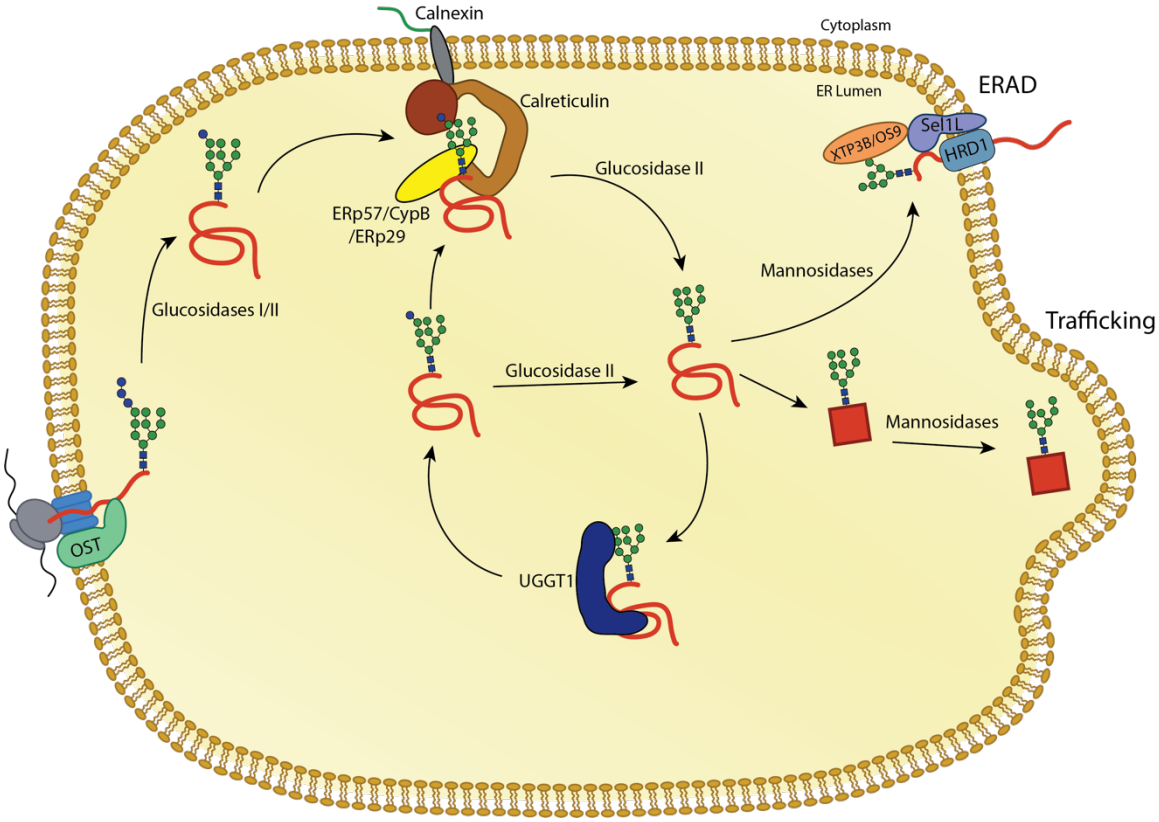

Figure 1.4 The calnexin/calreticulin substrate binding cycle

Proteins targeted to the ER receive N-linked glycans that are transferred by the OST complex to acceptor sites. The first two glucoses are trimmed by glucosidases I and II, leaving a monoglucosylated glycan. In this state, the glycan is a substrate for calnexin and calreticulin. Release from calnexin/calreticulin and trimming of the final glucose by glucosidase II leaves the glycan in a non-glucosylated state. Productive folding and adoption of a native state allows for trafficking of the glycoprotein from the ER. Glycoproteins that do not adopt a native fold can be recognized by the folding sensor UDP-glucose: glycoprotein glucosyltransferase 1 (UGGT1). UGGT1 reglucosylates substrates, allowing for rebinding to calnexin/calreticulin or trimming by glucosidase II. Glycoproteins that continue to non-productively fold can be removed from the calnexin/calreticulin cycle through trimming by mannosidases and targeting to ER associated degradation (ERAD) machinery.

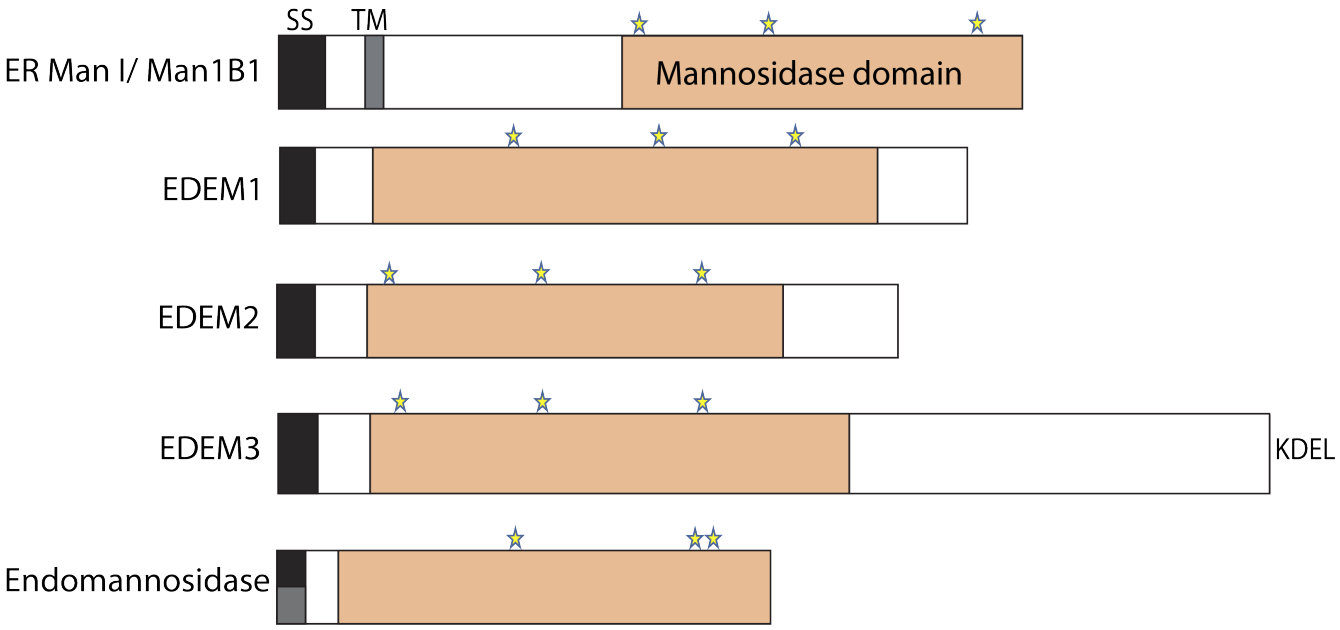

Figure 1.5 The architecture of mannosidases involved in quality control

Domain architecture of mannosidases: ER Man1, EDEM1, EDEM2, EDEM3, and endomannosidase. The signal sequences (black), predicted transmembrane domains (grey), the mannosidase domains (orange), and putative catalytic residues (stars) are designated. Endomannosidase possesses a predicted non-cleavable signal sequence, as shown by a half black and half grey box.

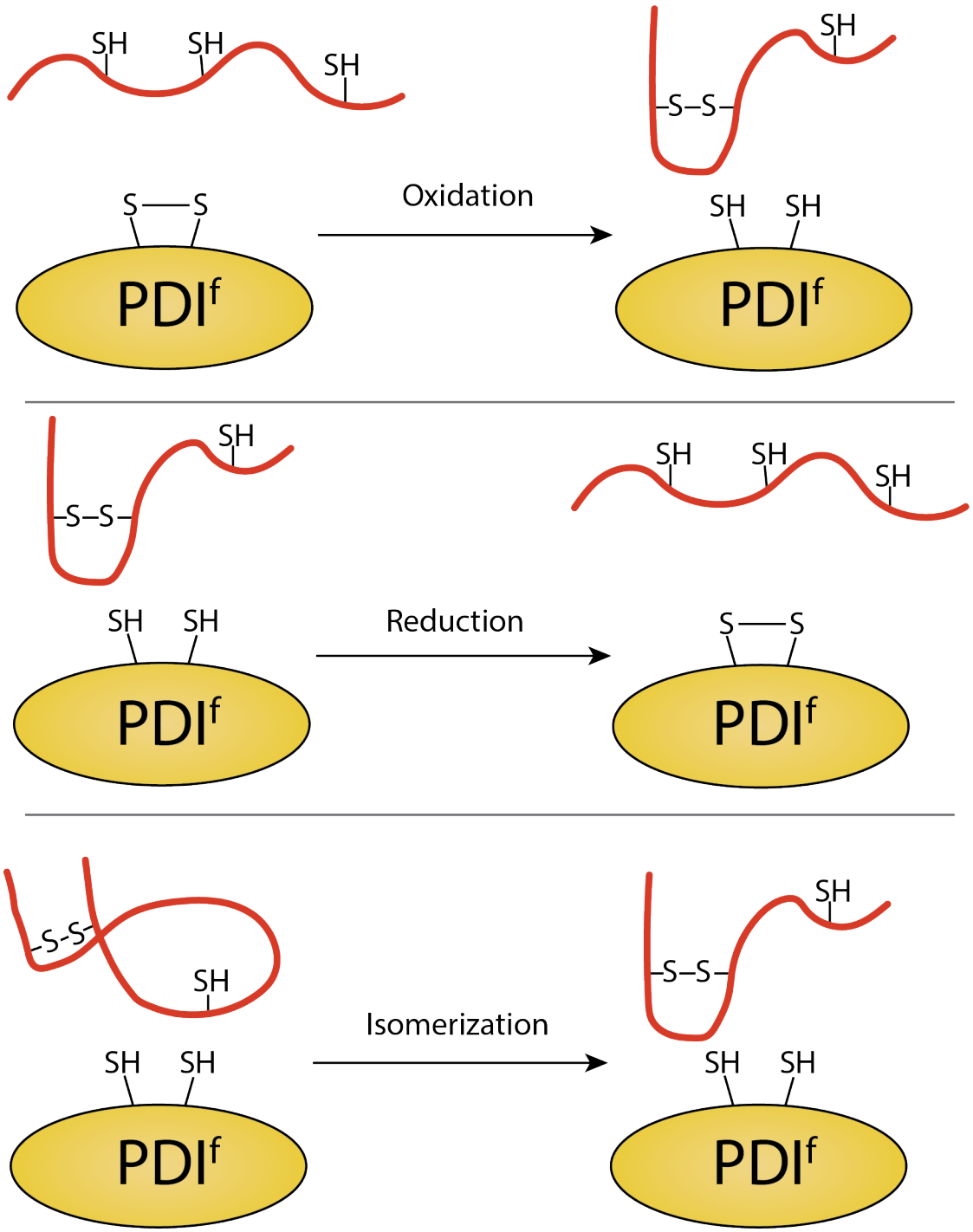

Figure 1.6 Redox reactions catalyzed by PDI family members

PDI, or members of the protein disulfide isomerase family (PDI^f), can oxidize, reduce and isomerize disulfide bonds of substrates. In all cases, a transient intermolecular disulfide bond is formed between PDI^f and the substrate (not pictured). While the various PDI^f members have varying numbers of catalytic domains, a single catalytic site of PDI^f is displayed for simplification.

CHAPTER 2

PROPER SECRETION OF THE SERPIN ANTITHROMBIN RELIES STRICTLY ON THIOL-DEPENDENT QUALITY CONTROL

Adams B.M[1,2]., Ke H[1]., Gierasch L.M[1.2.3]., Gershenson A., Hebert D.N.[1,2]

Published in Journal of Biological Chemistry, 2019

[1]Department of Biochemistry and Molecular Biology, University of Massachusetts, 240 Thatcher Road, Amherst, MA, 01003, USA
[2]Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, MA, 01003, USA
[3]Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003

Abstract

The protein quality control machinery of the endoplasmic reticulum (ERQC) works to ensure that clients are properly folded. ERQC substrates may be recognized as non-native by the presence of exposed hydrophobic surfaces, free thiols, and/or processed N-glycans. How these features dictate which ERQC pathways engage a given substrate is poorly understood. We used the human serpin antithrombin III (ATIII) to explore the interaction of ERQC systems in cells. Although ATIII has N-glycans and a hydrophobic core, we found that its quality control depended solely on free thiol content. Mutagenesis of all six Cys residues to Ala resulted in efficient secretion even though the product was not natively folded. ATIII variants with free thiols were retained in the ER but not degraded. These results provide insight into the hierarchy of ERQC systems and reveal a fundamental vulnerability of ERQC in a case of reliance on the thiol-dependent quality control pathway.

Introduction

Protein maturation is an error-prone process which, if allowed to proceed unchecked, would cause major cellular dysfunction. To prevent such calamities, quality control processes monitor the integrity of maturing polypeptide chains. For proteins that traverse the endoplasmic reticulum (ER), including secretory and plasma membrane proteins and proteins that reside in endomembrane compartments, proteins evaluated as native are allowed to exit to the Golgi. Proteins deemed non-native are initially targeted for ER retention and potential repair. If the non-native properties persist and the proteins accumulate in the ER, two processes can be initiated (Lamriben et al., 2016; Lederkremer, 2009). The unfolded protein response (UPR) is activated, causing a transcription-based remodeling of the ER contents in an attempt to maintain protein homeostasis (Kozutsumi et al., 1988; Hwang & Qi, 2018). If the interrogated proteins are determined to be irreparably or terminally misfolded, they are eventually targeted for destruction in order to recycle components and ensure that a potentially toxic misfolded substrate is not released from the ER or cell (McCracken & Brodsky, 1996; Olzmann et al., 2013).

A small number of quality control factors is responsible for evaluating the thousands of different client proteins that traverse the ER (Huh et al., 2003; Määttänen et al., 2010; Tamura et al., 2010**)**. Therefore, general protein hallmarks of foldedness must be queried to provide an efficient and plastic quality control process that supports the evaluation of a large number of substrates. Regardless of cellular location, exposed hydrophobic regions appear to be a defining signature of folding intermediates, misfolded proteins or unassembled oligomers – all forms of protein that are frequently recognized

by molecular chaperones and should be retained in the ER (Ellgaard & Helenius, 2003; Y. E. Kim et al., 2013). The ER is also an oxidizing environment that supports the formation of disulfide bonds assisted by a family of ER-resident oxidoreductases (Bulleid & Ellgaard, 2011). These oxidoreductases are involved in a thiol-dependent quality control process (Isidoro et al., 1996; Z. V. Wang et al., 2007; Anelli et al., 2015). The vast majority of proteins that travel through the mammalian secretory pathway are also modified in the ER with multiple 14-residue N-linked oligosaccharides with the composition $Glc_3Man_9GlcNAc_2$ (Helenius & Aebi, 2004). The maturation of the glycan provides a quality control code that reports on the fitness of the attached protein (Hebert & Molinari, 2012). Glycosidases and transferases orchestrate the glycan composition based on protein structural features, supporting binding and sorting of glycoproteins by quality control carbohydrate-binding factors. Among these are the lectins calnexin and calreticulin, which bind monoglucosylated glycans (Hammond et al., 1994; Hebert et al., 1995; Peterson et al., 1995). While the use of multiple quality control interrogation mechanisms that are based on the features of the specific protein is thought to maximize coverage and minimize mistakes, the interplay between these various quality control mechanisms is poorly understood. For instance, if there is overlap or redundancy in coverage, how is it determined which quality control process dominates? Or can multiple pathways query the same protein?

Serpins have been used extensively as model substrates to study protein quality control, as mutations in several inhibitory serpins are associated with degradation and accumulation of these proteins in the ER (Y. Liu et al., 1997; Gooptu & Lomas, 2009). Functional inhibitory serpins are metastable; they fold to a kinetically trapped state that

allows them to store potential energy, which is deployed to inhibit their cognate protease (Huntington et al., 2000; Dementiev et al., 2006; Gooptu & Lomas, 2009; Corral et al., 2018). This metastability and the need to avoid alternative conformations that might be more stable when folding likely make serpin maturation particularly problematic and may provide an explanation for why serpin misfolding is associated with a large number of pathologies or serpinopathies (Davies & Lomas, 2008). Functional, properly folded, inhibitory serpins form inactive covalent complexes with their cognate proteases providing a simple activity assay to determine whether a secreted serpin has successfully folded.

The human serpin antithrombin III (ATIII) is an inhibitory serpin that is modified by multiple N-linked glycan. The three native ATIII disulfide bonds have been used to map its cellular folding pathway (Chandrasekhar et al., 2016). ATIII inhibits thrombin thus playing an essential role in the blood coagulation cascade. Mutations in ATIII are commonly associated with thrombosis, and some of these mutations also lead to accumulation of ATIII in the ER (Perry & Carrell, 1996; Corral et al., 2018). Therefore, ATIII provides an apt model substrate to explore various ER quality control processes. Here, we made use of ATIII variants with disrupted folding by mutating native Cys residues to Ala. We hypothesized that these variants would be targets for the thiol-based quality control pathway. We also predicted that these variants would be compromised in their folding in such a way that they would present non-native hydrophobic surface and potentially glycan signals. Thus, these variants should reveal the interplay and hierarchy of ER quality control systems.

We found that quality control of ATIII relied solely on thiol-dependent quality control and diversion from the thiol-dependent quality control pathway led to improper secretion of misfolded and inactive protein. These results demonstrate that ER quality control pathways do not necessarily act redundantly and the general features of a substrate do not automatically dictate which quality control pathways will be engaged.

Results

Misfolded and inactive Cys-less ATIII is efficiently secreted

As shown previously, the three intramolecular disulfide bonds of ATIII (C8/C128, C21/C95, and C247/C430) are required for folding to the functional, metastable conformation (Chandrasekhar et al., 2016). We began to explore how the ERQC handles misfolded ATIII by determining the fate of ATIII with all its Cys mutated to Ala (Cys-less ATIII). Chinese hamster ovary (CHO) cells transfected with wild type (WT) or Cys-less ATIII were pulsed for 30 min with [^{35}S]-Met/Cys and chased for the designated times (Figure 2.1A). ATIII from the lysates and media was then immuno-isolated using antibodies directed towards the C-terminal Myc-tag present on the ATIII constructs. Samples were analyzed by reducing SDS-PAGE and phosphorimaging. As expected, secreted ATIII shows two bands due to incomplete glycosylation at Asn135 (Picard et al., 1995). Surprisingly, the level of secretion of WT and Cys-less ATIII was similar, reaching levels in the media after 3 hr of chase of 48.0% and 43.6%, respectively (Figure 2.1A and B). Strikingly, the remaining in-cell ATIII fraction for both constructs remained stable as degradation did not appear to occur even for Cys-less ATIII (Figure 2.1A, lanes 1-4, 9-12).

Normally, the ERQC machinery interrogates proteins as they traverse the secretory pathway sensing whether their native structures have been reached (Hurtley & Helenius, 1989; Määttänen et al., 2010; Adams, Oster, et al., 2019), and only properly folded proteins are packaged into COPII vesicles for eventual secretion, while misfolded or non-native structures are directed for ER retention (C. K. Barlowe & Miller, 2013). To assess the completeness folding of Cys-less ATIII, the activity of the product secreted into the media was analyzed by gel shifts due to its ability to form a covalent inhibitory complex with its target protease, thrombin. When the protease thrombin cleaves the reactive center loop of the serpin ATIII, an acyl bond is formed between the Ser in the loop of ATIII and thrombin. The formation of the covalent complex can be visualized by a gel shift on SDS-PAGE. Inactive ATIII does not demonstrate this gel shift, while partially active ATIII, though not functional as a thrombin inhibitor, can act as a substrate for thrombin while not forming a covalent bond. This leads to ATIII being cleaved and running at a lower molecular weight (Figure 2.2A). Cells were pulsed for 30 min, chased for 3 hr, and ATIII from lysates and media samples were immune-isolated. Protease inhibition by serpins requires large conformational changes (Huntington et al., 2000; Dementiev et al., 2006). Therefore, prior to performing the activity assay, ATIII was eluted from the beads and antibody using excess Myc-peptide to remove steric constraints. The eluted fraction was equally divided between thrombin-treated and non-treated samples and analyzed by reducing SDS-PAGE. While WT ATIII in the media was active as shown by the formation of a thrombin and ATIII complex (Figure 2.2A, lane 4), Cys-less ATIII was completely inactive (Figure 2.2A, lane 8).

To further characterize the state of folding of the secreted Cys-less ATIII, we used protease sensitivity. A natively folded protein should be relatively resistant to protease digestion as compared to an unfolded protein as, generally, natively folded proteins adopt more compact conformations that expose fewer sites for cleavage. Cells were pulsed for 30 min, chased for 3 hr, and ATIII was immunoprecipitated from media samples. Samples were then divided equally between untreated samples and trypsin treated for either 0, 15, 30, or 60 minutes. Samples were then analyzed by SDS-PAGE. Secreted Cys-less ATIII showed significantly higher protease sensitivity compared to WT (Figure 2.2B and 2.2C). This result indicates that Cys-less ATIII is secreted like a natively folded protein, despite the fact that it is an inactive and significantly misfolded protein. Both the secretion of Cys-less ATIII and its stability in the cell suggests that it is evaluated as properly folded by the ERQC network even though it has not achieved its native, functional fold.

There is a formal possibility that Cys-less ATIII is being improperly secreted and misevaluated by the ER quality control network because it is secreted via an unconventional secretory pathway, thereby escaping ERQC. In traversing the full secretory pathway, glycoproteins trafficked from the ER pass through the Golgi where their glycans are extensively remodeled to receive complex glycans. However, proteins that pass through recently identified pathways for unconventional secretion, both including and excluding the ER, can bypass portions of the canonical secretory pathway such as the Golgi and thus lack remodeled complex glycans (Fatal et al., 2002; Rabouille, 2017). The differential susceptibility of secretory proteins to glycosidases PNGaseF, which can cleave both complex and high-mannose glycans acquired in the Golgi, and

EndoH, which can only cleave high-mannose glycans, can be used to ask whether a given protein has traversed the canonical secretory pathway. Proteins carrying glycans found in the ER or on secreted proteins that have bypassed the Golgi are sensitive to both PNGaseF and EndoH, while proteins carrying glycans that have passed through the Golgi are sensitive to PNGaseF and resistant to EndoH. Glycosidase sensitivity, analyzed by size shifts using SDS-PAGE, can therefore be used as an assay for unconventional secretion. As expected, both WT and Cys-less ATIII in the cell lysate were sensitive to PNGaseF and EndoH, indicating that cellular ATIII mainly resides in the ER. In contrast, both WT and Cys-less ATIII from the cell media were sensitive to PNGaseF and resistant to EndoH (Figure 2.2D, lanes 5, 6, 11 and 12), showing that both had received complex glycans due to passage through the Golgi. Therefore, the Cys-less variant of ATIII utilizes the conventional secretory pathway.

Taken together, these results indicate that Cys-less ATIII, an inactive and misfolded protein, evades ER protein quality control and instead is secreted similarly to WT ATIII. These results support the hypothesis that free thiols are essential for ERQC recognition of misfolded ATIII and its consequent retention in the ER. Concomitantly, evasion of thiol-dependent quality control in this case allowed improper secretion rather than attempts by a complementary quality control branch to correct the misfolding or action of the ER-associated degradation pathway to eliminate the misfolded product.

Cellular retention and characterization of ATIII disulfide mutants

According to the recently discovered cellular pathway of disulfide bond formation for ATIII, the C-terminal disulfide, C247/C430, must form first in order for the two N-

terminal disulfides, C8/C128 and C21/C95, to form properly (Chandrasekhar et al., 2016). If the C-terminal disulfide is mutated, this model predicts that the remaining two disulfides would not form, leading to the generation of free thiols that may support ER retention. In contrast, this model predicts that mutating either pair of Cys residues that comprise the two N-terminal disulfides, C8A/C128A or C21A/C95A, allows the two remaining disulfides to form, reconciling all thiols into disulfides. This previous work, which elucidated the disulfide formation pathway of homogeneously glycosylated ATIII missing the partially recognized N-linked glycan at position Asn135, led to the predictions that ATIII quality control is mediated, at least in part, by free thiols directing ER retention (Chandrasekhar et al., 2016).

To characterize the disulfide requirements for proper ATIII quality control and secretion, and to test the predictions for ATIII disulfide mutants, the secretion rates of the three disulfide mutants (C8A/C128A, C21A/C95A, C247A/C430A) and WT ATIII (all constructs possessing all four native glycosylation sites) were analyzed via pulse-chase in cells. Transfected cells were pulsed for 30 min with [^{35}S]-Met/Cys and chased for the designated times. Samples were analyzed by non-reducing and reducing SDS-PAGE and phosphorimaging (Figure 2.3A).

Secretion of C247A/C430A ATIII was significantly reduced compared to WT, C8A/C128A and C21A/C95A ATIII, (Figure 2.3A, 13-15 and Figure 2.1B). After a 2 hr chase, WT secretion levels reached 50% compared to 8% for C247A/C430A ATIII. The secretion of the N-terminal disulfide mutants, ATIII C8A/C128A and C21A/C95A, was similar to WT secretion levels. Non-reducing SDS-PAGE displayed an increase in disulfide linked adducts for all mutants compared to WT ATIII. Similar levels of

aggregated ATIII were found in the triton insoluble fractions for all constructs (Figure 2.3A, lanes 16-18).

While secreted WT ATIII was active, as observed by the formation of the thrombin-ATIII complex (Figure 2.4A, lane 4), the disulfide mutants displayed little activity and did not form an inhibitory complex (Figure 2.4A, lanes 8, 12 and 16). This demonstrates that all of the ATIII disulfide mutants analyzed are inactive proteins, including C8A/C128A and C21A/C95A ATIII, which were efficiently secreted into the cell media. Thus, like Cys-less ATIII, the ATIII N-terminal disulfide mutants were mistakenly evaluated as natively folded by the ER quality control process. By contrast, the inactive C247A/C430A ATIII construct, which lacked the key early forming disulfide previously shown to be critical for initiation of proper folding, was properly retained.

A question that emerges from these data is whether secretion correlates with protection or inaccessibility of thiols. To address this question a polyethylene glycol-maleimide (PEG-Mal) modification gel-shift assay was performed on the ATIII mutants. PEG-Mal modifies exposed free thiols, and the bulky PEG group causes an increase in mass that can be visualized by a gel shift. Transfected cells were pulsed for 30 min then chased for 30 min. Lysate and media samples were split equally, and one half was treated with PEG-Mal, and analyzed by reducing SDS-PAGE.

For all ATIII constructs analyzed, secreted ATIII found in the media had insignificant levels of PEG-Mal modification, which in turn indicates near absence of accessible free thiols (Figure 2.4B, lanes 4, 8, 12 and 16; and Figure 2.4C for quantification). This result suggests that in ATIII constructs that are successfully secreted to the media, any thiols had either formed disulfide bonds or were otherwise inaccessible,

most likely through partial folding. The level of accessible thiols modifiable by PEG-MAL for WT, C8A/C128A, and C21A/C95A ATIII constructs in cell lysates was similar, ranging from 55 to 72% (Figure 2.4E). In contrast, the level of C247A/C430A ATIII modified by PEG-Mal was 89%, arguing that in this construct there is little sequestration of the Cys thiols. This result supports the model that the disulfide between C247 and C430 is critical for folding, and its absence leads to poor formation of the two N-terminal disulfides (Chandrasekhar *et al*., 2016). Free thiols are present in all forms of ATIII in the cell lysates, including WT, because the intracellular pool of protein is sampled during formation and isomerization of intermediate disulfide bonds. These results suggest that C247A/C430A ATIII is retained in the ER because its folding is impaired so that its free thiols cannot form disulfide bonds, while C8A/C128A and C21A/C95A ATIII are secreted because they can initiate folding and hence partially form disulfides, which together with partial folding, protect their free thiols. These findings support both the hypotheses that the C-terminal disulfide of ATIII must form before the N-terminal disulfides in order for ATIII to fold correctly and that free thiols lead to retention in the ER.

A single Cys is sufficient to retain ATIII in the ER

As the results in Figures 2.1, 2.2, 2.3, and 2.4 suggest that free thiols play a significant role in the ER retention of ATIII, we next investigated whether the presence of a single Cys would be sufficient to retain ATIII in the ER. To this end, six single Cys mutants were created by adding the individual Cys back at their natural sites in the Cys-less ATIII background (A8C, A21C, A95C, A128C, A247C, A430C) and their secretion

was analyzed by pulse-chase. Cells were pulsed for 30 min and chased for 10, 60, and 120 min. When secretion was analyzed by reducing SDS-PAGE, three single Cys mutants, A8C, A21C, and A247C, were poorly secreted as compared to Cys-less ATIII. A128C and A95C were secreted at a similar level to Cys-less ATIII. While there was no significant difference between secretion of A430C and Cys-less, A430C secreted to a lower level than Cys-less (Figure 2.5A, B). When secretion was analyzed by non-reducing SDS-PAGE, a high molecular weight band of secreted single Cys ATIII was present, suggesting that a portion of secreted single Cys ATIII mutants was in the form of a redox-dependent complex in which the Cys was no longer a free thiol but rather was in an intermolecular disulfide (Figure 2.5A, top NR image). Altogether, these results demonstrate that thiol-dependent quality control is a robust retention mechanism as a single free thiol led to the cellular retention of ATIII. These results also suggest that the location of the free thiol has relatively little influence as the majority of the thiols lead to retention of ATIII.

C247A/C430A ATIII is stably retained in the ER in a redox-dependent complex

We sought to obtain a better understanding of the cellular fate of the C247A/C430A ATIII, which was inefficiently secreted. CHO cells expressing WT, Cys-less, or C247A/C430A ATIII were imaged by immunofluorescence confocal microscopy, and co-localization with ER (KDEL) and Golgi (giantin) markers was monitored. All three variants were found throughout the ER as observed by extensive co-localization with the KDEL ER marker (Figure 2.6A). Some co-localization of all three constructs with the Golgi marker was also observed. These results indicated that regardless of their

different fates and properties, after 16 h of expression, these three ATIII variants localize throughout the ER.

Despite their similar cellular localization, we hypothesized that misfolded ATIII variants may not be monomeric within the ER. We therefore investigated whether ATIII variants were present in complexes within the ER. Cell lysate was layered on top of a 10-40% sucrose gradient and subjected to ultracentrifugation. Fractions were then analyzed by SDS-PAGE and immunoblotting. Both Cys-less and C247A/C430A ATIII are found in complexes that are larger than WT (Figure 2.7). Addition of the reducing agent dithiothreitol (DTT) post cell lysis causes C247A/C430A ATIII to shift to lower molecular weight fractions while addition of DTT had no effect on WT and Cys-less ATIII, suggesting that C247A/C430A is in a complex formed via accessible thiols and that retention of C247A/C430A is redox-dependent. Cys-less ATIII is also in a complex in the media, which may be due to secretion as an aggregate or secretion while bound to a partner. These data suggest that C247A/C430A ATIII is retained in a multimeric, disulfide-dependent complex in the ER while ATIII Cys-less is secreted in a non-monomeric state.

WT ATIII is more efficiently reglucosylated than Cys-less and C247A/C430A ATIII

The carbohydrate-binding chaperone calnexin and its soluble paralogue calreticulin play an important role in glycoprotein folding, quality control, and ER retention (Hebert et al., 1996; Vassilakos, Cohen-Doyle, et al., 1996; Lamriben et al., 2016; Kozlov et al., 2017b). Within the calnexin/calreticulin cycle, the folding sensor UDP-glucose:glycoprotein glucosyltransferase 1 (UGGT1) acts as a gatekeeper of

secretion by reglucosylating misfolded or incompletely folded proteins, which then allows calnexin/calreticulin to rebind the glycoprotein, leading to ER retention (Sousa & Parodi, 1995b; Pearse et al., 2008; Lamriben et al., 2016). Furthermore, recent x-ray crystal structures have shown that UGGT1 proteins from both *Thermomyces dupontii* and *Chaetomium thermophilum* have four thioredoxin-like domains (Satoh et al., 2017; Roversi et al., 2017). Oxidoreductases are frequently comprised of multiple thioredoxin domains, contributing to their oxidizing, reducing or isomerizing activities (Holmgren et al., 1975; Marin, 1995). Although the active oxidoreductase motif, CysXxxXxxCys, is not present in any of these domains, the multiple Cys residues present in UGGT1 and the thioredoxin-like folds likely aid UGGT1 in substrate recognition. We therefore investigated whether UGGT1 reglucosylation, and by extension the N-glycan quality control pathway, contributed to the ER retention of C247A/C430A ATIII.

We have previously developed a cell-based reglucosylation assay (Pearse et al., 2008, 2010; Tannous et al., 2015). Briefly, an *Alg6* defective CHO cell line, MI8-5 CHO, generates glycans lacking glucoses on their A-branches (Quellhorst et al., 1999). Therefore, in this cell line a monoglucosylated glycan can only be generated by UGGT1. In contrast, in WT cells, the presence of a monoglucosylated glycan could indicate either trimming of two glucoses from the original $Glc_3Man_9GlcNAc_2$ glycan or reglucosylation of an unglucosylated side chain by UGGT1. MI8-5 CHO cells were treated with N-butyl deoxynojirimycin (DNJ) for 30-min prior to the 30-min pulse and then throughout the indicated chase times. DNJ is a glucosidase inhibitor and therefore traps monoglucosylated proteins in their monoglucosylated state. Monoglucosylated proteins

were first pulled down using recombinant glutathione *S*-transferase (GST)-calreticulin, and from this pull-down ATIII was immunoprecipitated using anti-Myc antibodies.

While all three variants of ATIII, WT, Cys-less and C247A/C430A, were reglucosylated by UGGT1, WT ATIII was found to be modified most efficiently followed by C247A/C430A ATIII and then Cys-less ATIII (Figure 2.8A and B). As WT ATIII was recognized at a higher level than the other two ATIII variants, UGGT1 modification does not explain the greatly increased ER retention of C247A/C430A.

To explore the timing of reglucosylation, DNJ was added for only 15 min prior to each time point, which allows for detecting reglucosylation at that specific window of time. Using this experimental scheme, C247A/C430A ATIII was found to be reglucosylated more than either WT or Cys-less ATIII after 2 hr (Figure 2.8C and D). These results suggest that UGGT1 best recognizes WT ATIII, but once WT has folded and trafficked out of the ER, ATIII C247A/C430A becomes a better substrate for modification as it is still present in the ER where UGGT1 has access to it.

In order to confirm that UGGT1 is not retaining the free thiol-carrying mutant, C247A/C430A ATIII, the secretion of WT and C247A/C430A ATIII was examined in WT and *Uggt1*$^{-/-}$ mouse embryonic fibroblast (MEF) cell lines using a pulse-chase approach. Cells were pulsed for 30 min and chased for 0 or 2 hr before the media and lysate were collected. ATIII was immunoprecipitated from the media and lysate, and the percent fraction of ATIII in each pool was analyzed as previously described. WT ATIII was secreted similarly in both WT and *Uggt1*$^{-/-}$ MEF cells, suggesting that repetitive rounds of calnexin/calreticulin binding were not needed for efficient ATIII secretion (Figure 2.9A and B). In contrast, ATIII C247A/C430A was poorly secreted in either WT

or *Uggt1*$^{-/-}$ MEF cells, indicating that in the absence of UGGT1, the misfolded ATIII mutant C247A/C430A was still efficiently retained within the cell. Therefore, all together UGGT1 does not appear to play the determinative role in C247A/C430A ATIII quality control.

C247A/C430A ATIII is a poor ER-associated degradation (ERAD) substrate

When a misfolded protein is persistently retained in the ER, the ERAD process generally degrades the protein to maintain proper secretory pathway flow and protein homeostasis (Brodsky, 2012; Olzmann et al., 2013). Because C247A/C430A ATIII is retained in the ER, we would be expected it to be degraded. The stability of C247A/C430A ATIII was therefore compared to WT ATIII and the classical ERAD substrate alpha-1-antitrypsin *null* Hong Kong (A1AT NHK), which is also an inhibitory serpin and serves as a positive control for ERAD.

After an 8-hr chase, both WT and C247A/C430A ATIII were stable with 74% and 115% of the total protein remaining relative to zero time, respectively (Figure 2.10A and B). WT ATIII accumulated in the media, while C247A/C430A ATIII largely accumulated in the Triton-insoluble fraction. In contrast, A1AT NHK was turned over rapidly with 48% total protein remaining after 8-hr of chase and only a small fraction being secreted. Interestingly, though ATIII C247A/C430A was retained in the ER, it is a poor ERAD substrate as it was not efficiently degraded.

Overexpression of ATIII constructs activates the IRE1α arm of the UPR pathway

One possible explanation for the lack of degradation of C247A/C430A ATIII retained in the ER is that expression of this impaired protein does not activate the unfolded protein response (UPR). In turn, the basal level UPR pathways would be unable to process the high load of retained ATIII mutant. Thus, we sought to examine whether UPR pathways were induced by overexpression of ATIII variants in CHO cells. The UPR consists of three branches controlled by the ER sensors IRE1α, ATF6, and PERK, and each sensor initiates a signaling cascade, which can increase the folding capacity of the ER, decrease the protein load in the ER, or increase the capacity of ERAD (Walter & Ron, 2011). Thus, we carried out experiments to test whether any of the three UPR branches were upregulated upon expression of WT, C247A/C430A, or Cys-less ATIII or NHK A1AT.

Upon activation, IRE1α splices an unconventional intron from X-box protein 1 mRNA (*Xbp1*), creating a spliced form of *Xbp1* (*Xbp1s*). A frameshift caused by the splicing supports the translation of an active transcription factor, XBP1s, which leads to the up-regulation of multiple genes involved in protein folding and ERAD (Calfon et al., 2002; Plate et al., 2016). Therefore, the production of *Xbp1s* is indicative of IRE1α activation. Tunicamycin (Tm), which inhibits N-linked glycosylation and leads to a strong activation of the UPR, was used as a positive control. CHO cells were transfected either with the serpin variants or treated with Tm for 24 hr. RNA was then collected and cDNA was generated. In all cases, *Xbp1s* was generated, suggesting that IRE1α is activated by ATIII/A1AT overexpression regardless of the construct (Figure 2.11A). The level of activation during NHK ATIII and A1AT overexpression appears to be less than

that of Tm, suggesting that overexpression of the proteins was not maximally activating IRE1α.

Next, the second branch of the UPR, activation of the kinase PERK, was tested. During activation by ER stress, PERK undergoes trans-autophosphorylation, which activates PERK and leads to multiple downstream effects including up-regulating chaperone expression, translational attenuation, and cell cycle arrest (Shi et al., 1998; Hetz & Papa, 2018). In order to test for the activation of PERK, CHO cells were either transfected or treated with Tm for 24 hr. Cells were then lysed and trichloroacetic acid (TCA) precipitated before immunoblotting. Phosphorylated PERK (pPERK) can be distinguished from non-phosphorylated PERK by an increase in mass on a gel. Overexpression of ATIII variants and A1AT NHK did not generate pPERK (Figure 2.11B). This suggested that PERK was not activated by the overexpression of these proteins.

The third branch of the UPR, activation of the transcription factor ATF6, was also examined. Activation of ATF6 leads to the up-regulation of multiple ER chaperones and ERAD factors (Haze et al., 1999; Yoshida et al., 2001). One of the prime targets of ATF6 is the Hsp70 family member BiP (Plate et al., 2016). As such, the level of BiP expression was examined as an indicator of ATF6 activation. CHO cells were either transfected or treated with Tm for 24 hr. Cells were then lysed and TCA precipitated before immunoblotting. BiP expression was not found to be upregulated by the overexpression of any of the constructs when compared to the positive control Tm (Figure 2.11C and D). Therefore, ATF6 was not activated by the overexpression of ATIII or A1AT variants. In

total, these results indicate that the only branch of UPR activated by overexpression of ATIII variants and A1AT NHK was the IRE1α pathway.

C247A/C430A ATIII interacts poorly with ERAD factors

IRE1α activation leads to up-regulation of numerous ER-resident proteins including the ERAD factors ER degradation enhancing α-mannosidase-like (EDEM) proteins EDEM1, EDEM2, and EDEM3 (Yoshida et al., 2003; Bernasconi et al., 2008; Plate et al., 2016). Up-regulation of ERAD factors is expected to lead to robust degradation of misfolded and ER retained proteins as they are thought to be responsible for directing glycosylated misfolded proteins like C247A/C430A ATIII to the ERAD pathway (Molinari et al., 2003; Olivari et al., 2005; Hirao et al., 2006; Cormier et al., 2010; Araki & Nagata, 2011; Lamriben et al., 2018).. However, C247A/C430A ATIII was degraded significantly less than the ERAD substrate A1AT NHK (Figure 2.10A and B). A possible explanation is that C247A/C430A ATIII does not bind well to EDEM1, EDEM2 and EDEM3. To test this, we co-transfected CHO cells with either WT or C247A/C430A ATIII, or A1AT NHK, and EDEM1, EDEM2 and EDEM3. Cells were pulsed for 30 min and chased for the indicated times. Cells were lysed with buffer containing Triton X-100 (MNT). Interactions were queried by performing co-immunoprecipitations . EDEM1, EDEM2 and EDEM3 all associated efficiently with NHK A1AT as demonstrated by their co-immunoprecipitation with the A1AT pulldowns (Fig 2.12A-F, lanes 21 and 22). In contrast, neither WT nor C247A/C430A ATIII showed significant association with any of the EDEMs. We conclude that the misfolded and ER retained C247A/C430A ATIII is a poor substrate for recognition by these ERAD sorting

factors, thus providing a possible explanation for the stability of C247A/C430A ATIII retained in the ER.

Discussion

A central question regarding protein quality control in the ER is how thousands of proteins that pass through the secretory pathway are evaluated by a small number of quality control factors. The ERQC must efficiently and accurately enable natively folded proteins to exit to their final destination, and identify incompletely or misfolded proteins so that they are retained in the ER to allow fresh starts on proper folding or degradation. ERQC must rely on features of secretory pathway clients that indicate those with non-native characteristics, such as exposed hydrophobic residues, mispaired or free Cys residues, or processed N-glycans (Hammond et al., 1994; Anelli et al., 2015; Parodi & Caramelo, 2015; Behnke et al., 2016) (Figure 2.13). Exposed hydrophobic residues are most notably monitored by the BiP/ERdj network in the ER (Hendershot, 2004; Pobre et al., 2018; Preissler & Ron, 2018), while disulfides are formed and monitored for their integrity by a group of some 20 oxidoreductases or PDI family members (Hatahet & Ruddock, 2007; Määttänen et al., 2010; Tsunoda et al., 2014). Glycosidases, transferases, lectin chaperones and sorting factors comprise an N-linked carbohydrate-dependent glycoprotein quality control system (Parodi & Caramelo, 2015; Lamriben et al., 2016). How these different quality control pathways work together to ensure that only native proteins are passed along the secretory pathway is poorly understood.

In this work, we have focused on a member of the serpin family of secreted proteins to gain insight into ERQC. The serpin native fold is complex and challenged by the need to adopt a metastable state that is required for inhibitory activity. Many

examples of mutant serpins have been characterized to have secretion defects; often theses are associated with diseases arising from loss of function and/or toxicity caused by retention and aggregation in the ER(Chandrasekhar et al., 2016; Gooptu & Lomas, 2009; Ronzoni et al., 2016; Stein & Carrell, 1995). While the fold is conserved in the serpin family, the number and locations of Cys and disulfides bonds, as well as N-linked glycans, vary among serpins. Therefore, we anticipated that different serpins would engage different quality control systems. We selected ATIII for this study as it is a glycoprotein with six Cys that are paired into three intramolecular disulfides bonds in its properly folded native, functional state (Zhou & Smith, 1990; Chandrasekhar et al., 2016). Therefore, it is expected to engage the three quality control pathways: hydrophobic-, thiol- and glycan-dependent.

Our previous work provided insight into the cellular folding pathway of ATIII: the C-terminal disulfide between Cys 247 and Cys 430 must form first for the protein to achieve its native fold with the next two disulfides in place (Chandrasekhar *et al.*, 2016). Cys residues in proteins that traverse the ER are generally either buried or reside in disulfide bonds once a protein is natively folded and assembled (Oka & Bulleid, 2013; Anelli et al., 2015). Exposed, unpaired Cys present in unpaired oligomers can be recognized by the thiol quality control pathway and retrieved from the ER-Golgi intermediate compartment and retained in the ER via a pH-dependent interaction with ERp44, as in the case of IgM (Sitia et al., 1990; Anelli et al., 2015), acetylcholinesterase (Kerem et al., 1993), SUMF1 (Fraldi et al., 2008), and adiponectin (Z. V. Wang et al., 2007), while IgG C_H1 domain mutants are free thiol quality control substrates retained by an unclear mechanism (Elkabetz et al., 2005). However, little is known about the role of

the thiol dependent quality control process for monomeric, secreted proteins. Previous work has also been limited by a lack of a clear method to detect the folded status of secreted protein, thereby leaving the possibility that Cys mutants are properly folded and secreted rather than evading quality control as inactive and misfolded mutants. Therefore, we sought to delineate the role of oligomerization from thiol-dependent quality control while monitoring folded status by examining the quality control of a monomeric protein, ATIII.

Strikingly, mutation of all six Cys of ATIII to Ala resulted in the efficient secretion of a significantly misfolded and inactive protein, adding back a single native Cys to the Cys-less construct resulted in efficient ER retention in all but one case, and all disulfide mutants with free thiols were retained while disulfide mutants without free thiols were not. These compelling results indicate that ATIII relies almost completely on thiol-dependent quality control for its interrogation within the ER. Our data therefore expands the clientele of thiol dependent quality control by demonstrating that ATIII, a monomeric and secreted protein, is a substrate of thiol dependent quality control. Furthermore, it is notable that no other features of ATIII that signal incomplete folding or misfolding were exploited by ER quality control in order to retain the non-native species, demonstrating a lack of redundancy between ER quality control pathways.

Interestingly, there are other proteins that possess unpaired Cys residues in their native sequences and nonetheless are efficiently secreted, including the serpin A1AT which possesses a single Cys (Ronzoni et al., 2016). Additionally, when Cys 128 was added back to Cys-less ATIII, the resulting protein was secreted at a level near to that of the Cys-less variant. The observation of a higher molecular weight species in the non-

reducing gel for this variant suggests that a population formed a disulfide bond either in an ATIII dimer or with another luminal species, cloaking the free thiol and allowing secretion of this ATIII variant (Fig 2.5). Thus, not all free thiols designate ER retention, and there must be additional determinants that contribute to a retention outcome. The most obvious factor is the solvent accessibility of a Cys, but even for ATIII an active mutant with an extra solvent accessible Cys (R57C) is properly secreted, while other mutants with a seventh Cys (Y63C, F402C, Y2166C), are retained (Perry et al., 1995; Chandrasekhar et al., 2016), and some data suggest that Cys solvent accessibility does not impede secretion of functional WT A1AT (Griffiths et al., 2002; Patschull et al., 2011). How accessible single Cys residues are to thiol-reactive proteins in the cell is not well understood. Additional factors involved in recognition may include the pKa of a given Cys, as Cys displays a wide variety of reactivity which is dictated by interactions with residues in the local environment (Weerapana et al., 2010), the presence of local hydrophobic domains, or oxidative modifications of a Cys. Future work should address these questions in order to determine the specific requirements of a thiol dependent quality control substrate.

Why did the other quality control pathways that monitor structural features associated with protein misfolding or incomplete folding not recognize the Cys-less ATIII variant? It is clear from our results that this construct is not natively folded, as it is inactive and protease labile. A major cellular strategy to recognize “unfoldedness” in a substrate is the exposure of hydrophobic surfaces or sequences. In the ER lumen, the most well studied quality control mediators that utilize hydrophobic surface belong to the BiP-ERdj network. Serpins have an ellipsoidal, watermelon shaped fold which increases

their surface to volume ratio relative to other globular proteins that are more spherical (Dima & Thirumalai, 2004; L. Liu et al., 2014), suggesting the serpins are likely to be more hydrophilic than more common, more spherical protein folds. This higher hydrophilicity and lower hydrophobicity could interfere with BiP binding. Future work may explore the possibility of a poor interaction of Cys-less ATIII with BiP directly.

The carbohydrate chaperone system of the ER is responsible for directing the folding and retention of aberrant glycoproteins (Parodi & Caramelo, 2015; Adams, Oster, et al., 2019). While retention is mediated by binding to the ER-resident lectin chaperones, calnexin and calreticulin, the decision for chaperone rebinding or ER retention is made by UGGT1. Though UGGT1 is understood to reglucosylate proteins that it determines to be non-native, the parameters by which UGGT1 select substrates remains incompletely understood. One general possibility is that UGGT1 recognizes generally misfolded proteins, as in the case of misfolded mutants of A1AT (Tannous et al., 2015). A second hypothesis suggests that UGGT1 specifically recognizes on-pathway folding intermediates, so as to promote productive folding rather than futile rounds of reglucosylation of irreparable substrates (Caramelo et al., 2004). While it is hard to envision distinguishing features of on-pathway proteins from off-pathway targets, our cellular results are consistent with this model as WT ATIII was reglucosylated at a higher level than the misfolded off-pathway mutants of ATIII (Figure 2.9A and B). Studies using purified components have found that UGGT1 favors modifying glycopeptides that have hydrophobic patches C-terminal to the glycan (Taylor et al., 2003). Analyzing ATIII with this pattern in mind did not reveal such hydrophobic sequences C-terminal to the N-linked glycans; by comparison, A1AT, which is a better substrate of UGGT1 (Tannous et

al., 2015), displayed two strong hydrophobic patches following glycans at Asn 70 and 271 (Figure 2.14). This observation would account for the apparent poor recognition of ATIII by the glycan-dependent pathway of ERQC and underline its reliance on the thiol-dependent quality control pathway. The presence of the robust thiol dependent pathway for monitoring ATIII may have alleviated any evolutionary pressure to maintain motifs suitable for reglucosylation and the reliance on the UGGT1 directed carbohydrate-dependent quality control pathway. Further work examining the substrate preferences of UGGT1 in a cellular context are required to elucidate the selection process for this intriguing folding sensor.

The remaining puzzle in our findings is why persistent retention of misfolded ATIII variants does not correlate with degradation. We postulate that stability of retained ATIII mutants is promoted by weak interactions with the ERAD promoting proteins EDEM1/2/3 (Figure 2.12). Why C247A/C430A ATIII is not recognized by the EDEMs is unclear. However, the presence of higher molecular weights of C247A/C430A ATIII, as demonstrated by sucrose gradients (Fig 2.7), may indicate that C247A/C430A ATIII is present in small, disulfide linked aggregates with which the EDEMs are unable to interact. Despite weak degradation, cells expressing C247A/C430A ATIII did not show higher UPR activation as compared to cells expressing other ATIII variants (Fig 2.11). As such, it is unclear how cells are capable of maintaining proteostasis during retention of this substrate, though it is possible long term retention may more strongly activate UPR or potentially other mechanisms such as autophagy.

It is notable that even other members of the serpin family, which fold to very similar structures, do not use the same quality control pathway as ATIII. Neuroserpin,

though trafficked through the secretory pathway, lacks Cys residues and as such does not engage thiol-dependent quality control. Multiple serpins, such as serpinB3 (squamous cell carcinoma antigen 1 (SCCA1)), the human serpin with the highest sequence identity (39%) to ATIII, are cytoplasmic and therefore also must use unique quality control pathways as compared to secretory serpins. This suggests that universal protein quality control mechanisms are not generally applicable as even highly related proteins do not use the same quality control pathways. Rather, slight differences in specific substrate features dictate which quality control pathways are engaged.

Together these results point to hierarchies in ERQC that are tailored to particular proteins through the process of evolution. We found that the serpin ATIII, a protein containing multiple glycans and a hydrophobic core, and characterized in its native state by three disulfide bonds, is solely reliant on the thiol-dependent quality control pathway. In the case of ATIII, this heavy reliance on the thiol-dependent quality control pathway introduced a vulnerability, which was revealed by removing all six Cys residues from ATIII. The resulting protein was completely incapable of folding to a native, functional state, but nonetheless bypassed ERQC and was secreted efficiently. These results highlight the potential lack of redundancy between ER quality control pathways and demonstrate that general substrate features do not necessarily predict the quality control pathways of a substrate. Future studies should explore the reasons that ATIII variants were poorly recognized by key players in the other quality control systems as well as conduct a more detailed analysis of other serpins in order to elucidate the nuances of protein quality control and accurately predict the quality control pathways a substrate will

engage. This predictive understanding will open the door to therapeutic approaches to improve ERQC and thus address defects in protein maturation and secretion.

Experimental Procedures

Cell Culture

MI8-5 Chinese hamster ovary (CHO) cells were a gift from S. Krag (Johns Hopkins University, Baltimore, MD). CHO-K1 (Lot# 62960170) cells were purchased from ATCC. Cells were authenticated by universal mycoplasma detection kit (Cat. No. 30-012K, ATCC). WT MEF and *Uggt1*$^{-/-}$ MEF cells were a gift from R.J. Kaufman and were generated as previously described (Molinari et al., 2005; Soldà et al., 2007). CHO and MI8-5 CHO cells were grown in alpha-MEM media supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin at 37 °C and 34 °C, respectively, and 5% CO_2. MEF cells were grown in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin at 37 °C and 5% CO_2. All cell culture reagents were purchased from ThermoFisher Scientific.

Reagents

The plasmid for pGEX-3X GST-calreticulin was from M. Michalak (University of Alberta, Edmonton, Canada). Antibodies used were: Monoclonal mouse Myc-tag 9B11 antibody, rabbit monoclonal C33E10 PERK (Cell Signaling), monoclonal mouse HA-probe 12CA5 (Santa Cruz Biotechnology), monoclonal mouse KDEL 10C3 DyLight 488 (Enzo Life Sciences), Goat anti-mouse secondary antibody Alexa Fluor 594 (ThermoFisher), monoclonal mouse anti-FLAG M2 F1804 (Millipore-Sigma), rabbit

polyclonal anti-A1AT (Dako), and mouse monoclonal GAPDH 374 (Millipore-Sigma). Rabbit polyclonal affinity-purified anti-BiP was from L. Hendershot (St. Jude Children's Research Hospital). S-tag agaraose was purchased from Millipore-Sigma. All ATIII constructs, EDEM1, and EDEM2 were cloned into a pcDNA3.1(+) vector. pcDNA3.1(+)EDEM3-HA was from N. Hosokawa (Kyoto University). All chemicals were purchased from Millipore-Sigma, except where indicated.

Metabolic Labeling

Cells were pulse labeled for 30 min with 60 μCi of EasyTag Express^{35}S Protein Labeling Mix [^{35}S]-Cys/Met (PerkinElmer; Waltham, MA) in 3 cm plates and 120 μCi of [^{35}S]-Cys/Met in 6 cm plates. Immediately after pulse, cells were washed with PBS and either lysed in lysis buffer with protease inhibitors (MNT; 20 mM MES, 100 mM NaCl, 30 mM, 0.5% Triton X-100, 50 μM calpain inhibitor I, 1μM pepstatin, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 400 μM phenylmethylsulfonyl fluoride, and 20 mM NEM) or chased for indicated time using regular growth media. Media and Triton X-100 insoluble fractions were collected where indicated.

Immunoprecipitations and SDS-PAGE

After lysis, samples were vortexed at high speed at 4 °C for 5 min then centrifuged at high speed at 4 °C for 5 min. The supernatant was then pre-cleared using protein-A sepharose beads (GE Healthcare) by end-over-end rotation for 1 hr at 4 °C. Samples were then centrifuged at 3,000 rpm for 5 min at 4°C and the beads were discarded. Samples were then incubated with protein-A sepharose beads and the indicated

antibody overnight at 4 °C under end-over-end rotation. Samples were then washed with Connie's Wash (100 mM Tris-HCl pH 8.6, 300 mM NaCl, 0.1% SDS, 0.05% Triton X-100) or lysis buffer without protease inhibitors, where indicated. The Triton-X 100 insoluble pellet was either discarded or solubilized in 1% SDS in 100mM Tris-HCl, pH 8 by trituration followed by high speed vortexing at room temperature, heating for 10 min at 95 °C, dilution in lysis buffer and sonication. Samples were eluted from beads using Werner's sample buffer (30 mM Tris-HCl pH 6.8, 9% SDS, 15% glycerol, 0.05% Bromophenol Blue), and SDS-PAGE was performed. Radiolabeled samples were imaged by phosphorimaging using a GE Typhoon FLA 9500 phosphorimager (GE Healthcare) and quantified using ImageQuant (Fujifilm).

Secretion assay

Cells were seeded onto 3 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 min of pulse and the indicated time of chase, media and lysate portions were collected and immunoprecipitations were performed as previously described. After washing with Connie's Wash, sample buffer was added and samples were analyzed by SDS-PAGE and imaged by phosphorimaging.

Activity assay

Cells were seeded into 6 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 min of pulse and 3 hr of chase, the media and lysate portions were collected and immunoprecipitations were performed as previously described. After washing with 0.5% Chaps HBS (0.5% Chaps, 50mM HEPES

pH 7.5, 200mM NaCl), samples were eluted from beads with 10 ul of 0.5 mg/ml c-Myc peptide (Millipore-Sigma) for 1 hr at 37 °C. The beads were then discarded and the sample was split equally between treated and non-treated, 2 units of thrombin was added to treated samples and an equal volume of water to non-treated samples. Samples were then incubated for 1 hr at 37 °C before the addition of sample buffer. All samples were then analyzed by SDS-PAGE and imaged by phosphorimaging.

Protease sensitivity assay

Cells were seeded into 6 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 min of pulse and 3 hr of chase, the media and lysate portions were collected and split equally into treated and non-treated samples. Samples were immunoprecipitated as previously described. After washing with 0.5% Chaps HBS, 0.05 ug of trypsin was added to treated samples and an equal volume of water was added to non-treated samples. Samples were then incubated for 15, 30, or 60 mins at 37 °C before the addition of sample buffer. All samples were then analyzed by SDS-PAGE and imaged by phosphorimaging.

Glycosylation assay

Cells were seeded onto 3 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 min of pulse and 2 hr of chase, the media and lysate portions were collected and immunoprecipitations were performed as previously described. After washing, samples were treated with either EndoH or

PNGaseF (New England Biolabs), according to the manufacturer's instructions. Samples were then analyzed by SDS-PAGE and imaged by phosphorimaging.

Assay for the presence of free thiols

Cells were seeded onto 3 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 min of pulse and 30 min of chase, the media and lysate portions were collected and split equally into treated and non-treated samples. To treated samples, 100 µl of Polyethylene glycol maleimide (Millipore-Sigma) was added to a final concentration of 1.4 mM and to the non-treated samples 100 µl of N-ethyl maleimide was added to a final concentration of 5 mM. Samples were then incubated at room temperature for 30 min. Treated samples were then quenched by adding Dithiothrietol (DTT) (ThermoFisher) to a final concentration of 100 mM. Immunoprecipitations were then performed as previously described. All samples were then analyzed by SDS-PAGE, imaged by phosphorimaging, and quantified using ImageQuant (Fujifilm).

Immunofluorescence

CHO cells were seeded onto glass coverslips in alpha-MEM supplemented with 10% FBS. 24 hr later cells were transfected with the indicated plasmid using polyetheyleneimine (PEI) (Polysciences Incorporated; Warrington, PA) according to the manufacturer's instructions. After 16 hr, cells were washed with phosphate buffered saline (PBS) (2.7 mM KCl, 1.5 mM KH_2PHO_4, 136.9 mM NaCl, 8 mM Na_2PHO_4, pH 7.4) and fixed using 3.7% formaldehyde for 15 min, permeabilized using 0.1% Triton X-

100 in PBS for 15 min, then washed with PBS and blocked with 10% FBS in PBS for 30 min. Cells were incubated in Myc-tag antibody in 10% FBS in PBS (1:500) for 1 hr, washed with PBS, blocked with 10% FBS, and incubated with highly cross-adsorbed Goat-anti Mouse IgG secondary antibody conjugated to Alexa Fluor 594 (1:500) for 1 hr. Cells were then incubated with KDEL antibody conjugated to DyLight 488 (Enzo Life Sciences) (1:200) for 1 hr. Cells were then washed with PBS and mounted to glass slides using vectashield with DAPI (VectorLabs; Burlingame, CA) and sealed with nail polish. All images were acquired using an Olympus Fluoview FV1000 confocal microscope. Images were taken at 100X oil immersion. Images were processed with Adobe Photoshop.

Sucrose gradients

Cells were seeded onto 10 cm plates and transfected. Media and lysate fractions were collected. Lysate fractions were then split equally between DTT treated and non-treated samples. Treated samples contained 100 mM DTT. Sucrose gradients were made by solubilizing sucrose into MNT to a concentration of 10% or 40%. 6 ml of 10% sucrose was then laid on top of an equal volume of 40% sucrose and the gradient was established using a BioComp Model 153 gradient station (BioComp Instruments; Fredericton, NB). Samples were then laid on top of the 10-40% sucrose gradient. Gradients were then centrifuged at 38K rpm for 18 hrs in a Beckman Optima L-90K Ultracentrifuge. Samples were taken from the gradient by pipetting 1 ml from the top of the gradient using a wide-bore pipette tip. Samples were treated with trichloroacetic acid to a final concentration of 10% and incubated for 30 mins. Samples were then centrifuged at 14K rpm in a bench-

top centrifuge for 10 minutes at 4°C, washed with acetone, and centrifuged at 14K rpm for 10 minutes at 4°C. Reducing sample buffer was added to all samples followed by analysis by SDS-PAGE and western blot.

Western Blot

After SDS-PAGE, gels were washed with ultra-pure water and transfer membrane (Immobilon-FL; Millipore) was pre-treated according to the manufacturer's instructions. Transfer was then conducted using a wet-transfer apparatus (Invitrogen Novex mini-cell). Blots were blocked in 5% milk, 2% BSA in PBS solution for 1 hr under gentle shaking at room temperature. Primary antibodies at the indicated concentration in a 5% milk, 2% BSA, PBST (2.7 mM KCl, 1.5 mM KH_2PO_4, 136.9 mM NaCl, 8 mM Na_2PHO_4, 0.5% Tween 20, pH 7.4) solution were added to blots and incubated overnight at 4 °C under gentle rotation. Blots were then washed with PBST and incubated with IRDye 800CW conjugated goat secondary antibody against the appropriate species (LI-COR; Lincoln, NE) in a 5% milk, 2% BSA, 0.02% SDS, PBST solution for 1 hr at room temp under gentle shaking. Blots were then washed with PBST and imaged using a LI-COR Odyssey CLx imaging system.

GST-calreticulin pull down

Recombinant GST-calreticulin was expressed in *Escherichia Coli* and purified as previously described (Baksh & Michalak, 1991; Pearse et al., 2008). MI8-5 cells were seeded into 3 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as

previously described. Cells were either treated with N-butyl deoxynojirimycin (Toronto Research Chemicals) for 30 mins prior to the experiment and then continuously throughout or for 15 mins prior to each time point, as indicated. 80% of the cell lysate was incubated with 8 μg of GST-Calreticulin pre-bound to glutathione beads (GE Life Sciences) overnight at 4°C while 20% of the lysate was immunoprecipitated with Myc antibody, as previously described. All samples were washed with Connie's wash and reducing sample buffer was added to the immunoprecipitations. GST-calreticulin samples were treated with 20 μl of elution buffer (1% SDS in 10 mM Tris, pH 7.5, 150 mM NaCl) at 95°C for 10 min, centrifuged for 2 min at 2000 rpm, the supernatant was collected and quenched with MNT buffer, followed by immunoprecipitation with protein-A sepharose beads with Myc antibody, as previously described. Beads were then washed with Connie's wash and the sample was treated with reducing sample buffer. Sample was then analyzed by SDS-PAGE, imaged with phosphorimaging, and quantified using ImageQuant (Fujifilm).

XBP1 Splicing

Cho cells were seeded onto 10 cm plates and transfected for 24 hr. Untransfected cells were either treated with tunicamycin in DMSO at a concentration of 5 μg/ml or DMSO alone for 24 hr. Total RNA was then collected using an RNeasy kit (Qiagen). cDNA was then generated using AMV Reverse Transcriptase (NEB) and mouse *Xbp1* primers (Lee et al., 2002), according to the manufacturer's instructions. The resulting cDNA product was then separated using a 2.5% agarose gel and visualized using a G:BOX (Syngene).

BIP expression

Cells were seeded onto 10 cm plates and transfected for 24 hr. Untransfected cells were either treated with tunicamycin (5 μg/ml) in DMSO or DMSO alone for 24 hr. Cells were then lysed using MNT and trichloroacetic acid (TCA) precipitated. TCA precipitation was conducted by adding TCA to cell lysate to a final concentration of 10%. Cell lysate was then briefly rotated and allowed to incubate on ice for 15 min before spinning at 14,000 rpm for 10 min at 4 ºC. Supernatant was then aspirated and washed with acetone and centrifuged for 10 min at 4 ºC. Supernatant was then aspirated and the remaining precipitant was allowed to dry for 5 minutes at room temperature and briefly at 65 ºC. Reducing sample buffer was then added and 5% of total lysates were resolved on a 9% reducing SDS-PAGE. Gels were then imaged by western blotting, as previously described, using the following antibodies; anti-BiP (1:1000), Myc (1:2,500), A1AT (1:500), and GapDH (1:1000).

PERK phosphorylation

Cells were seeded onto 10 cm plates and transfected for 24 hr. Untransfected cells were either treated with tunicamycin (5 μg/ml) in DMSO or DMSO alone for 24 hr. Cells were then lysed using MNT and TCA precipitated as previously described. 5% of total lysates were resolved on a 9% reducing SDS-PAGE. Gels were then imaged by western blotting, as previously described, using the following antibodies: PERK (1:500), Myc (1:2,500) and A1AT (1:500).

Co-Immunoprecipitations

Cells were seeded onto 3 cm plates, transfected, and metabolically labeled using [^{35}S]-Cys/Met, as previously described. After 30 mins of pulse and the indicated chase times, cells were lysed and the lysate was collected. For EDEM co-IPs, half of the plates were lysed in MNT. For OS9.2 and XTP3B, half of the plates were lysed in buffer containing 150 mM NaCl, 5 mM EDTA pH 7.4, 50 mM Tris pH 7.4, 1% Triton X-100, 50 μM Calpain inhibitor I, 1μM pepstatin, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 400 μM phenylmethylsulfonyl fluoride, and 20 mM NEM. In all cases, the other half of plates were lysed with 2% CHAPS HBS buffer. Cell lysates were then split equally and immunoprecipitated for either substrate or ERAD factor. ATIII was immunoprecipitated as previously described. AlAT NHK was immunoprecipitated using A1AT antibody and protein-A agarose beads EDEM1/2 were immunoprecipitated using Flag-tag and protein-A agarose beads while EDEM3 was immunoprecipitated using HA-tag and protein-A agarose beads. OS9.2/XTP3B were affinity purified using S-tag agarose beads (Millipore-Sigma). IPs were then washed with Connie's buffer if lysed using 1% Triton X-100, or 0.5% CHAPS HBS buffer if lysed using 2% CHAPS HBS. After washes, sample buffer was added and samples were analyzed by 9% reducing SDS-PAGE and imaged by phosphorimaging.

Statistics

Percentages of reglucosylation were calculated by dividing the number obtained from quantification of bands in the GST-calreticulin lanes by the number obtained from quantification of bands in the single immunoprecipitation lanes multiplied by four to

correct for the amount of cell lysate used in each pull-down. The comparative percentage of reglucosylation was then obtained by dividing the amount of reglucosylation of mutant protein over that of wild type. For all quantifications, error bars were calculated by determining the standard deviation of three independent samples. Statistical significance was determined by using an unpaired *t*-test with a confidence interval of 95%.

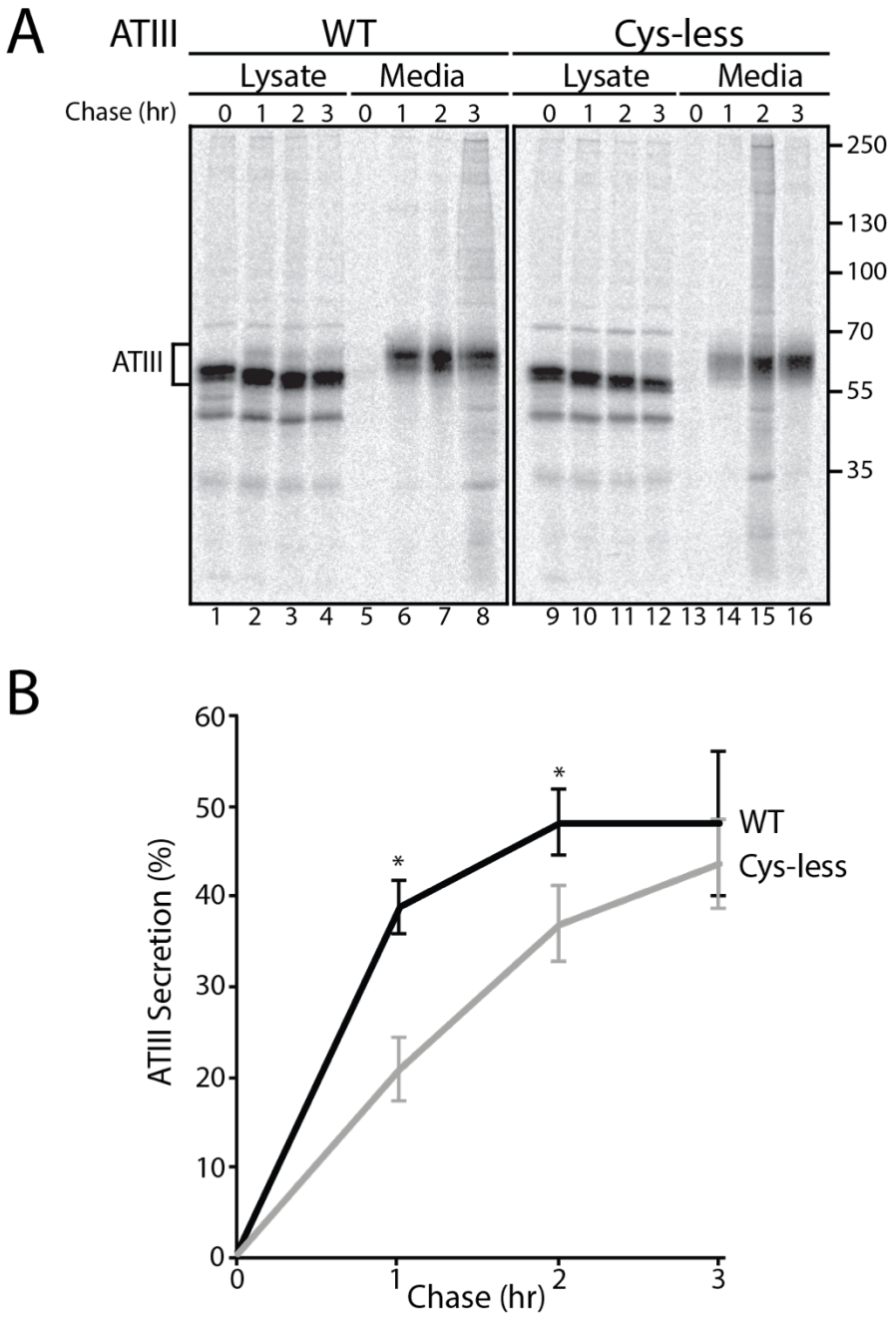

Figure 2.1 ATIII Cys-less is efficiently secreted

(A) ATIII and ATIII Cys-less were expressed in CHO cells. Cells were radiolabeled with [35S]-Cys/Met for 30 min and chased for the indicated times. At each time point, cell lysate and media were collected. Cells were lysed in MNT buffer. ATIII was immunoprecipitated using anti-Myc antibodies. Samples were resolved by reducing 9% SDS-PAGE. (B) Quantification of ATIII secretion from panel A. The lysate and media was quantified and ATIII secretion is presented as a percentage of ATIII in the media to ATIII in the 0 hr lysate.

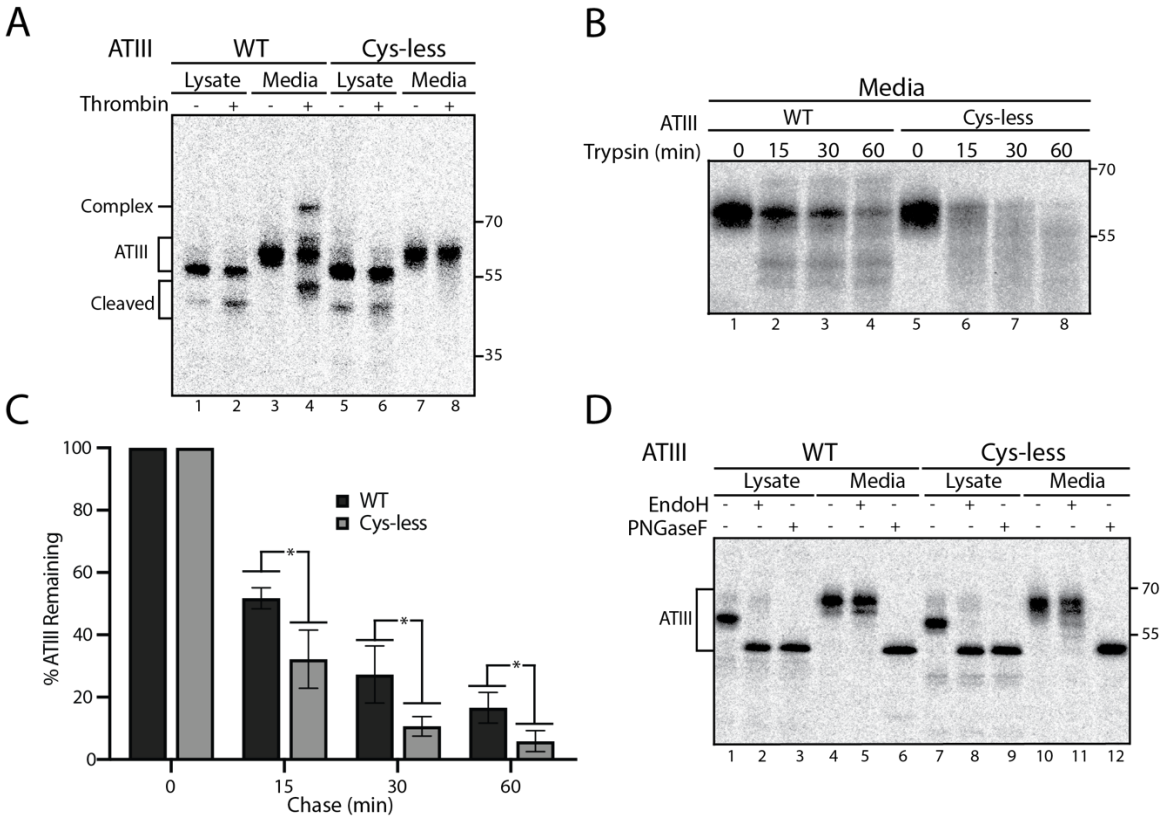

Figure 2.2 Secreted ATIII Cys-less is inactive and misfolded

(A) ATIII WT and ATIII Cys-less were expressed in CHO cells and radiolabeled with [35S]-Cys/Met for 30 min and chased for 3 hr. Cells were lysed in MNT buffer. Cell lysate and media were collected and ATIII was immunoprecipitated using anti-Myc antibodies and washed with buffer containing 0.5% CHAPS. ATIII was then eluted from the immunoprecipitation beads by incubation with 0.5 mg/ml c-Myc peptide for 1 hr at 37 °C. Sample was then evenly split between treated and non-treated samples. 0.648 mg (2 units) of thrombin were added to treated samples while non-treated samples received an equal volume of water. Samples were then incubated at 37 °C for 1 hr and resolved by reducing 9% SDS-PAGE. (B) ATIII WT and Cys-less were expressed in CHO cells. Cells were radiolabeled with [35S]-Cys/Met for 30 min, chased for 3 hr, and cell lysate and media were collected. Cells were lysed in MNT buffer. ATIII was immunoprecipitated using anti-Myc antibodies. 0.05 μg of trypsin was added to treated samples and an equal volume of water was added to non-treated samples. Samples were then incubated for the indicated times at 37 °C and resolved by reducing 9% SDS-PAGE. (C) Quantification of ATIII remaining post-trypsin degradation from panel B. (D) ATIII WT and Cys-less were expressed in CHO cells. Cells were radiolabeled for 30 mins with [35S]-Cys/Met, chased for 2 hr, and cell lysate and media were collected. Cells were lysed in MNT buffer. ATIII was immunoprecipitated using anti-Myc antibodies. Lysate and media samples were treated with EndoH, PNGaseF, or left untreated. Samples were then resolved on a reducing 9% SDS-PAGE. All experiments are representative experiments of three independent experiments.

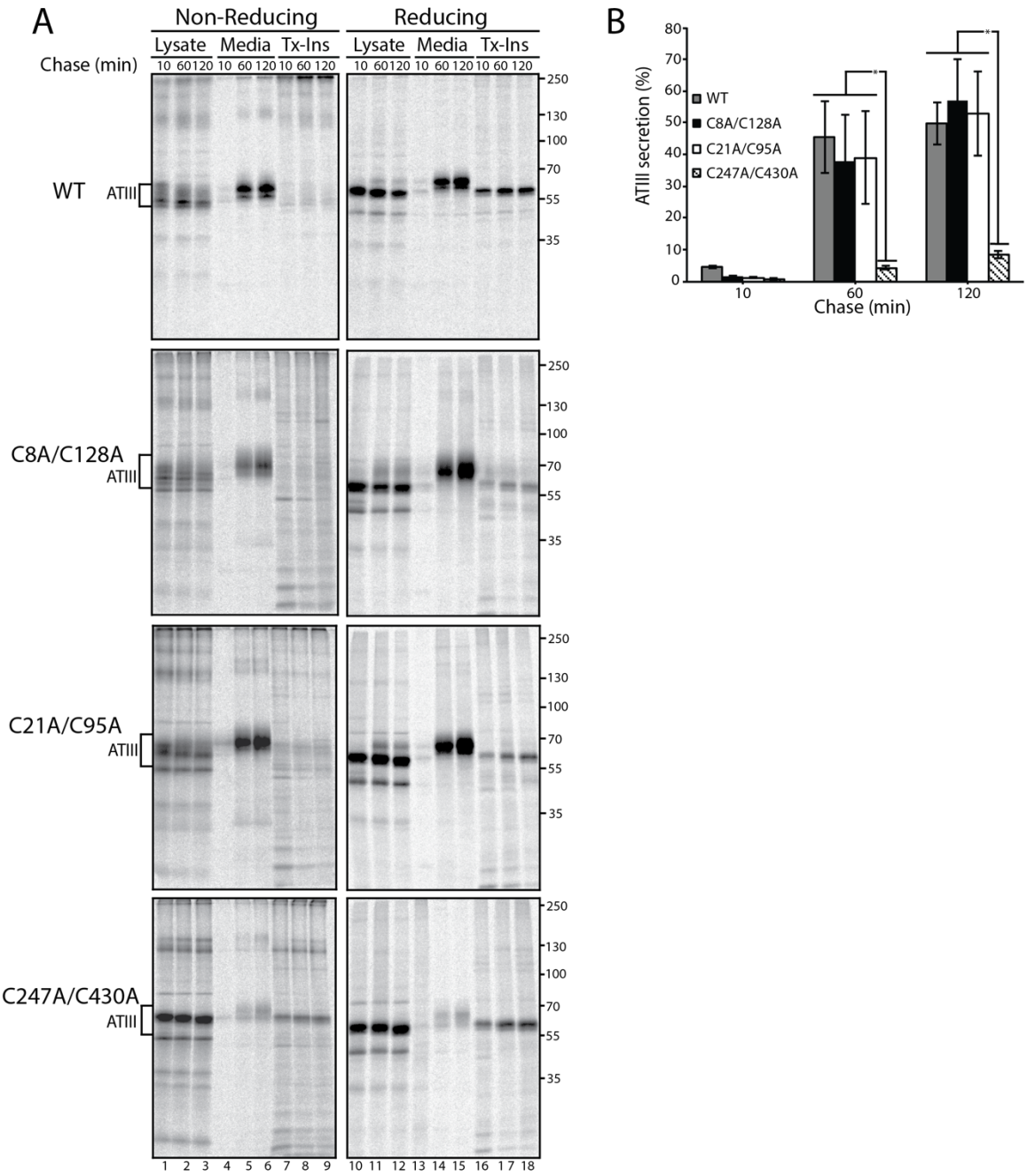

Figure 2.3 ATIII C247A/C430A is retained in the cell while remaining disulfide mutants are secreted

(A) ATIII and ATIII disulfide mutants were expressed in CHO cells. Cells were radiolabeled with [35S]-Cys/Met for 30 min and chased for the indicated times. At each time point, cell lysate and media were collected. Cells were lysed in MNT buffer. ATIII was immunoprecipitated using anti-Myc antibodies. Samples were resolved by both reducing and non-reducing 9%

SDS-PAGE. (B) Quantification of ATIII secretion from panel A. The lysate and media were quantified and ATIII secretion is presented as a percentage of ATIII in the media to ATIII in the 10 min lysate.

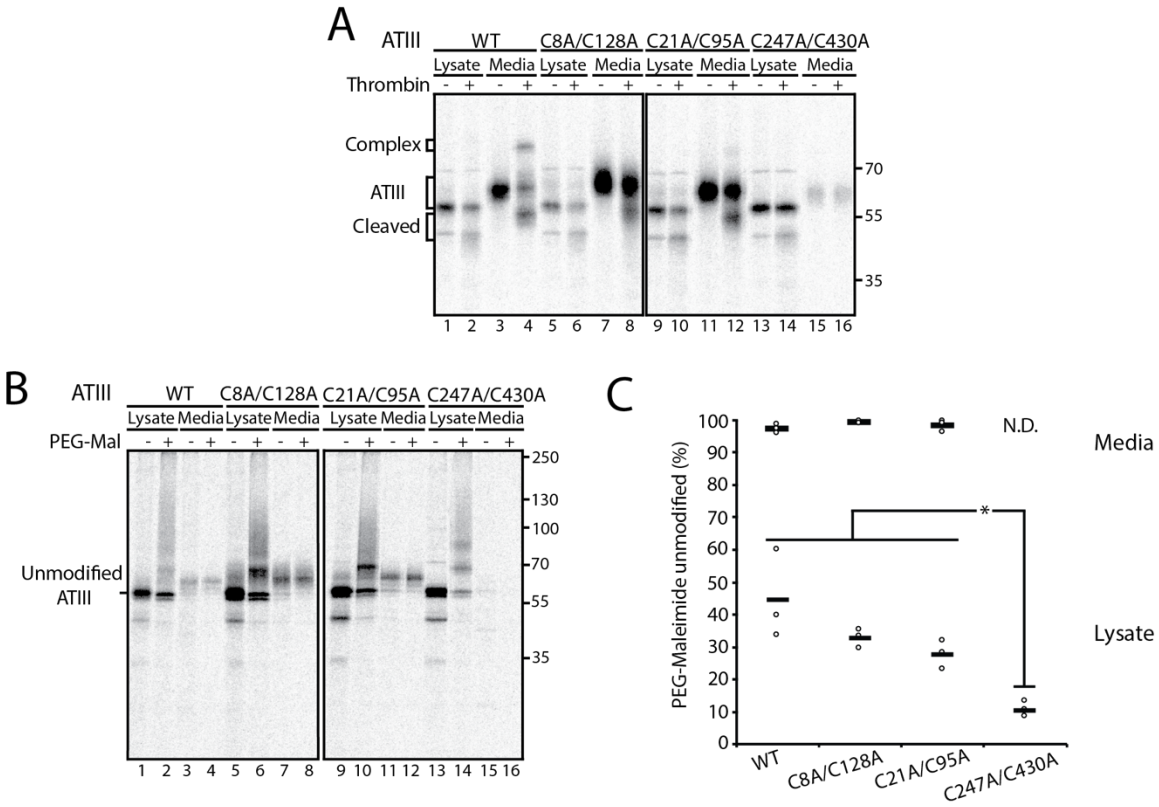

Figure 2.4 Cellularly retained ATIII C247/430A is inactive and contains a higher level of free thiols than secreted disulfide mutants

(A) ATIII and ATIII disulfide mutants were expressed in CHO cells, radiolabeled with [35S]-Cys/Met for 30 min and chased for 3 hr. Cells were lysed in MNT buffer. Cell lysate and media were collected and ATIII was immunoprecipitated using anti-Myc antibody and washed with 0.5% CHAPS buffer. ATIII was then eluted from the immunoprecipitation beads by incubation with 0.5 mg/ml c-Myc peptide for 1 hr at 37 °C. Samples were then evenly split between treated and non-treated samples. 0.648 mg (2 units) of thrombin were added to treated samples while non-treated samples received an equal volume of water. Samples were then incubated at 37 °C for 1 hr and resolved by reducing 9% SDS-PAGE. (B) ATIII and ATIII disulfide mutants were expressed in CHO cells, radiolabeled with [35S]-Cys/Met for 30 min and chased for 30 min. Cells were lysed in MNT buffer. To treated samples, PEG-maleimide was added to a final concentration of 1.4 mM and to the non-treated samples N-ethyl maleimide was added to a final concentration of 5 mM. Samples were incubated at room temperature for 30 min. Treated samples were then quenched by adding dithiothreitol to a final concentration of 100 mM. ATIII was immunoprecipitated using anti-Myc antibodies and samples were resolved by reducing 9% SDS-PAGE. (C) Quantification of PEG-Maleimide unmodified ATIII from panel B. Percent PEG-Maleimide unmodified was determined by quantifying the amount of unmodified ATIII in PEG-Maleimide treated and non-treated samples and dividing the treated sample by the untreated sample. All experiments are representative of three independent experiments.

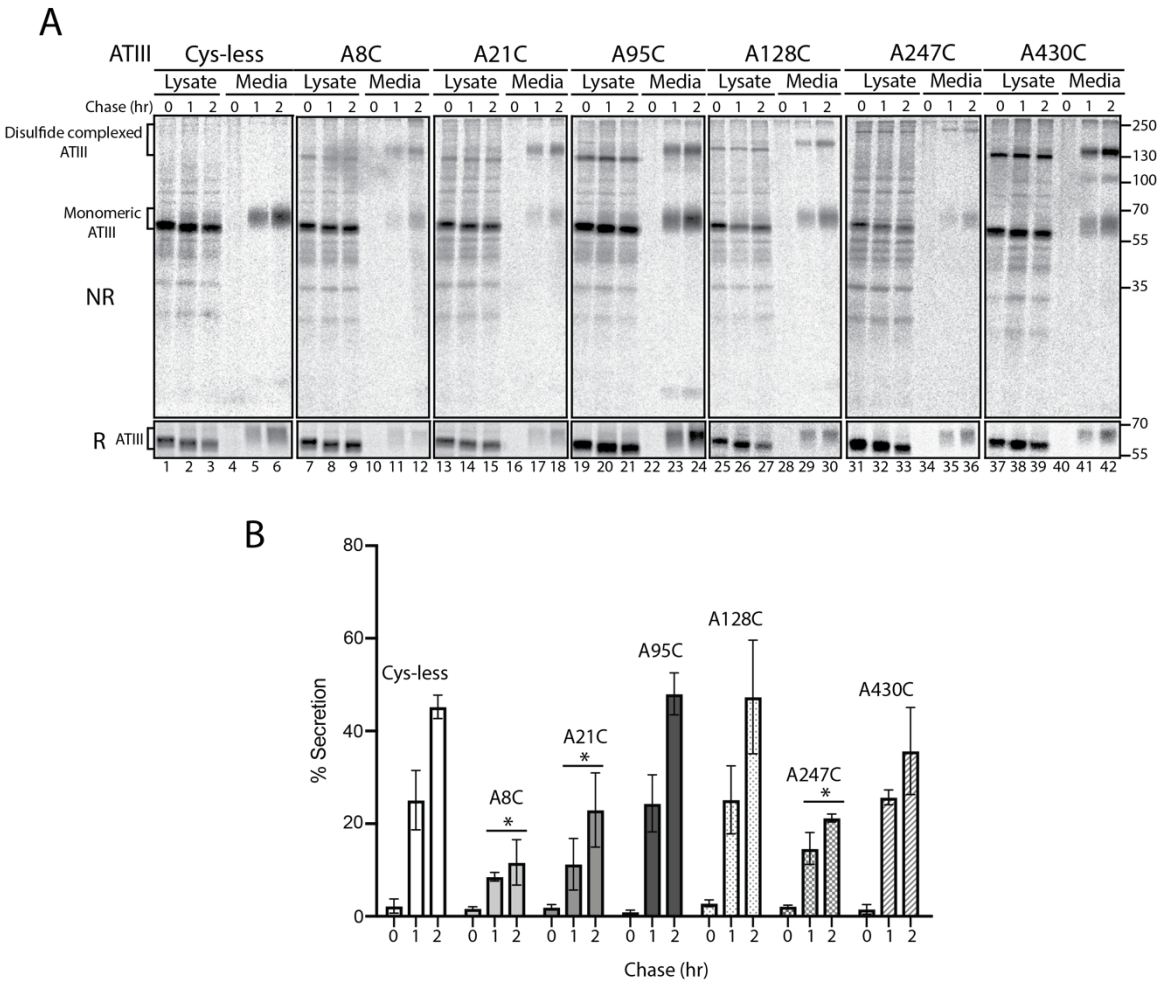

Figure 2.5 ATIII with a single free thiol is retained in the ER

(A) ATIII Cys-less and ATIII single Cys mutants were expressed in CHO cells. Cells were radiolabeled with [35S]-Cys/Met for 30 mins and chased for the indicated times. At each time point, cell lysate and media were collected and processed as previously described in Figure 2.1A. Samples were resolved by both reducing (R) and non-reducing (NR) 9% SDS-PAGE. (B) Quantification from panel A. Asterisks denote statistical significance relative to Cys-less. All experiments are representative of three individual experiments.

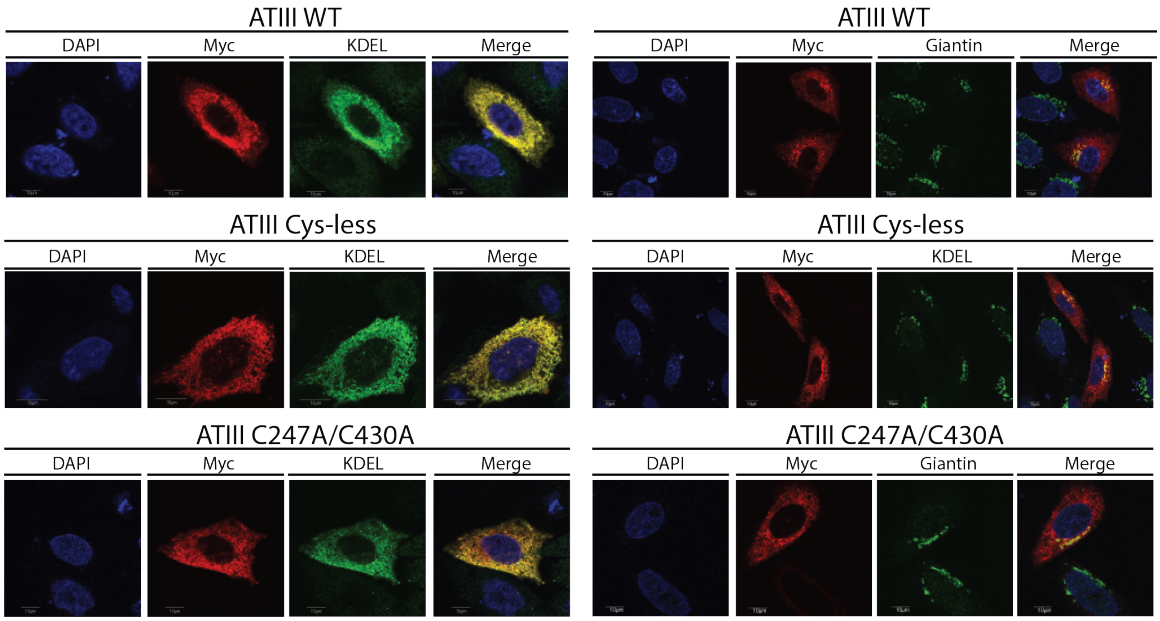

Figure 2.6 ATIII C247A/C430A is retained and diffuse throughout the ER

ATIII WT, ATIII Cys-less and ATIII C247A/C430A were expressed in CHO cells. Cells were fixed in buffer containing 3.7% formaldehyde, permeabilized in buffer containing 0.1% Triton X-100 and stained using anti-Myc antibodies, KDEL, and giantin primary antibodies, as indicated, goat anti-mouse IgG secondary antibody, and DAPI. Cells were imaged using a confocal epifluoresence microscope at 100x oil immersion.

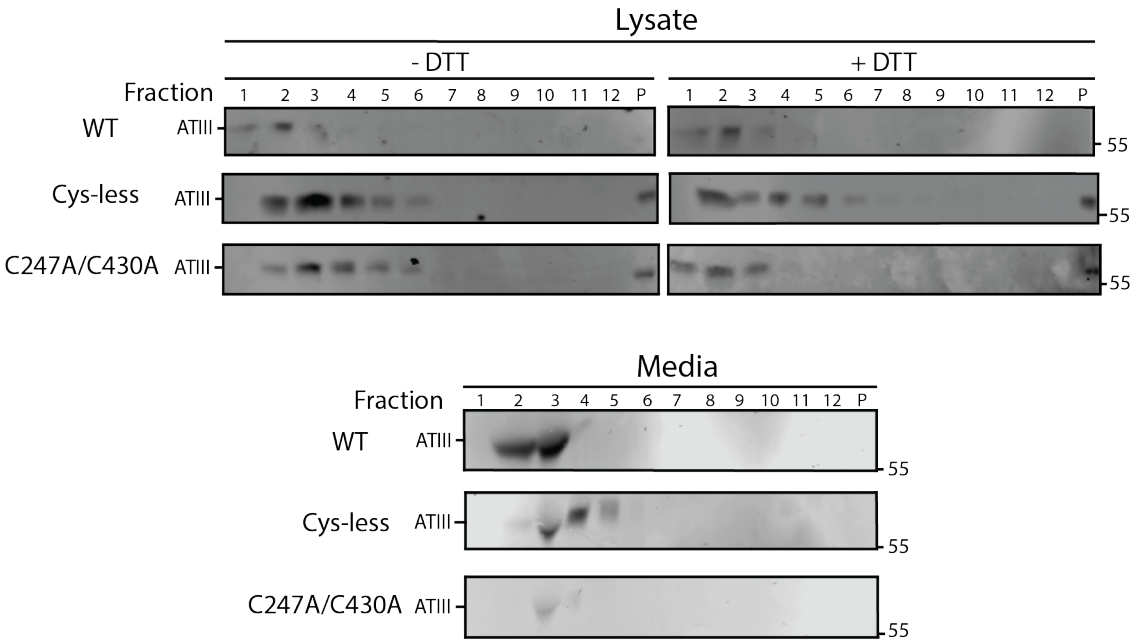

Figure 2.7 ATIII C247A/C430A is retained in a disulfide-dependent multimer

ATIII WT, ATIII Cys-less, and ATIII C247A/C430A were expressed in CHO cells. Cells were lysed in MNT. Media and lysate fractions were collected and split equally between DTT-treated and non-treated fractions. Sucrose gradients were generated by solubilizing sucrose into MNT and a 10-40% gradient was established. Samples were then laid on top of the sucrose gradient. Gradients were then centrifuged at 38K rpm for 18 hr. Samples were taken from the gradient by pipetting 1 ml from the top of the gradient. Samples were TCA precipitated and then resolved on a 9% reducing SDS-PAGE and imaged by western blot with Myc-tag antibody.

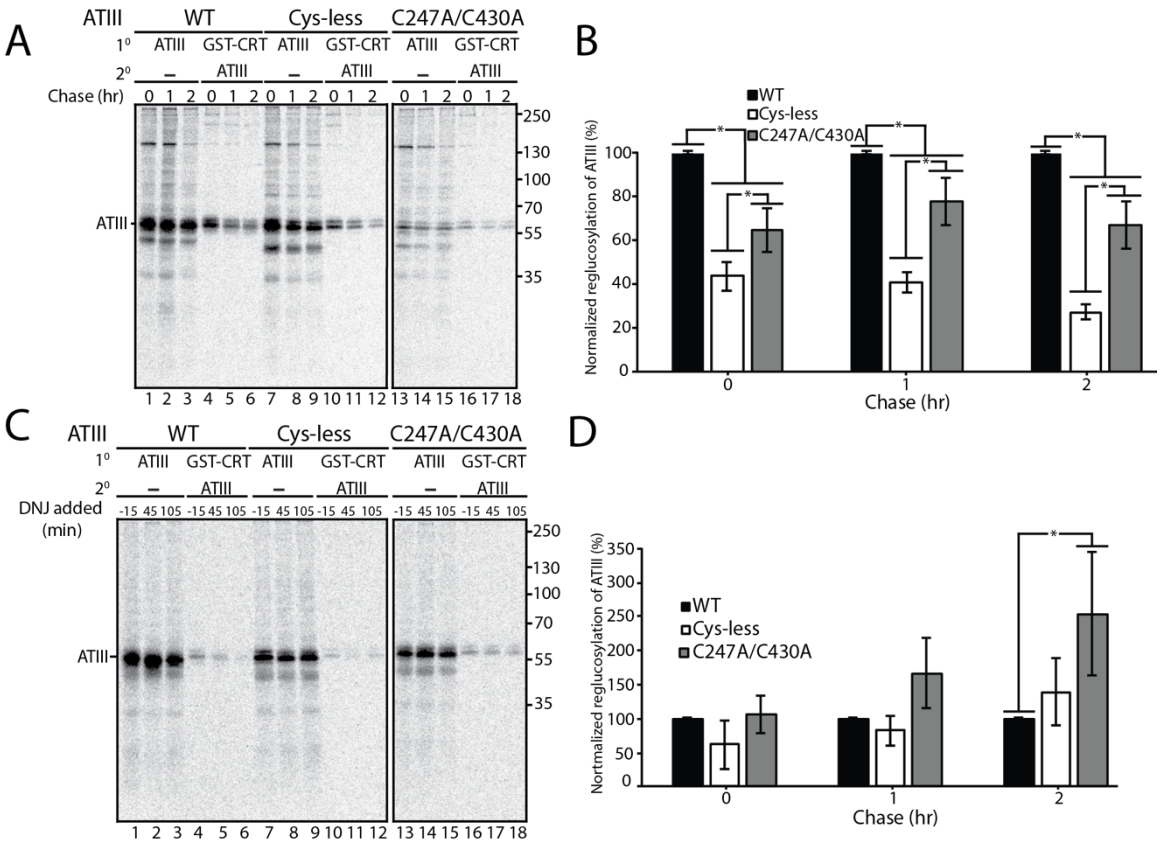

Fig 2.8 ATIII C247A/C430A is not a strong UGGT1 substrate

(A) ATIII variants were expressed in MI8-5 CHO cells. 30 minutes prior to the pulse and throughout the chase, cells were treated with 0.5 mM N-butyl deoxynojirimycin (DNJ). Cells were radiolabeled with [35S]-Cys/Met for 30 mins and chased for the indicated times. At each time point, cells were lysed in MNT buffer. 80% of the cell lysate was affinity purified with Glutathione S-transferase-tagged calreticulin (GST-CRT) while 20% of the cell lysate was immunopurified with anti-Myc antibody. GST-CRT affinity purifications were then eluted in buffer containing 1% SDS, diluted in MNT, and immunopurified using anti-Myc-Tag antibody. Samples were resolved by reducing 9% SDS-PAGE. (B) Quantification of reglucosylation in panel A. Percent reglucosylation for each ATIII variant is calculated by quantifying the bands corresponding to ATIII, multiplying the lysate band by 4, and dividing the amount of ATIII in the sequential IP by the amount of ATIII in the non-sequential IP at each time point. All reglucosylation values are normalized to WT ATIII. (C) Same as panel A, except DNJ was added 15 mins prior to each time point, not throughout the experiment. (D) Quantification of reglucosylation in panel C. Percent reglucosylation was calculated as described in panel B.

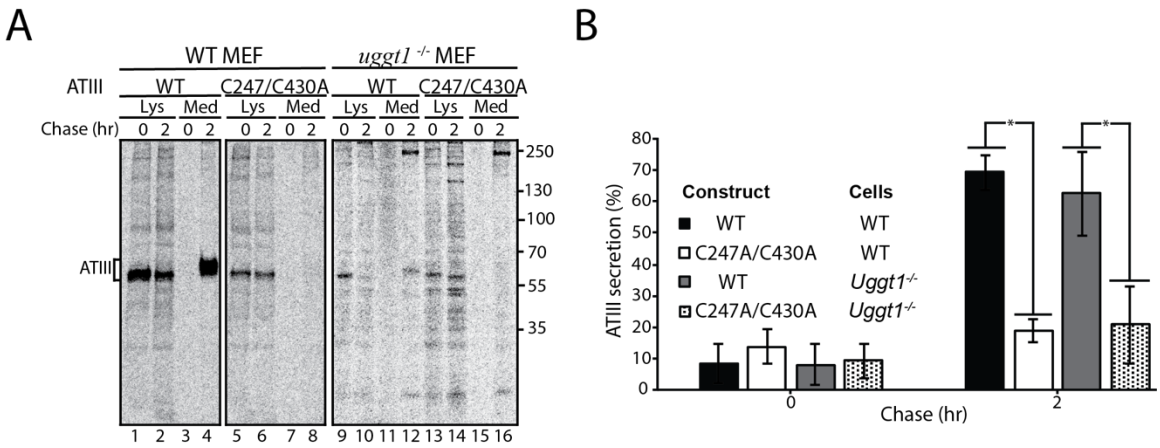

Figure 2.9 UGGT1 is not required for ATIII C247A/C430A to be retained

(A) ATIII and ATIII C247A/C430A is expressed in both WT and UGGT1-/- MEF cells. Cells were radiolabeled with [35S]-Cys/Met for 30 mins and chased for the indicated times. At each time point, cells were lysed in MNT buffer. At each time point, cell lysate and media were collected. Cells were lysed in MNT buffer. ATIII was immunoprecipitated using anti-Myc antibody. Samples were resolved by both reducing and non-reducing 9% SDS-PAGE. (B) Quantification of ATIII secretion from panel A. The lysate and media were quantified and ATIII secretion is presented as a percentage of ATIII in the media to ATIII in the 0 hr lysate. All experiments are representative of three independent experiments.

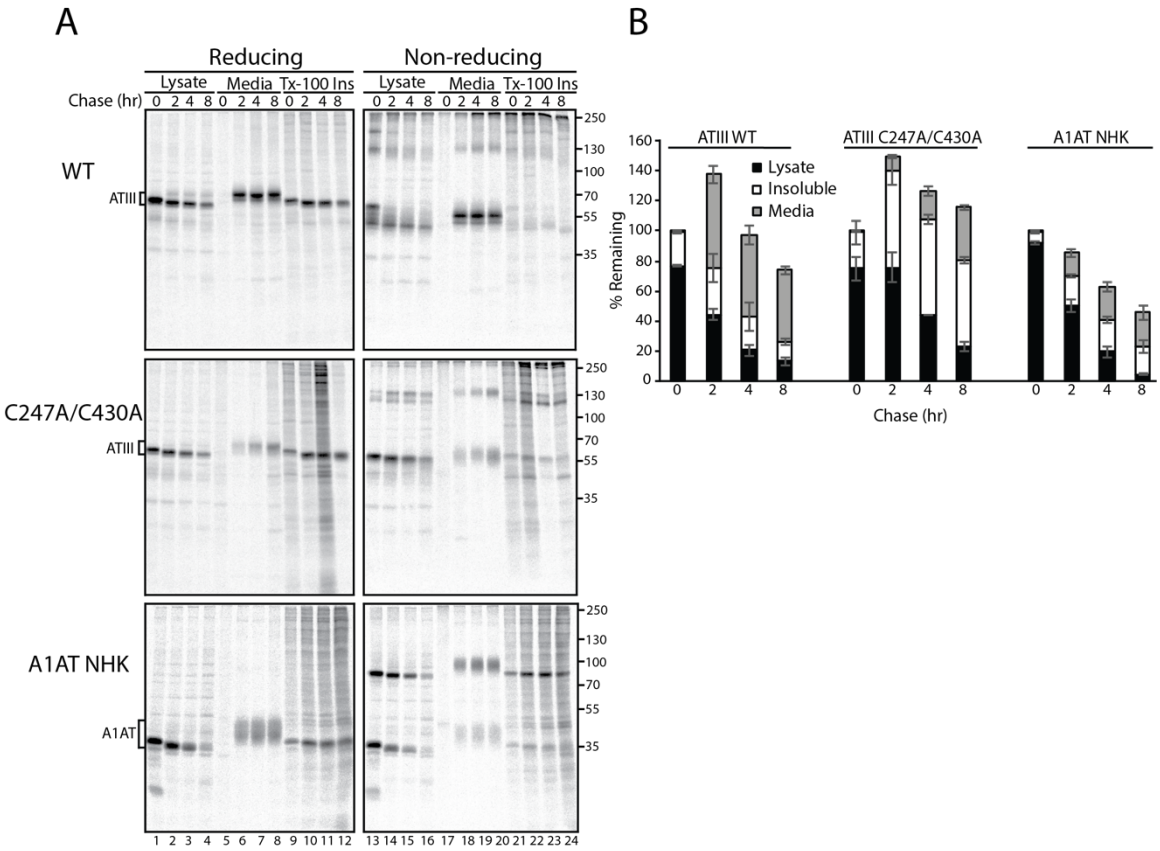

Figure 2.10 ATIII C247A/C430A is poorly degraded

(A) ATIII WT and disulfide mutants were expressed in CHO cells. Cells were radiolabeled with [35S]-Cys/Met for 30 min and chased for the indicated times. At each time point, cell lysate, media, and Triton X-100 insoluble fractions were collected and processed as described in Figure 2.1A. (B) Quantification of A. The percent of each protein remaining was calculated by quantifying ATIII present in each fraction of the reducing gel and dividing by the amount of ATIII immediately after the chase (0 hr). The amount of protein in the indicated fraction at each time point is represented as a fraction of the total amount of protein present at that time point. All experiments are representative of three independent experiments.

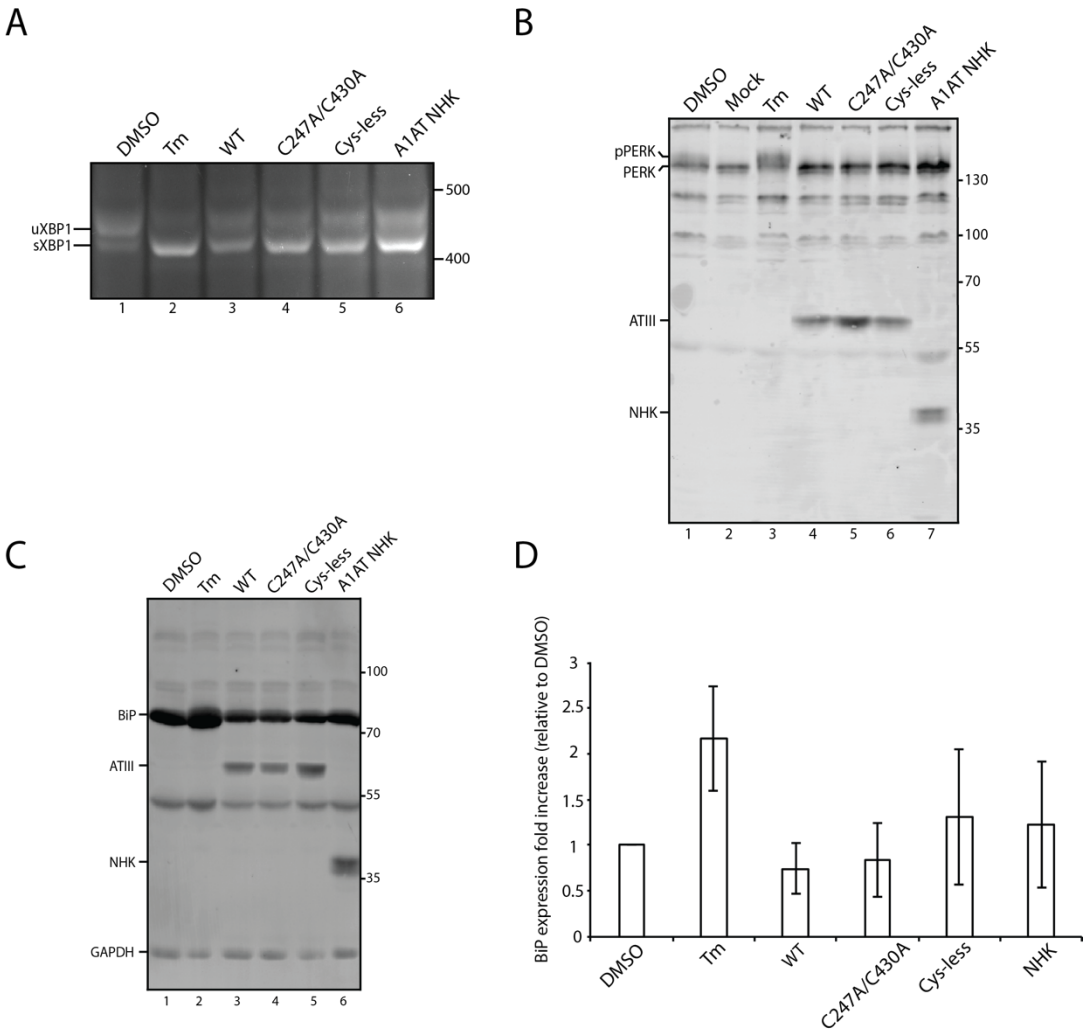

Figure 2.11 IRE-1 but not ATF6 or PERK is activated by ATIII overexpression

(A) ATIII variants and A1AT NHK were expressed in CHO cells for 24 hr. Control cells were treated with DMSO or tunicamycin (Tm) for 24 hr. Cells were lysed and RNA was collected. cDNA was generated and amplified via PCR using XPB-1 specific primers. (B) ATIII variants and A1AT NHK were expressed in CHO cells for 24 hr. Control cells were treated with either DMSO or Tm for 24 hr. Cells were then lysed with MNT and TCA precipitated. 5% whole cell lysate was then resolved on a 9% SDS-PAGE and imaged by western blot using anti-Myc antibody (ATIII), anti-Perk antibody, and anti-A1AT antibody (C) ATIII variants and A1AT NHK were expressed in CHO cells for 24 hr. Control cells were treated with either DMSO or Tm for 24 hr. Cells were then lysed with MNT and TCA precipitated. 5% whole cell lysate was then resolved on a 9% SDS-PAGE and imaged by western blot using the indicated antibodies. (D) Quantification of panel B. Relative BiP expression was calculated by normalizing all BiP levels to DMSO, using GAPDH as a loading control. All experiments are representative of three independent experiments.

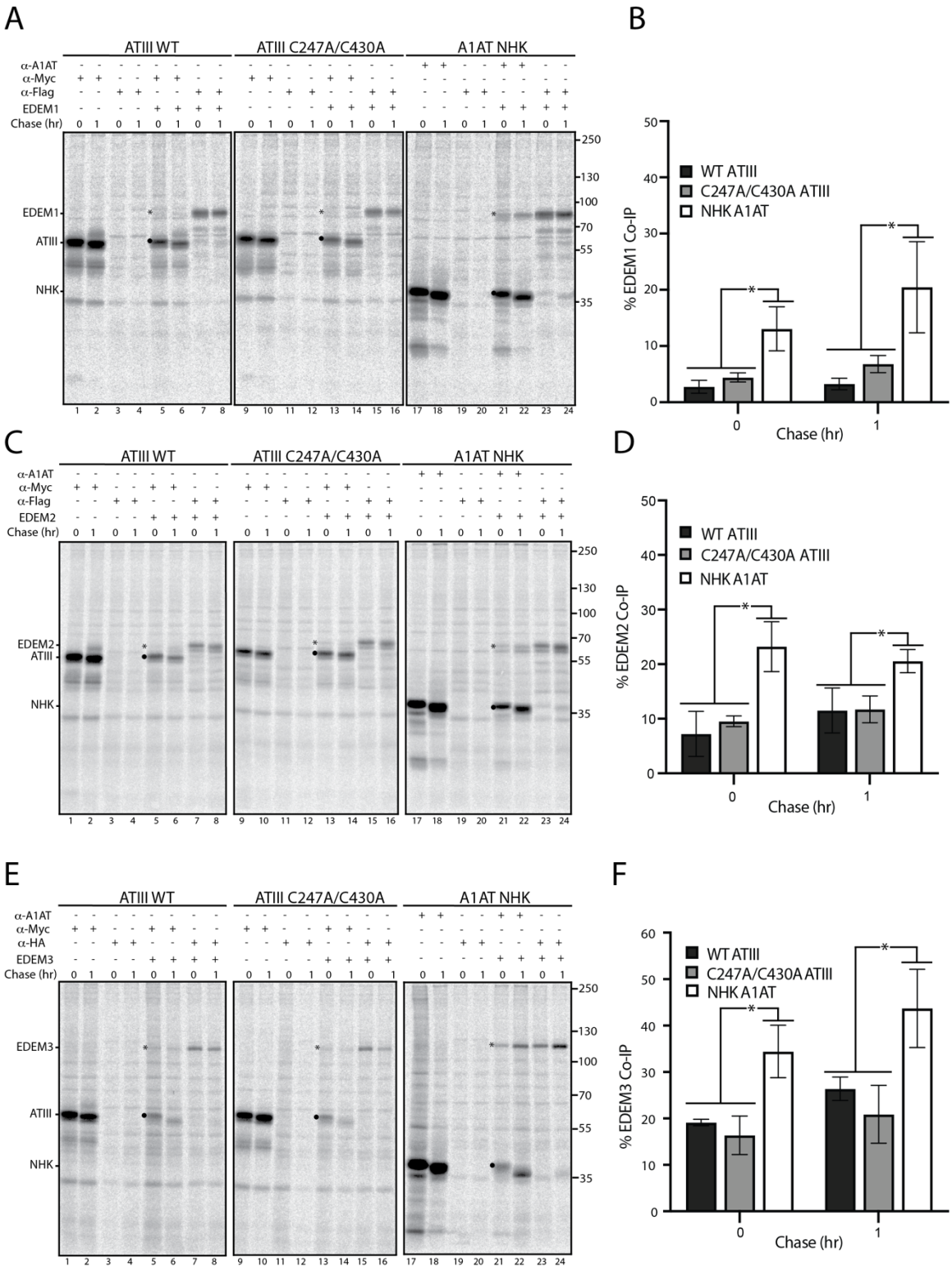

Figure 2.12 ATIII C247A/C430A binds poorly to EDEM1, EDEM2, and EDEM3

(A) Flag-tagged EDEM1 was co-expressed with Myc-tagged ATIII WT,

Myc-tagged ATIII C247A/C430A, and A1AT NHK in CHO cells, as indicated. Cells were radiolabeled with [35S]-Cys/Met for 30 min and chased for the indicated times. Cells were lysed in MNT buffer. Lysates were then split equally and immunoprecipitated with either anti-Myc, Flag, or A1AT antibodies, as indicated. Samples were resolved by reducing 9% SDS-PAGE and imaged by phosphorimaging. EDEM1 is denoted by an asterisk, while ATIII and A1AT NHK are denoted by a filled circle. (B) Quantification of EDEM1 co-immunoprecipitation from panel A. Percent co-immunoprecipitation was determined by dividing the amount of EDEM1 immunoprecipitated by ATIII or NHK to total EDEM1. (C-F) Either Flag-tagged EDEM2 or HA-tagged EDEM3 were co-expressed with Myc-tagged ATIII WT, Myc-tagged ATIII C247A/C430A, and A1AT NHK in CHO cells. Lysates were treated the same as previous panels and immunoprecipitated using the indicated antibodies. Quantifications were conducted as described in panel B. All experiments are representative of three independent experiments.

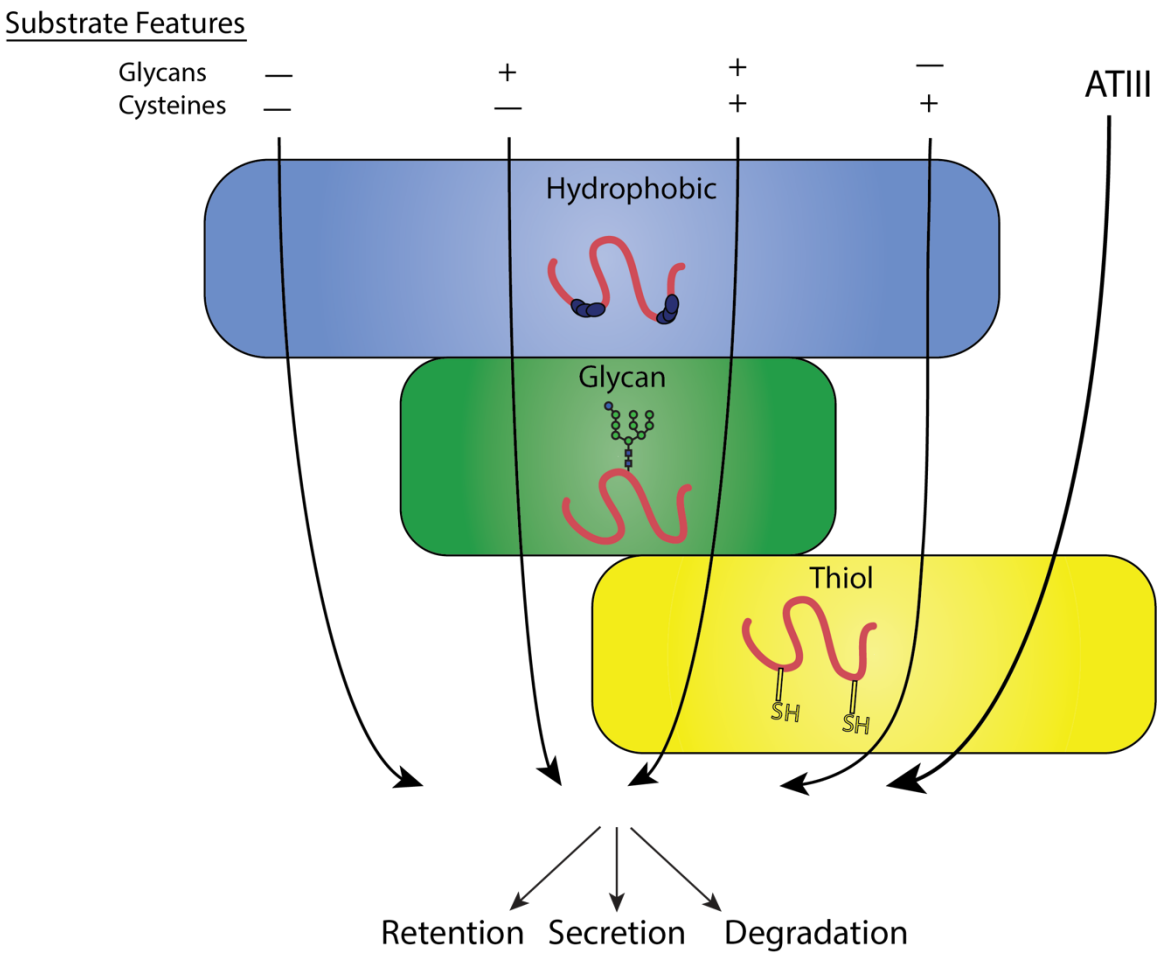

Figure 2.13 ATIII quality control model

Substrate features are generally understood to dictate the ER quality control pathways substrates engage, as depicted by arrows corresponding to different combinations of substrate features traversing quality control pathways. ATIII does not follow expected quality control pathways, but rather engages only the thiol-dependent quality control.

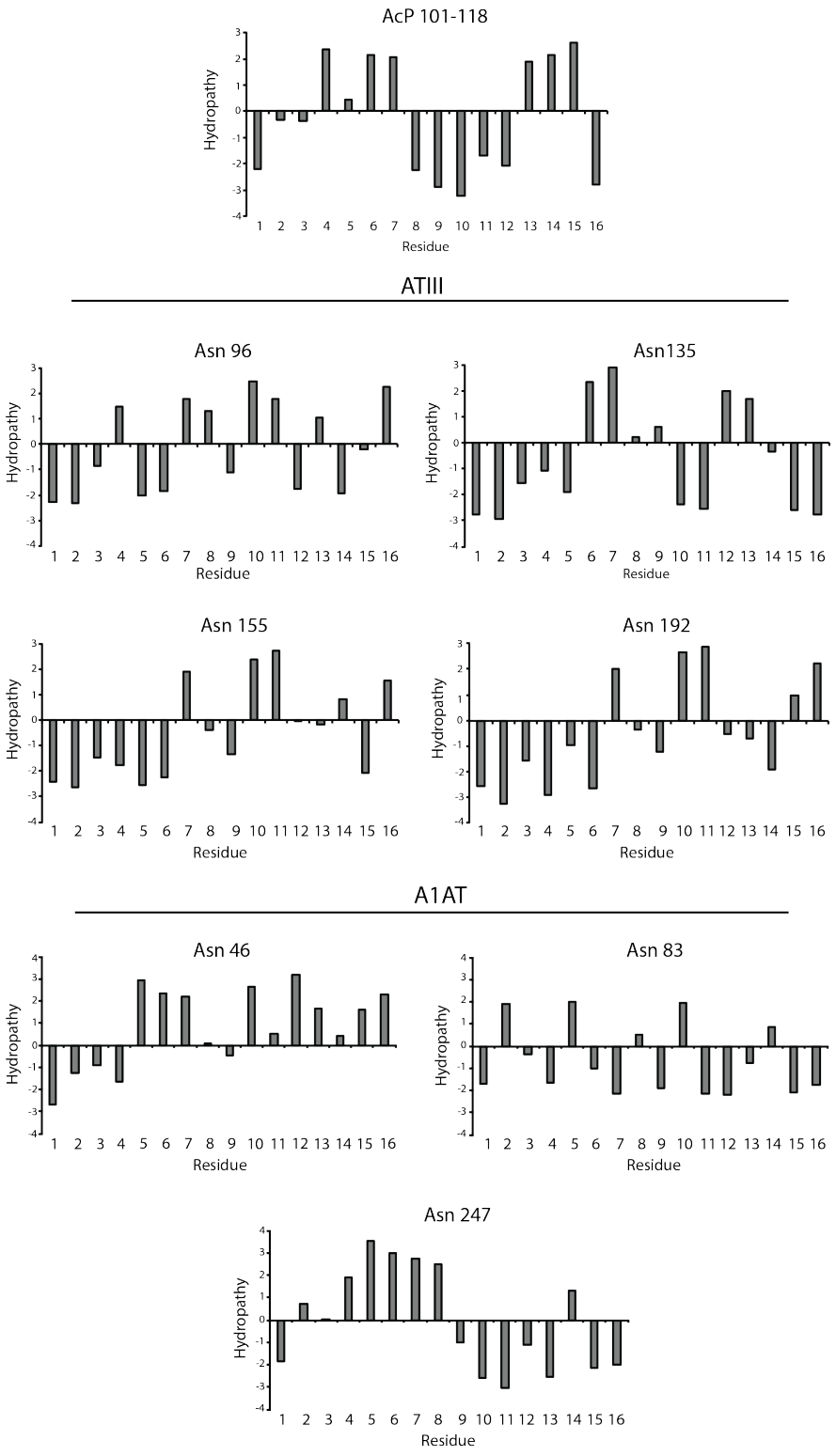

Figure 2.14 Hydropathy plots of acid phosphatase (AcP 101-118), ATIII and A1AT glycan regions

Kyte-Doolittle scores of amino acids C-terminal to the glycan using a window size of 5 amino acids are depicted. The asparagine residue for each glycan is positioned at 0. Positive values represent hydrophobicity. Oscillating hydrophobicity profiles with two hydrophobic patches of three or more amino acids are correlated with increased reglucosylation by UGGT1. AcP 101-118 is depicted as an example of a well-reglucosylated substrate possessing a characteristic hydropathy profile (Taylor et al, 2003). Note that ATIII does not contain any such hydropathy profiles while A1AT possesses one.

CHAPTER 3

Cellular substrate selectivity of the endoplasmic reticulum protein quality control sensors UGGT1 and UGGT2

Benjamin M. Adams[1,2], Nathan P. Canniff[1,2], Kevin P. Guay[1,2], Ida Signe Bohse Larsen[3,4], and Daniel N. Hebert[1,2]

[1]Department of Biochemistry and Molecular Biology
[2]Program in Molecular and Cellular Biology; University of Massachusetts, Amherst, MA, 01003
[3]Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark.
[4]Copenhagen Center for Glycomics, University of Copenhagen, Copenhagen, Denmark.

This work is currently in submission for publication

Abstract

The protein quality control sensors UDP-glucose: glycoprotein glucosyltransferase (UGGT) 1 and 2 are proposed to act as gatekeepers of the early secretory pathway. They initiate rebinding to the carbohydrate-dependent chaperones calnexin and calreticulin that associate with proteins possessing monoglucosylated glycans. The UGGTs control glycoprotein exit from the endoplasmic reticulum (ER) for trafficking to the Golgi or ER retention to provide additional folding opportunities. A quantitative glycoproteomics strategy was used to identify cellular glycoproteins modified by the UGGTs at endogenous levels and delineate the specificities of UGGT1 and UGGT2. UGGT substrates were comprised of seventy-one mainly large multidomain and heavily glycosylated proteins when compared to the general N-glycome. UGGT1 was the dominant glucosyltransferase with a preference towards large plasma membrane proteins whereas UGGT2 favored the modification of smaller, soluble lysosomal

proteins. This study provides insight into the cellular secretory load that utilizes multiple rounds of carbohydrate-dependent chaperone intervention for proper maturation.

Introduction

Protein folding in the cell is an error-prone process and protein misfolding is the basis for a large number of disease states (Hebert & Molinari, 2007; Hartl, 2017). A significant fraction of the proteome in mammalian cells passes through the secretory pathway by first being targeted to the endoplasmic reticulum (ER) where folding occurs (Uhlén et al., 2015; Itzhak et al., 2016; Adams, Oster, et al., 2019). Molecular chaperones of the ER help to guide secretory pathway cargo along a productive folding pathway by directing the trajectory of the folding reaction, inhibiting non-productive side reactions such as aggregation or by retaining immature or misfolded proteins in the ER until they can properly fold or be targeted for degradation. Understanding how chaperone binding controls the maturation and flux of proteins through the secretory pathway is of important fundamental biological concern and will impact our knowledge of protein folding diseases and the development of potential therapeutics including the production of biologics that are frequently secretory proteins.

Proteins that traverse the secretory pathway are commonly modified with N-linked glycans as they enter the ER lumen (Zielinska et al., 2010). These carbohydrates serve a variety of roles including acting as quality control tags or attachment sites for the lectin ER chaperones calnexin and calreticulin (Helenius & Aebi, 2004; Hebert et al., 2014). N-glycosylation commences co-translationally in mammals and the first round of binding to calnexin and calreticulin is initiated shortly thereafter by the rapid trimming of

glucoses by glucosidases I and II to reach their monoglucosylated state (W. Chen et al., 1995; Cherepanova et al., 2019). Lectin chaperone binding is multifunctional as it has been shown to: (1) direct the folding trajectory of a protein by acting as a holdase that slows folding in a region-specific manner; (2) act as an adapter or platform to recruit folding factors including oxidoreductases (ERp57 and ERp29) and a peptidyl-prolyl *cis trans* isomerase (CypB) to maturing nascent chains; (3) diminish aggregation; (4) retain immature, misfolded or unassembled proteins in the ER; and (5) target aberrant proteins for degradation (Rajagopalan et al., 1994; Hebert et al., 1996; Daniels et al., 2003; Molinari et al., 2003; Oda et al., 2003; N. Wang et al., 2008; Kozlov & Gehring, 2020). For glycoproteins, the lectin chaperones appear to be the dominant chaperone system as once an N-glycan is added to a region on a protein, it has been shown to be rapidly passed from the ER Hsp70 chaperone BiP to the lectin chaperones, further underscoring their central role in controlling protein homeostasis in the secretory pathway (Helenius & Hammond, 1994).

N-glycan trimming to an unglucosylated glycoform by glucosidase II supports substrate release from the lectin chaperones. At this stage, if the protein folds properly, it is packaged into COPII vesicles for anterograde trafficking (C. Barlowe & Helenius, 2016). Alternatively, substrates that are evaluated to be non-native are directed for rebinding to the lectin chaperones by the protein folding sensor UDP-glucose: glycoprotein glucosyltransferase 1 (UGGT1) that reglucosylates immature or misfolded proteins (Helenius, 1994; Sousa & Parodi, 1995b). Since UGGT1 directs the actions of this versatile lectin chaperone system and thereby controls protein trafficking through the

ER, it acts as a key gatekeeper of the early secretory pathway. Therefore, it is vital to understand the activity of UGGT1 and the scope of substrates it modifies.

Our current knowledge of the activity of UGGT1 relies largely on studies using purified components. UGGT1 was found to recognize non-native or near-native glycoproteins with exposed hydrophobic regions using *in vitro* approaches where the modification of glycopeptides, engineered or model substrates by purified UGGT1 was monitored (Ritter & Helenius, 2000; Taylor et al., 2003; Caramelo et al., 2004). Recent crystal structures of fungal UGGT1 have shown that it possesses a central, hydrophobic cavity in its protein sensing domain, which may support hydrophobic-based interactions for substrate selection (Roversi et al., 2017; Satoh et al., 2017).

Cell-based studies of UGGT1 have relied on the overexpression of cellular and viral proteins (Soldà et al., 2007; Pearse et al., 2008; Ferris et al., 2013; Tannous et al., 2015). *Uggt1* knockout studies have found that the roles of UGGT1 appear to be substrate specific as UGGT1 can promote, decrease or not affect the interaction between substrates and calnexin (Soldà et al., 2007). Prosaposin, the only known cellular substrate of UGGT1 when expressed at endogenous levels, grossly misfolds in the absence of *Uggt1* and accumulates in aggresome-like structures (Pearse et al., 2010). Work in animals has further emphasized the importance of UGGT1 as the deletion of *Uggt1* in mice is embryonically lethal (Molinari et al., 2005).

UGGT1 has a paralogue, UGGT2, but it has no demonstrated cellular activity (Arnold et al., 2000). Domain swapping experiments have demonstrated that UGGT2 possesses a catalytically active glucosyltransferase domain when appended to the folding sensor domain of UGGT1 (Arnold & Kaufman, 2003). *In vitro* experiments using

purified, chemically glycosylated interleukin-8 (IL-8), which is not glycosylated in cells, have found that UGGT2 can glucosylate IL-8 (Y. Takeda et al., 2014). This suggests that UGGT2 may be an additional reglucosylation enzyme or protein folding sensor of the ER.

Unlike the classical ATP-dependent chaperones that directly query the conformation of their substrates (Balchin et al., 2016), binding to the lectin chaperones is dictated by enzymes that covalently modify the substrate (Helenius & Aebi, 2004; Hebert et al., 2014). Rebinding to the carbohydrate-dependent chaperones is initiated by the UGGTs that interrogate the integrity of the structure of the protein. Therefore, the proteome-wide detection of cellular UGGT substrates provides the unprecedented opportunity to identify clients that require multiple rounds of chaperone binding and are more reliant on lectin chaperone binding for proper maturation and sorting. Therefore, we designed a cell-based quantitative glycoproteomics approach to identify high-confidence endogenous substrates of UGGT1 and UGGT2 by the affinity purification of monoglucosylated substrates in CRISPR/Cas9-edited cells. UGGT1 and UGGT2 substrates were found to display multiple features of complex proteins including extended lengths plus large numbers of Cys residues and N-glycans. Specific substrates of either UGGT1 or UGGT2 were also discovered, therefore determining that UGGT2 possessed glucosyltransferase activity and identifying its first natural substrates. UGGT1 demonstrated a slight preference for transmembrane proteins, especially those targeted to the plasma membrane, while UGGT2 modification favored soluble lysosomal proteins. The identification of reglucosylated substrates improves our understanding of their

folding and maturation pathways and has implications regarding how folding trajectories may be altered in disease states.

Results

Experimental design

To identify the substrates that are most dependent upon persistent calnexin/calreticulin cycle binding, we isolated and identified endogenous substrates for the ER protein folding sensors UGGT1 and UGGT2. As the product of a reglucosylation by the UGGTs is a monoglucosylated N-glycan, the presence of the monoglucosylated glycoform was used as a readout for substrate reglucosylation. N-glycans are originally transferred to nascent glycoproteins containing three glucoses, therefore a monoglucosylated glycan can be generated either through trimming of two glucoses from the nascent N-linked glycan or through reglucosylation by the UGGTs. In order to isolate the reglucosylation step from the trimming process, a gene edited cell line was created that transfers abbreviated unglucosylated N-linked glycans to nascent chains. The N-linked glycosylation pathway in mammalian cells is initiated through the sequential addition of monosaccharides, mediated by the *ALG* (Asn-linked glycosylation) gene products, to a cytosolically exposed dolichol-P-phosphate embedded in the ER membrane (Aebi, 2013; Cherepanova et al., 2016) (Figure 3.1A). The immature dolichol-P-phosphate precursor is then flipped into the ER lumen and sequential carbohydrate addition is continued by additional ALG proteins. The completed N-glycan ($Glc_3Man_9GlcNAc_2$) is then appended to an acceptor Asn residue in the sequon Asn-Xxx-Ser/Thr/Cys (where Xxx is not a Pro) by the oligosaccharyl transferase (OST) complex

(Cherepanova et al., 2016). Initially, a Chinese Hamster Ovary (CHO) cell line with a defect in *Alg6* was employed to establish the utility of this approach to follow (re)glucosylation (Quellhorst et al., 1999; Cacan et al., 2001; Pearse et al., 2008, 2010; Tannous et al., 2015). As the CHO proteome is poorly curated compared to the human proteome, CRISPR/Cas9 was used to knock-out the *ALG6* gene in HEK293EBNA1-6E cells to provide a cellular system that transferred non-glucosylated glycans ($Man_9GlcNAc_2$) to substrates. In these *ALG6*$^{-/-}$cells, a monoglucosylated glycan is solely created by the glucosylation by the UGGTs providing a suitable system to follow the glucosylation process (Figure 3.1B).

To aid in substrate identification, an inhibitor of glucosidases I and II, deoxynojirimycin (DNJ), was added 1 hr prior to cell lysis to block glucose trimming and trap monoglucosylated products. Monoglucosylated substrates were then isolated by affinity purification using recombinant glutathione S-transferase-calreticulin (GST-CRT), as calreticulin binds monoglucosylated proteins. To account for non-specific binding, a lectin-deficient construct (GST-CRT-Y109A) was used as an affinity purification control (Kapoor et al., 2004). Affinity purified substrates were reduced, alkylated, and trypsin digested. The resulting peptides were labeled with tandem mass tags (TMT) (Rauniyar & Yates, 2014), deglycosylated using PNGaseF, and analyzed by mass spectrometry to identify substrates of the UGGTs. The use of TMT, as well as the control GST-CRT-Y109A affinity purification, allows for robust, quantitative identification of substrates of the UGGTs. The resulting data was analyzed by calculating the fold change in abundance of the TMT associated with proteins identified through affinity purification using wild type GST-CRT over affinity purification using GST-CRT-Y109A. To be considered a

UGGT substrate, a cutoff of three-fold (wild type GST-CRT/GST-CRT-Y109A) was applied. This conservative cutoff was set to give a high level of confidence in the identified substrates, as below this cutoff, increasing fractions of non-secretory pathway proteins were found.

Substrate identification of the UGGTs

In order to determine the cellular substrates of the UGGTs, the above glycoproteomics protocol was followed using *ALG6*$^{-/-}$ cells. A restricted pool of thirty-seven N-linked glycosylated proteins was identified as substrates of the UGGTs (Figure 3.1C). Prosaposin, the only previously known endogenous substrate of the UGGTs, was included in this group, supporting the utility of the approach (Pearse et al., 2010). Integrin β-1 showed the most significant fold change (wild type GST-CRT/GST-CRT-Y109A) of ~26-fold, indicating there is a large dynamic range of reglucosylation levels.

The cell localizations of UGGT substrates were then determined by using their Uniprot classification. Approximately two thirds of the UGGT substrates are destined for the plasma membrane or lysosomes (Figure 3.2C and 3.2D). Additional substrates are secreted or are resident to the ER or nuclear membrane. Nuclear pore membrane glycoprotein 210 (NUP210) was the only nuclear membrane protein found to be reglucosylated and it is the sole subunit of the nuclear pore that is N-glycosylated (Beck & Hurt, 2016). The nucleus and ER share a contiguous membrane. Proteins targeted to the nuclear membrane are first inserted into the ER membrane, then move laterally to the nuclear membrane (Katta et al., 2014). Four proteins were designated as ‘multiple localizations’ including cation-independent mannose-6-phosphate receptor (CI-M6PR),

which traffics between the Golgi, lysosome and plasma membrane (Dell'Angelica & Payne, 2001).

To distinguish the general pool of substrates that the UGGTs are expected to be exposed to, N-glycosylated proteins of the secretory pathway proteome (N-glycome) were computationally defined (Supplemental Table 2). The N-glycome is comprised of proteins that are targeted to the ER either for residency in the secretory/endocytic pathways or for trafficking to the plasma membrane or for secretion. The reviewed UniprotKB *H. sapiens* proteome (20,353 total proteins) was queried to identify all proteins annotated as N-glycosylated, resulting in a set of 4,520 proteins. This set was then curated to remove proteins predicted to be mitochondrial, contain less than 50 amino acids or redundant isoforms. The resulting N-glycome contained 4,361 proteins, predicting ~21% of the proteome is N-glycosylated. Comparing UGGT substrates to the N-glycome allows for the characterization of feature preferences of substrates for the UGGTs.

The majority of the N-glycome was either localized to the plasma membrane (37%) or was secreted (20%) according to their Uniprot designations. Smaller fractions of the N-glycome reside in the ER (5%), Golgi (4%) or lysosomes (2%). UGGT substrates are therefore significantly enriched for lysosomal proteins compared to the N-glycome, while all other localizations display a similar distribution to their availability. In total, these results demonstrate the ability to identify substrates of the UGGTs proteomically and suggest that the UGGTs display substrate preferences.

Determination of UGGT1 and UGGT2 specific substrates

There are two ER glucosyltransferase paralogues, UGGT1 and UGGT2, though currently there is no evidence that UGGT2 acts as a protein sensor or a glucosyltransferase in the cell. Therefore, we sought to determine if UGGT2 has glucosyltransferase activity in the cell, and if so, do these two paralogues have different substrate specificities. To address this concern, GST-CRT affinity purification and TMT mass spectrometry were used to identify substrates of UGGT1 in *ALG6*/*UGGT2*$^{-/-}$ cells and potential UGGT2 substrates in *ALG6*/*UGGT1*$^{-/-}$ cells.

With the *ALG6*/*UGGT2*$^{-/-}$ cells, 66 N-glycosylated proteins were identified as reglucosylation substrates using the three-fold cutoff (GST-CRT/CST-CRT-Y109A) (Figure 3.3A). Nearly double the number of UGGT1 substrates were identified through this approach compared to using *ALG6*$^{-/-}$ cells where both UGGT1 and UGGT2 were present. This expansion in substrate number is likely due to the ~50% increase in expression of UGGT1 in *ALG6*/*UGGT2*$^{-/-}$ cells (Figure 3.4A). The substrate demonstrating the most significant fold change (23.5-fold) was CD164, creating a similar dynamic range for reglucosylation to that observed in *ALG6*$^{-/-}$ cells.

To identify possible UGGT2 specific substrates, *ALG6/UGGT1*$^{-/-}$ cells were used to isolate UGGT2 modified substrates. Thirty-four proteins passed the three-fold GST-CRT/GST-CRT-Y109A cutoff, with 33 of these proteins predicted to be N-glycosylated and localized to the secretory pathway (Figure 3.3B). Importantly, this demonstrated for the first time that UGGT2 was a functional glycosyltransferase capable of reglucosylating a range of cellular substrates. The glycoprotein with the most significant fold change was arylsulfatase A (10.4-fold). Notably, 8 of the 9 strongest UGGT2 substrates or, 15 of 33 substrates overall, are lysosomal proteins (Figure 3.3B and Figure 3.5A). While UGGT1

was also observed to engage a significant percentage of lysosomal proteins (27%), 45% of UGGT2 substrates are lysosomal. Both of these percentages are significantly enriched when compared to the N-glycome for which only 2% is comprised of resident lysosome proteins (Figure 3.2B).

UGGT1 substrates were enriched for plasma membrane localized proteins (35%) when compared to UGGT2 substrates (18%), while plasma membrane proteins were found to compose a similar percent of the N-glycome (37%) compared to UGGT1 substrates. Similar percentages of UGGT1 and UGGT2 substrates localize to the ER (18%), are secreted (12%), or are found in multiple localizations (6%) (Figure 3.5A). Even though 4% of the N-glycome is composed of Golgi proteins (Figure 3.2B), neither UGGT1 nor UGGT2 appeared to modify Golgi localized proteins.

The number of UGGT1 substrates was double that of UGGT2 suggesting that UGGT1 carried the main quality control load. Only three out of thirty-three UGGT2 substrates were specific to UGGT2. These three UGGT2 specific substrates included arylsulfatase A, α-N-acetylgalactosaminidase and β-hexosaminidase subunit β (HexB), three soluble lysosomal enzymes (Figure 3.5B and C). Thirty substrates overlapped between UGGT1 and UGGT2, while thirty-six substrates were found to be specific to UGGT1 (Figure 3.5B). The preference for the shared substrates was explored by plotting all proteins identified as a substrate of either glucosyltransferase on a $\log_{10}$ scale of the associated TMT value in *ALG6/UGGT2*$^{-/-}$ cells divided by the values in *ALG6/UGGT1*$^{-/-}$ cells (Figure 3.5C). Proteins enriched as UGGT2 substrates therefore possess positive values while UGGT1 enriched substrates have negative values.

The three substrates found to be specific to UGGT2 clustered away from all other proteins (Figure 3.5C at the top left). The remaining UGGT2 enriched substrates, except for one ER localized protein, localized to the lysosome. All the UGGT2 favored substrates were soluble proteins. In contrast, UGGT1 favored proteins were greater in number and displayed a diversity of localizations with a preference for plasma membrane proteins. These results indicate that UGGT2 is a functional glucosyltransferase, which preferentially engages soluble lysosomal proteins while UGGT1 modifies a wider variety of proteins with a preference for plasma membrane and transmembrane domain-containing proteins in general.

Validation of UGGT substrates

Having identified numerous novel substrates of the UGGTs, a select number of these substrates were tested for reglucosylation to validate the identification approach. Substrates were chosen based on a diversity of topologies, lengths, differences in propensities as UGGT1 or UGGT2 substrates and reagent availability. Monoglucosylated substrates were affinity isolated from *ALG6*$^{-/-}$, *ALG6*/*UGGT1*$^{-/-}$, *ALG6*/*UGGT2*$^{-/-}$ and *ALG6*/*UGGT1*/*UGGT2*$^{-/-}$ cells using GST-CRT compared to CST-CRT-Y109A. Substrates were then identified by immunoblotting with the percent reglucosylation determined by subtracting the amount of protein bound by GST-CRT-Y109A from that of GST-CRT, divided by the total amount of substrate present in the whole cell lysate, and multiplying by 100.

CI-M6PR and insulin-like growth factor type 1 receptor (IGF-1R) are both large type I membrane protein that possess multiple N-glycosylation sites (Figure 3.6D and H).

Overall 10% of CI-M6PR was reglucosylated in *ALG6*$^{-/-}$ cells (Figure 3.6B). The modification level of CI-M6PR was significantly reduced in *ALG6*/*UGGT1*$^{-/-}$, but not *ALG6*/*UGGT2*$^{-/-}$ cells. As a control, reglucosylation was not observed in *ALG6*/*UGGT1*/*UGGT2*$^{-/-}$ cells. A similar profile was observed for IGF-1R where reglucosylation levels reached 12% in *ALG6*/*UGGT2*$^{-/-}$ cells (Figure 3.6E-G). Altogether, these findings were consistent with the quantitative glycoproteomics isobaric labeling results (Figure 3.6C and G), confirming that CI-M6PR and IGF-1R are efficient substrates of UGGT1.

Next, the reglucosylation of the type II membrane protein, ectonucleotide pyrophosphatase/phosphodiesterase family member 1 (ENPP1) was analyzed (Figure 3.7D). ENPP1 was found to be reglucosylated at similar levels in *ALG6*$^{-/-}$ (7%) and *ALG6*/*UGGT1*$^{-/-}$ (7%) cells. In *ALG6*/*UGGT2*$^{-/-}$ cells, reglucosylation increased to 12%, while in *ALG6*/*UGGT1*/*UGGT2*$^{-/-}$ cells reglucosylation decreased to 1% (Figure 3.7A and B). These results suggest that ENPP1 can be reglucosylated by both UGGT1 and UGGT2, with a slight preference for UGGT1, supporting the TMT mass spectrometry results (Figure 3.7C).

The reglucosylation of the smaller soluble lysosomal protein, HexB, was also tested (Figure 3.7E-H). HexB is processed into three disulfide-bonded chains in the lysosome (Mahuran et al., 1988). Only immature or ER localized proHexB was affinity purified by GST-CRT (Figure 3.7E, lanes 2, 5, 8 and 11). HexB was reglucosylated at 34% in *ALG6*$^{-/-}$ cells (Figure 3.7F). No significant change in glucosylation levels were observed when UGGT1 was also knocked out (35%). However, a reduction to 20% reglucosylation of HexB was observed in *ALG6*/*UGGT2*$^{-/-}$ cells, and complete loss of

reglucosylation was observed in *ALG6*/*UGGT1*/*UGGT2*$^{-/-}$ cells. *ALG6*/*UGGT1*$^{-/-}$ cells consistently displayed increased levels of expression of HexB (Figure 3.7E, lane 4), which was supported by RNAseq data (Figure 3.8B). These results confirm the mass spectrometry results, which showed HexB to be a favored substrate of UGGT2 (Fig 3.7G). It is also notable that HexB, as the first validated substrate of UGGT2, is highly reglucosylated. Taken together, these results demonstrate that the mass spectrometry screen accurately identified substrates of the UGGTs, as well as differentiated between substrates specific to either UGGT1 or UGGT2.

Analysis of UGGT substrates

To investigate the properties of the substrates modified by the UGGTs and identify potential types of proteins UGGT1 and UGGT2 modify, a systematic analysis of the substrates of the UGGTs was performed and compared to the general properties of the N-glycome. All characteristics were analyzed using UniprotKB annotations. Initially, the length of substrates was compared to the N-glycome. The N-glycome ranged widely in size, from elabela (54 amino acids) to mucin-16 (14,507 amino acids). The overall amino acid distribution of the N-glycome was significantly shifted smaller compared to the size of UGGT substrates (Figure 3.9A). The median size of the N-glycome was 443 amino acids, compared to 737 for UGGT substrates found in *ALG6*$^{-/-}$ cells. Substrates of both UGGT1 (718 amino acid median) and UGGT2 (585 amino acids) are significantly larger when compared to the N-glycome. This increase in length may lead to more complex folding trajectories, requiring increased engagement with the lectin chaperones for efficient maturation.

The distribution of the number of N-glycans possessed by the N-glycome (median of 2 glycans per glycoprotein) was also shifted significantly smaller than that of UGGT1 (7 glycans) or UGGT2 (5 glycans) substrates (Figure 3.9B). All the UGGT substrates displayed both a larger shifted peak and a prominent extended shoulder compared to the N-glycome. Despite the identification of UGGT1 and UGGT2 substrates generally containing high numbers of N-glycans, multiple substrates possessed as few as two N-glycans, suggesting that the experimental approach did not require a high number of monoglucosylated glycans for GST-CRT affinity isolation.

The ER maintains an oxidizing environment that supports the formation of disulfide bonds. Complex folding pathways can involve the engagement of oxidoreductases, such as the calnexin/calreticulin-associated oxidoreductase ERp57, to catalyze disulfide bond formation and isomerization (Margittai & Sitia, 2011; Kozlov & Gehring, 2020). The most common number of Cys residues in proteins identified as UGGT substrates was 2, which was similar to the N-glycome Cys content (Figure 3.9C). However, there are variations in the median number of Cys residues as for the N-glycome it is 11, which is smaller than that found in *ALG6*$^{-/-}$ cells (16 Cys), and for UGGT1 substrates observed in *ALG6/UGGT2*$^{-/-}$ cells (13 Cys). In contrast, a median of 9 Cys was observed for UGGT2 substrates. Therefore, UGGT1 appears to display a slight preference for proteins with high Cys content, when compared to the N-glycome and UGGT2 substrates.

UGGT1 or UGGT2 substrates displayed similar pI distributions with pIs predominantly near a pH of 6.0, while a second smaller peak centered around a pH of 8.5. Interestingly, a pronounced valley was observed at pH 7.9 under all conditions,

presumably due to the instability of proteins with pIs of a similar pH to that of the ER. The N-glycome displayed a more bimodal distribution with significant population of both acidic and basic pIs (Figure 3.9D). These results suggest that both UGGT1 and UGGT2 preferentially engage proteins with low pIs.

The predicted topologies of the substrates of the UGGTs and the N-glycome were also analyzed. Approximately 70% of the N-glycome is comprised of membrane proteins, with half of these membrane proteins possessing multiple transmembrane domains, followed by single membrane pass proteins with a type I orientation (a third) with the remainder being type II membrane proteins (Figure 3.9E). A total of 43% of UGGT substrates in *ALG6*$^{-/-}$ cells contained a transmembrane domain with the vast majority of these substrates having their C-terminus localized to the cytosol in a type I orientation, while two substrates possessed the reverse type II orientation and a single multi-pass membrane substrate (NPC1) was identified. When the UGGTs were considered separately, about half of the UGGT1 substrates (*ALG6/UGGT2*$^{-/-}$ cells) possessed at least one transmembrane domain, with 70% of these membrane proteins being in the type I orientation, a quarter in a type II orientation and two being multi-pass proteins (NPC1 and scavenger receptor class B member 1 (SR-BI)). In contrast to UGGT1, the majority of UGGT2 substrates were soluble proteins (72%) with the breakdown of remaining transmembrane proteins being similar to that of UGGT1 with the majority being type I membrane proteins. The preference of UGGTs for type I transmembrane proteins is likely caused by their larger luminal-exposed domains and N-glycan numbers compared to multi-pass membrane proteins (Figure 3.10A and B). Notably, substrates of the UGGTs had significantly larger luminal domains than the

membrane proteins of the N-glycome, though especially for the multi-pass membrane proteins (Figure 3.10A). Furthermore, while the pIs of type II and polytopic membrane proteins were bimodal, they were overall more basic, which appears to be a property disfavored by UGGT substrates (Figure 3.10C). Overall, these results show that UGGT1 efficiently modifies both soluble and membrane associated proteins, while UGGT2 strongly favors soluble substrates.

Efficient IGF-1R trafficking requires lectin chaperone engagement

A number of natural substrates of the UGGTs were identified using a glycoproteomics approach with gene edited cell lines. As reglucosylation by the UGGTs can direct multiple rounds of lectin chaperone binding, the necessity for reglucosylation to support the efficient maturation of a reglucosylated substrate was investigated. IGF-1R is proteolytically processed in the *trans*-Golgi by proprotein convertases including furin, facilitating the monitoring of IGF-1R trafficking from the ER to the Golgi (Lehmann et al., 1998). The requirement for lectin chaperone binding and reglucosylation to aid IGF-1R trafficking was analyzed.

Initially, cells were treated without or with the inhibitor of α-glucosidases I and II, DNJ, to accumulate IGF-1R in the triglucosylated state to bypass entry into the calnexin/calreticulin binding cycle (Helenius & Hammond, 1994; Hebert et al., 1995). At steady state as probed by immunoblotting of cell lysates, IGF-1R accumulated in the ER localized pro form relative to the mature form after DNJ treatment (Figure 3.11A) resulting in a 19% decrease in the level of the *trans*-Golgi processed mature protein

(Figure 3.11B). This indicated that the lectin chaperone binding cycle helps support efficient IGF-1R trafficking.

There are two modes for engaging the lectin chaperone cycle: initial binding, which can potentially commence co-translationally for glycoproteins such as IGF-1R that have N-glycans located at their N-terminus through their trimming of the terminal two glucoses by glucosidases I and II; or by rebinding, which is directed by the reglucosylation of unglucosylated species by the UGGTs (Parodi & Caramelo, 2015; Lamriben et al., 2016). The contribution of each mode of monoglucose generation for the proper trafficking of IGF-1R was analyzed.

IGF-1R maturation was investigated in *ALG6*$^{-/-}$ cells as in these cells the N-glycan transferred to the nascent substrate is non-glucosylated, leading to a lack of initial glucosidase trimming mediated lectin chaperone binding. Reglucosylation by the UGGTs is required for lectin chaperone binding in *ALG6*$^{-/-}$ cells. Similar to DNJ treatment in wild type cells, *ALG6*$^{-/-}$ cells demonstrate a 20% decrease in mature IGF-1R relative to the pro form at steady state (Figure 3.11C, lanes 1 and 3, and Figure 3.11D). As hypoglycosylation can occur in a substrate dependent manner in *Alg6*$^{-/-}$ cells (Shrimal & Gilmore, 2015), the mobility of IGF-1R with and without N-glycans (PNGase F treated) was monitored by comparing the mobility of IGF-1R by SDS-PAGE and immunoblotting of wild type and *ALG6*$^{-/-}$ cell lysates. IGF-1R appeared to be fully glycosylated (Figure 3.11C). To confirm that the pro form of IGF-1R represented ER localized protein rather than protein trafficked out of the ER but not processed by proprotein convertases, IGF-1R from wild-type and *Alg6*$^{-/-}$ cells was treated with the endoglycosidase EndoH. As EndoH cleaves high-mannose glycans which are preferentially present in the ER or early Golgi,

an increase in mobility by SDS-PAGE suggests ER localization. In both wild-type and *Alg6*$^{-/-}$ cells, Pro IGF-1R was found to be EndoH sensitive, while mature IGF-1R was found to be largely EndoH resistant (Figure 3.11C, lanes 2 and 5), suggesting the accumulation of pro IGF-1R in *Alg6*$^{-/-}$ cells represents impaired ER trafficking rather than impaired processing in the *trans*-Golgi. Altogether, these steady state results suggest that lectin chaperone binding is important for efficient IGF-1R maturation.

As steady state results can be impacted by changes in protein synthesis and turnover, a radioactive pulse-chase approach was used to follow protein synthesized during a 1 hr [^{35}S]-Met/Cys pulse interval followed by chasing for up to 2-hr under non-radioactive conditions. Pulse-chase experiments are generally performed with overexpressed tag constructs to accumulate and isolate sufficient protein for monitoring. Here, endogenous IGF-1R was isolated by immunoprecipitation with anti-IGF-1R antibodies and analyzed by SDS-PAGE and autoradiography to determine the percent of IGF-1R that was properly processed to its mature form in the *trans*-Golgi. IGF-1R was found to traffic efficiently out of the ER and to the Golgi in wild type cells as 59% of the total protein after a 2-hr chase was mature IGF-1R (Figure 3.12A, lanes 1-3 and F). When lectin chaperone binding was inhibited by treatment with DNJ, mature IGF-1R was diminished to 22%, underscoring the importance of lectin chaperone binding (Figure 3.12A, lanes 4-6 and F).

To delineate the contributions of early compared to late lectin chaperone binding, IGF-1R trafficking was followed in gene edited cells that control the methods for lectin chaperone engagement. A single early round of lectin chaperone binding will be permitted in the absence of both UGGTs or rebinding would only be directed by the

UGGT present with knockouts of a single UGGT. Alternatively, early lectin chaperone binding as dictated by glucosidase trimming will be absent in the *ALG6*$^{-/-}$ cells where lectin chaperone binding is directed solely through glucosylation by the UGGTs. Monitoring the trafficking of IGF-1R in these cells will allow us to determine the contributions of the different steps in the lectin chaperone binding cycle for proper IGF-1R maturation.

When both UGGTs were absent in *UGGT1/2*$^{-/-}$ cells, the percent of mature IGF-1R after 2 hr of chase decreased to 42%. In agreement with early glycoproteomics and affinity isolation results showing IGF-1R was largely a UGGT1 substrate, UGGT2 knockout alone had little influence on IGF-1R trafficking while the knocking out of UGGT1 supported IGF-1R trafficking similar to the double UGGT deletion (Figure 3.12, 7-15 lanes and F). These results support a role for UGGT1 in optimizing IGR-1R trafficking.

To determine the importance of early chaperone binding directed by the glucosidases, IGF-1R trafficking was monitored in *ALG6*$^{-/-}$ cells that support reglucosylation but lack the ability for early binding to the lectin chaperones as directed by glucosidase trimming of the triglucosylated species. In *ALG6*$^{-/-}$ cells, the percent of mature IGF-1R was significantly decreased to 21%, indicative of an important contribution of the initial round of lectin binding, as was suggested by steady state data (Fig 5C). The addition of DNJ to *ALG6*$^{-/-}$ cells would be expected to trap IGF-1R in a monoglucosylated state after glucosylation, allowing the effect of prolonged interaction with the lectin chaperones to be observed. Under this condition, IGF-1R was strongly retained in the ER with no increase observed in the level of mature IGF-1R observed

even after 2 hr of chase (Figure 3.12, lanes 16-21 and F). Altogether these results demonstrate that while early (glucosidase-mediated) and late (UGGT-mediated) lectin chaperone binding contribute to the efficient trafficking from the ER and subsequent Golgi processing of IGF-1R, early lectin chaperone binding appears to be most critical for supporting proper IGF-1R maturation.

Discussion

As lectin chaperone binding is directed by the covalent modification of substrates by the UGGTs, the identification of *bona fide* substrates of the UGGTs is central to understand the impact the lectin chaperone network has on cellular homeostasis. Features of proteins alone cannot accurately predict which chaperones will be required for efficient folding and quality control (Adams, Ke, et al., 2019). Previous studies involving the UGGTs have focused mainly on the overexpression of biasedly selected substrates or using purified proteins, providing uncertain biological relevance (Ritter & Helenius, 2000; Taylor et al., 2003; Caramelo et al., 2004; Soldà et al., 2007; Pearse et al., 2008; Ferris et al., 2013; Tannous et al., 2015). Here, we used a quantitative glycoproteomics-based strategy to identify seventy-one natural cellular substrates of the UGGTs. When compared to the N-glycome that represents the total population of potential substrates (4,361 N-glycoproteins in human cells), the UGGTs favored the modification of more complex, multidomain proteins with large numbers of N-glycans. These results are in agreement with the common requirement of chaperones for the proper folding of more complex proteins (Balchin et al., 2016, 2020). The lectin chaperone system is part of the robust chaperone network necessary to promote the efficient folding and quality control

of substrates and mitigate harmful misfolding events that are associated with a large range of pathologies.

The discovery of 33 UGGT2 cellular substrates provides the first evidence of intact UGGT2 acting as a quality control factor in cells (Figure 3.3B). Previous work demonstrated that UGGT2 is enzymatically active against chemically engineered glycosylated substrates using purified components or when the catalytic domain of UGGT2 was appended to the folding sensor domain of UGGT1 (Arnold & Kaufman, 2003; Y. Takeda et al., 2014). The lower number of UGGT2 substrates compared to UGGT1 (66 substrates) is likely due, at least in part, to UGGT2 being expressed at a fraction of the level of UGGT1 (~4% in HeLa cells (Itzhak et al., 2016)). Of special note is the preference of UGGT2 for lysosomal substrates as 8 of the 9 preferential UGGT2 substrates are lysosomal proteins (Figure 3.5C). The preferential UGGT2 substrates are all soluble proteins, while half of the preferential UGGT1 substrates contained transmembrane domains indicative of a further preference of UGGT2 for soluble proteins (Figure 3.5C). Given the preference of UGGT2 for soluble lysosomal proteins, it would be of interest in future studies to examine lysosomes in *UGGT2*$^{-/-}$ cells as a number of the UGGT2 substrates are associated with lysosomal storage diseases including metachromatic leukodystrophy (arylsulfatase A), Sandhoff disease (β-hexosaminidase subunit β) and Schindler disease (α-N-acetylgalactosaminidase) (Mahuran, 1999; Cesani et al., 2016; Ferreira & Gahl, 2017).

UGGT1 serves as the predominant ER glycoprotein quality control sensor. While overall the 66 UGGT1 substrates are evenly distributed between soluble and membrane proteins, the majority of the most efficiently reglucosylated proteins are membrane

proteins (Figure 3.5C). Seventy percent of the membrane proteins modified by UGGT1 are in the type I orientation possessing luminal N-glycosylated domains of significant length. Only two substrates of the UGGTs are multi-pass membrane proteins (NPC1 and SR-BI). In contrast to most polytopic membrane proteins that have little exposure to the ER lumen (Figure 3.10A), both NPC1 and SR-BI have large heavily glycosylated luminal domains. The enrichment of UGGT1 for transmembrane proteins may be influenced through a weak association with the ER membrane or a general slower and more complex folding process for membrane proteins that provides a longer window for modification.

An important question to ask is what is the basis for the differing substrate specificities of UGGT1 and UGGT2? They display sequence identities that are high within the catalytic domains (83% identical) and lower in their folding sensor domains (49%) (Arnold & Kaufman, 2003). This sequence disparity within the folding sensor domain may drive altered substrate selection. In addition, UGGT1 and UGGT2 may reside in separate subdomains within the ER, which could contribute to substrate accessibility. The CLN6/CLN8 transmembrane complex appears to recognize lysosomal proteins within the ER for COPII packaging in support of a possible mechanism of lysosomal substrate selection (Bajaj et al., 2020). An additional possibility addressed was that the level of expression of the lysosomal proteins identified as UGGT2 substrates may be augmented in *ALG6/UGGT1*$^{-/-}$cells. However, only the mRNA expression level of β-hexosaminidase subunit β was increased relative to *ALG6*$^{-/-}$ or wild type cells, as supported by immunoblot data (Figure 3.7E) with the remaining preferential UGGT2 lysosomal substrates displaying no significantly change in mRNA expression levels (Figure 3.14). The increased expression of β-hexosaminidase subunit β in *ALG6/UGGT1*$^{-}$

$^{-/-}$ cells may be attributed to induction by UPR, as in these cells a slight induction primarily through the ATF6 branch of the UPR was observed (Figure 3.15). Further studies will be required to understand the varying selectivities of the UGGTs.

With some 4,350 possible N-glycosylated proteins as potential UGGT substrates, why were only 71 proteins identified as substrates of the UGGTs? First, many proteins are expected to fold in a chaperone independent manner, especially small, simple proteins. Second, our stringent isolation approach prioritized high quality substrates with at least a 3-fold induction for GST-CRT/GST-CRT-Y109A binding. Third, the profile of reglucosylated substrates is likely cell-type dependent with additional substrates expected to be identified in cell types with heavy secretory pathway loads such as pancreatic cells or hepatocytes, compared to the kidney line used here. Fourth, ~1,500 proteins of the N-glycome are multi-pass transmembrane proteins (Figure 3.9E). This class of protein was strongly de-enriched as substrates of the UGGTs, likely due to their limited luminal exposure and minimal N-glycan content (Figure 3.10A and B). This reduces the pool of favored substrates by a third. Fifth, the monoglucosylated protein isolation procedure may also be limited by possibly requiring multiple sites of reglucosylation for efficient binding to survive the pulldown protocol. However, multiple substrates with two N-glycans were identified, suggesting heavy glycosylation is not an absolute requirement. Additionally, protein expression levels are expected to play some role in substrate identification but it does not appear to be a major determining factor as multiple strong substrates were expressed at or below an average protein level for the N-glycome and no correlation between mRNA expression level and the TMT mass spectrometry fold increase for the GST-CRT/GST-CRT-Y109A fraction was observed (Figure 3.8). It

would be of interest to determine if proteotoxic stress would increase levels and the range of reglucosylated substrates as both the pool of non-native proteins and the amount of the UPR-induced substrates of the UGGTs would be expected to increase. It is also possible that some proteins identified as substrates of the UGGTs may misfold after missing the first round of calnexin/calreticulin binding in *ALG6*$^{-/-}$ cells and therefore engage the UGGTs more efficiently.

As carbohydrate binding can be dictated initially by glucosidase trimming followed by additional later rounds of binding dictated by UGGT reglucosylation, it is of importance to understand which stage of the binding cycle contributes most significantly to proper protein maturation and cell homeostasis. N-glycans in *Sacchromyces cerevisiae* and other single cell species are transferred post-translationally as they are missing the OST isoform subunit that interacts with the Sec61 translocon and supports early co-translational modification (Ruiz-Canada et al., 2009; Shrimal et al., 2019). A second OST isoform appears in multicellular organisms that is translocon-associated. In addition, reglucosylation activity was first observed in single cell parasites of *Trypanosoma cruzi* where glycans are transferred as $Man_9GlcNAc_2$ moieties thereby bypassing the initial glucosidase initiated binding step observed in metazoans (Parodi & Cazzulo, 1982). These seminal *T. cruzi* studies from Parodi and colleagues that first discovered the (re)glucosylation activity, later attributed to UGGT1, were the inspiration for the development of the experimental *ALG6*$^{-/-}$ system used in this study to isolate substrates of the UGGTs. Conservation analysis of glycosylation and the lectin chaperone pathway suggests that reglucosylation supporting the quality control function of the calnexin cycle evolved prior to its role in assisting in earlier folding events.

Using CRISPR edited cell lines, the contributions of the various steps for chaperone binding engagement for the UGGT1 substrate IGF-1R was experimentally explored as its processing in the Golgi provided a robust Golgi trafficking assay. Furthermore, IGF-1R is a target in cancer biology as it is important for cell growth (Sell et al., 1994; Desbois-Mouthon et al., 2006; Chng et al., 2006; King et al., 2014; Mutgan et al., 2018). When binding to the lectin chaperones was blocked in wild type cells by glucosidase inhibition with DNJ treatment, supporting the production of triglucosylated trapped species, the percent of processed IGF-1R strongly decreased compared to untreated cells, demonstrating a requirement of lectin chaperone engagement for the efficient maturation, trafficking and processing of IGF-1R. In *UGGT1/2*$^{-/-}$ cells, IGF-1R can enter the first round of glucosidase-mediated binding to the lectin chaperones but rebinding directed primarily by UGGT1 mediated reglucosylation cannot occur (Figures 3.12A, B and Figure 3.13). This led to a reduced efficiency in the accumulation of mature IGF-1R. The first round of lectin chaperone binding is bypassed in *ALG6*$^{-/-}$ cells as the N-glycans transferred to proteins do not contain glucoses (Figure 3.13). Therefore, only the rebinding events mediated by reglucosylation take place. More strikingly in *ALG6*$^{-/-}$ cells, this led to a dramatic reduction in IGF-1R processing at a greater level than in *UGGT1/2*$^{-/-}$ cells, indicating the first round of binding to the lectin chaperones was most critical for IGF-1R maturation. The addition of DNJ in *ALG6*$^{-/-}$ cells supported the trapping of reglucosylated side chains and severely reduced Golgi processing, suggesting that reglucosylation-mediated persistent interaction with the lectin chaperones delays IGF-1R exit from the ER.

Understanding the proteins that interact with or rely on chaperone systems will advance our understanding of protein homeostasis (Houry et al., 1999; Kerner et al., 2005). Large multi-domain proteins such as IGF-1R and many of the other substrates of the UGGTs have apparently evolved to utilize the lectin chaperone system to help direct their complex folding trajectories. The co-evolution of chaperones and their substrates has led to the expansion of the complexity of the proteome for multicellular organisms (Balchin et al., 2016; Rebeaud et al., 2020). The large group of substrates of the UGGTs identified here represents glycoproteins that utilize multiple rounds of lectin chaperone engagement for proper maturation and are likely more prone to misfold under stress. Future studies will determine if this increased vulnerability makes these substrates more susceptible to misfold under disease conditions where cell homeostasis is challenged.

Experimental Methods

Reagents

Antibodies used were: rabbit monoclonal IGF-1 receptor β (D23H3, Cell Signaling), rabbit monoclonal IGF-IIR/CI-M6PR (D3V8C, Cell signaling), rabbit monoclonal BiP (C50B12, Cell Signaling), rabbit monoclonal β-hexosaminidase subunit β (HEXB) (EPR7978, Abcam), rabbit polyclonal ENPP1 (N2C2, Genetex) rabbit polyclonal UGGT1 (GTX66459, Genetex), mouse monoclonal Glyceraldehyde 3-Phosphate (MAB374, Millipore Sigma), IRDye x anti-rabbit secondary (LiCor). All chemicals were purchased from Millipore-Sigma, except where indicated.

Cell culture

HEK293-EBNA1-6E cells were employed and used as the parental line to create all CRISPR/Cas9 edited lines (Tom et al., 2008). Cells were cultured in DMEM (Sigma) supplemented with certified 10% fetal bovine serum (Gibco) at 37 °C at 5% CO_2. Cells were tested for the presence of mycoplasma using a universal mycoplasma detection kit (ATCC, Cat # 30-012K).

CRISPR/Cas9-mediated knock outs

HEK293EBNA1-6E *ALG6*$^{-/-}$, *ALG6/UGGT1*$^{-/-}$, *ALG6/UGGT2*$^{-/-}$, *ALG6/UGGT1/UGGT2*$^{-/-}$, *UGGT1*$^{-/-}$, *UGGT2*$^{-/-}$, and *UGGT1/2*$^{-/-}$ cells were generated via CRISPR/Cas9 using gRNA plasmids gh260, gh172, and gh173, and Cas9-GFP plasmid CAS9PBKS (Lonowski et al., 2017; Narimatsu et al., 2018). Plasmids gh260 (106851), gh172 (106833), gh173 (106834), and CAS9PBKS (68371) were from Addgene. Knock-out cell lines were generated by co-transfecting HEK293-EBNA1-6E cells at 70% confluency in a 10-cm plate with 7 μg of both the associated gRNA and Cas9-GFP plasmid, using a 2.5 μg of PEI per 1 μg of plasmid. Cells were allowed to grow for 48 hr prior to trypsinization and collection. After trypsinization, cells were collected and washed twice with sorting buffer (1% FBS, 1mM EDTA, PBS). Cells were then resuspended in sorting buffer at approximately 1 million cells per ml. Cells were then bulk separated using flow assisted cell sorting based on the top 10% of Cas9-GFP expressing cells (FACS Aria II SORP, Becton Dickinson and Company). Cells were then plated at 5, 10, 20 thousand cells per 10 cm plate in pre-conditioned DMEM media with 20% FBS. Colonies derived from a single cell were isolated using cell cloning cylinders (Bellco Glass), trypsinized from the plate, and further passaged. Knock-outs were

confirmed by immunoblotting and staining for UGGT1 or, where antibodies were not available, isolating genomic DNA using a genomic DNA isolation kit (PureLink genomic DNA mini kit, Thermo Fisher), PCR amplification of the genomic DNA region of interest, and insertion of genomic DNA into pcDNA3.1-. Plasmids were then sequenced for conformation (Genewiz).

GST-CRT purification

The plasmid for pGEX-3X GST-CRT was from Prof. M. Michalak (University of Alberta). pGEX-3X GST-calreticulin-Y109A was generated by site-directed mutagenesis. GST-CRT was expressed in BL21 *E. Coli* cells in LB medium containing ampicillin at 100 μg/ml. Cultures were grown at 37 °C with shaking until an O.D. of A_{600}=0.6. Protein expression was then induced by treating cultures with 8.32 mg/L IPTG for 2 hr. Cultures were centrifuged at 1,000 g for 10 min. Cell pellets were lysed with cold lysis buffer (1 mM phenylmethylsulfonyl fluoride, 2% Triton X-100, PBS pH 7.4) and resuspended. Resuspended cells were lysed in a microfluidizer (110L, Microfluidics) at 18,000 psi for two passes. The cell lysate was centrifuged for 40 min at 8,000 g at 4 °C. Lysate was filtered through a 0.45 μm filter. Two ml bed volume glutathione sepharose beads (GE Lifesciences, Cat# GE17-0756-01) per liter of lysate was equilibrated in wash buffer (1% Triton X-100, 1 mM PMSF, PBS pH 7.4), added to cleared lysate, and rotated at 4 °C for 3 hr. Beads were precipitated through centrifugation at 1,000 g for 5 min at 4 °C. The supernatant was aspirated and beads were washed twice in wash buffer with gentle resuspension between washes. One ml of elution buffer (10 mM reduced glutathione, 1 mM PMSF, 50 mM Tris pH 8.5) was added to

beads and beads were gently resuspended and allowed to incubate for 5 min at 4 °C. Beads were precipitated by centrifugation at 1,000 g for 5 min 4 °C. The eluate was collected and a total of 6-elutions were collected. Resulting eluate was tested for purity and protein amount on a reducing SDS-PAGE and stained with Imperial protein stain (Thermo Fisher, Cat# 24617). Elutions were then combined and protein concentration was quantified by a Bradford assay (Bio-Rad). Purified protein was then stored at -80 °C in a 20% glycerol PBS buffer at 1 mg/ml.

GST-CRT isolation and TMT mass spectrometry sample preparation

Five 10 cm plates were seeded with 3.5 million cells and allowed to grow for 48 hr. Cells were treated with N-butyldeoxynojirimycin hydrochloride (DNJ) (Cat # 21065, Cayman Chemicals) at 500 μM for 1 hr. Prior to lysis, the media was aspirated and cells were washed once with filter sterilized PBS. Cells were lysed in 1 ml of lysis buffer (20 mM MES, 100 mM NaCl, 30 mM Tris-HCl pH 7.5, 0.5% Triton X-100) per plate. Samples were shaken at 4 °C for 5 min and centrifuged at 20,800 g at 4 °C for 5 min. Lysate was pre-cleared with 25 μl bed volume of buffer-equilibrated glutathione beads per 1 ml of lysate under rotation for 1 hr at 25 μl bed volume. Beads were precipitated by centrifugation at 950 g at 4 °C for 5 min. Glutathione beads were pre-incubated with either GST-CRT or GST-CRT-Y109A by equilibrating 25 μl bed volume/pull-down glutathione beads with lysis buffer. Beads were incubated with 100 μg of purified GST-CRT/pull-down under gentle rotation at 4 °C for 3 hr. Beads were centrifuged at 950 g at 4 °C for 5 min and washed twice with lysis buffer. Supernatant was collected and split in half, with one half incubated for 14 hr at 4 °C under gentle rotation with glutathione

beads pre-incubated with GST-CRT and the other half under the same conditions with GST-CRT-Y109A.

After incubation with GST-CRT beads, samples were washed once in lysis buffer without protease inhibitors and twice in 100 mM triethylammonium bicarbonate (Thermo Fisher Cat# 90114). After the final wash, samples were incubated with 10 µl of 50 mM DTT (Pierce, Cat# A39255) for 1 hr at room temperature under gentle agitation. Samples were treated with 2 µl of 125 mM iodoacetamide (Pierce, Cat# A39271) and incubated for 20 min under gentle agitation, protected from light. Samples were digested with 5 µg of trypsin (Promega, Cat# V5280) at 37 °C overnight under agitation. Peptide concentration was quantified using a BCA protein quantification kit (Pierce, Cat# 23227). 0.8 mg 10plex tandem mass tags (TMT) (Themo Fisher) were resuspended in mass spectrometry grade acetonitrile and was added to digested peptide and incubated for 1-hr at room temp, per manufacturer's instructions. Labeling was quenched by adding hydroxylamine to 0.25% and incubating for 15 min at room temp. Labeled samples were pooled, treated with 1,000 units of glycerol-free PNGaseF (NEB, Cat# P0705S), and incubated for 2-hr at 37 °C. Samples were cleaned using C18 tips (Pierce, Cat# 87784), and eluted in 75% mass spectrometry grade acetonitrile, 0.1% formic acid (TCI Chemicals). Sample peptide concentration was then quantified using a colorimetric assay (Pierce, Cat# 23275).

Mass spectrometry data acquisition

An aliquot of each sample equivalent to 3 µg was loaded onto a trap column (Acclaim PepMap 100 pre-column, 75 µm × 2 cm, C18, 3 µm, 100 Å, Thermo Scientific)

connected to an analytical column (Acclaim PepMap RSLC column C18 2 μm, 100 Å, 50 cm × 75 μm ID, Thermo Scientific) using the autosampler of an Easy nLC 1000 (Thermo Scientific) with solvent A consisting of 0.1% formic acid in water and solvent B, 0.1% formic acid in acetonitrile. The peptide mixture was gradient eluted into an Orbitrap Fusion mass spectrometer (Thermo Scientific) using a 180 min gradient from 5%-40%B (A: 0.1% formic acid in water, B:0.1% formic acid in acetonitrile) followed by a 20 min column wash with 100% solvent B. The full scan MS was acquired over range 400-1400 m/z with a resolution of 120,000 (@ m/z 200), AGC target of 5e5 charges and a maximum ion time of 100 ms and 2 s cycle time. Data dependent MS/MS scans were acquired in the linear ion trap using CID with a normalized collision energy 35%. For quantitation of scans, synchronous precursor selection was used to select 10 most abundant product ions for subsequent MS^3 using AGC target 5e4 and fragmentation using HCD with NCE 55% and resolution in the Orbitrap 60,000. Dynamic exclusion of each precursor ion for 30 s was employed. Data were analyzed using Proteome Discoverer 2.4.1 (Thermo Scientific).

Computational determination of the human N-glycome and substrates analyses

The human N-glycome was defined by the total predicted N-glycosylated proteins from the reviewed human proteome from the UniprotKB (accessed 8/10/2020). Both manual and automated curation of the data set was preformed to remove mitochondrial proteins as well as proteins smaller than 50 amino acids from the dataset. All annotations were derived directly from the UniprotKB information and annotations available for

these proteins were analyzed in R. Determination of the pI values were performed by the pI/MW tool on the Expasy database.

Reglucosylation validation assay

Five 10 cm plates were seeded with 3.5 million cells each and allowed to grow for 48 hr. Cells were treated with DNJ at 500 μM for 14 hr. Prior to lysis, the media was aspirated and cells were washed once with filter sterilized PBS. Cells were lysed in 1 ml of MNT (20 mM MES, 100 mM NaCl, 30 mM Tris pH 7.5, 0.5% Triton X-100) with protease inhibitors (50 μM Calpain inhibitor I, 1 μM pepstatin, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 400 μM PMSF) and 20 mM N-ethyl maleimide, shaken vigorously for 5 min at 4 °C, and centrifuged for 5 min at 17,000 g at 4 °C. 50 μl bed volume of glutathione beads was added to each pull-down and incubated for 1 hr at 4 °C under gentle rotation. Beads were then precipitated by centrifugation at 1,000 g for 5 min at 4 °C. Supernatant was collected with 10% used for WCL and the remainder split evenly between GST-CRT and GST-CRT-Y109A conjugated glutathione beads, which were generated as previously described, and incubated for 16 hr at 4 °C under gentle rotation. Beads were precipitated at 1,000 g for 5 min at 4 °C. Supernatant was aspirated and beads were washed twice with lysis buffer without protease inhibitors. Beads were treated with reducing sample buffer (30 mM Tris-HCl pH 6.8, 9% SDS, 15% glycerol, 0.05% bromophenol blue). WCLs were trichloroacetic acid (TCA) precipitated by adding TCA to cell lysate to a final concentration of 10%. Cell lysate was then briefly rotated and allowed to incubate on ice for 15 min before centrifugation at 17,000 g for 10 min at 4 °C. Supernatants were aspirated and washed twice with cold acetone and centrifuged at

17,000 g for 10 min at 4 °C. Supernatants were aspirated and the remaining precipitant was allowed to dry for 5 min at room temperature and briefly at 65 °C. Precipitated protein was resuspended in sample buffer. Samples were resolved on a 9% reducing SDS-PAGE and imaged by immunoblotting.

Metabolic labeling and IGF-1R immunoprecipitation

Two million cells were plated in 6 cm plates and allowed to grow for 40 hr. Cells were pulse labeled for 1 hr with 120 μCi of EasyTag Express^{35}S Protein Labeling Mix [^{35}S]-Cys/Met (PerkinElmer; Waltham, MA). Immediately after the radioactive pulse, cells were washed with PBS and either lysed in MNT with a protease inhibitor cocktail (Halt protease and phosphatase inhibitor single-use cocktail, Thermo Fisher) and 20 mM NEM, or chased for indicated time using regular growth media. Where indicated, cells were treated with 500 μM DNJ for 30 min prior to [^{35}S]-Cys/Met labeling and through the chase. Cell lysates were shaken for 5 min at 4 °C, centrifuged at 17,000 g for 5 min at 4 °C, and the supernatants were collected. Samples were pre-cleared with a 20 μl bed volume of protein-A sepharose beads (GE Healthcare) by end-over-end rotation for 1 hr at 4 °C. The supernatants were collected and incubated with a 30 μl bed volume of protein-A-sepharose beads and 1.5 μl of α-IGF-1 receptor β (D23H3) XP (Cell Signaling) per sample. Samples were washed with MNT without protease inhibitors or NEM and eluted in sample buffer. Samples were then resolved on a 9% reducing SDS-PAGE, imaged using a GE Typhoon FLA 9500 phosphorimager (GE Healthcare), and quantified using ImageJ.

Glycosylation assay

Three million cells for each indicated cell line were plated in a 10-cm plate and allowed to grow for 48 hr. Cells were lysed in 300 µl RIPA buffer (1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 150 mM NaCl, 50 mM Tris-HCl pH 8.0) with protease inhibitor cocktail and 20 mM NEM. Samples were then sonicated for 20-sec at 40% amplitude (Sonics vibra cell VC130PB), shaken vigorously for 5 min, and centrifuged for 5 min at 17,000 g. 20 µl of the resulting lysate was heated at 95 °C for 5 min, and treated with either 10 µl of PNGaseF or EndoH for 1 hr at 37 °C, according to the manufacturer's instructions (NEB). Samples were diluted 1:1 into sample buffer and imaged by immunoblotting.

RNAseq library preparation and Sequencing

Three million cells for each indicated cell line were plated in 10 cm plates and allowed to grow for 48 hr. Cells were then lysed in TRIzol buffer and RNA was isolated using RNA Clean Concentrate Kit with in-column DNase-I treatment (Zymo Research Corp), following manufacturer instructions. The quantity of RNA was assayed on Qubit using RNA BR assay (Life Technologies Corp), and quality was assessed on Agilent 2100 Bioanalyzer using RNA 6000 Nano Assay (Agilent Technologies Inc). Total RNA was used to isolate poly(A) mRNA using NEBNext Poly(A) mRNA Magnetic Isolation Module, and libraries were prepared using NEBNext UltraII Directional RNA Library Prep Kit for Illumina (New England Biolabs) following manufacturer instructions. The quantity of library was assayed using Qubit DNA HS assay (Life Technologies Corp), and quality was analyzed on Bioanalyzer (Agilent Technologies Inc). Libraries were

sequenced on Illumina NextSeq 500 platform using NextSeq 500/550 High Output v2 kit (150 cycles) with 76 bp paired-end sequencing chemistry.

Sequence quality was assessed using FastQC (Andrews, n.d.) and MultiQC (Ewels et al., 2016). Reads were aligned to the hg38 human reference genome using STAR (Dobin et al., 2013). Transcript abundance was quantified using RSEM (B. Li & Dewey, 2011) and normalized to counts per million (CPM) in R using the edgeR software package (Robinson et al., 2010). Analyses to compare gene expression between cell types was conducted in Excel by finding the average CPM in the pool of genes of interest for the associated cell type and determining the standard deviation away from the average for each gene of interest.

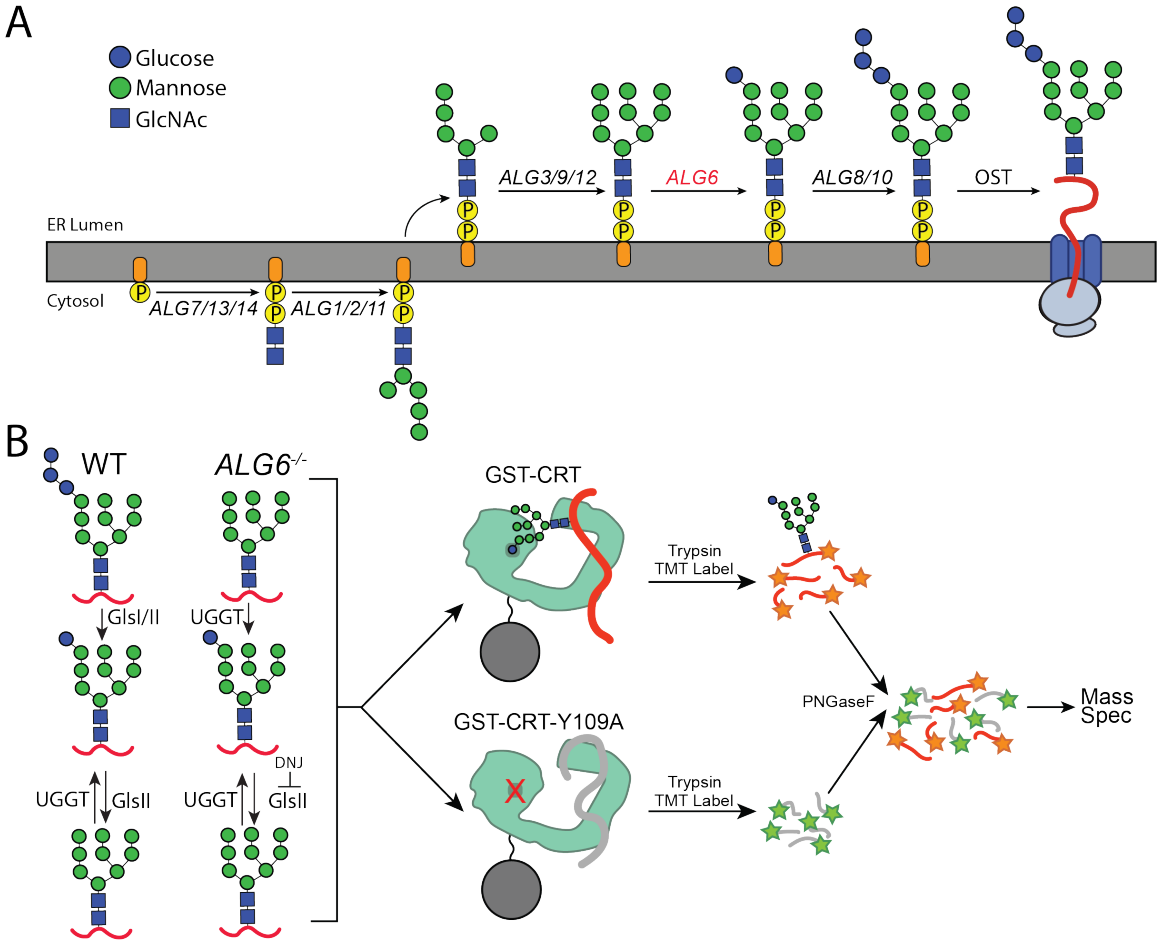

Figure 3.1 Reglucosylation substrate identification experimental design

(A) The pathway of N-glycosylation in eukaryotic cells is depicted. N-glycan synthesis is initiated in the outer ER membrane leaflet on a dolichol-P-phosphate facing the cytoplasm. Flipping of the precursor N-glycan to the ER luminal leaflet and further synthesis steps mediated by ALG proteins leads to eventual transfer of a $Glc_3Man_9GlcNAc_2$ N-glycan to a substrate by the OST complex. ALG6 (red lettering) catalyzes the transfer of the initial glucose onto the Man_9 precursor N-glycan. (B) In wild type (WT) cells, a $Glc_3Man_9GlcNAc_2$ N-glycan is transferred to substrates. Monoglucosylated substrates may therefore occur via trimming by glucosidases I/II (GlsI/II) or reglucosylation by UGGT1/2. In *ALG6*$^{-/-}$ cells, a $Man_9GlcNAc_2$ N-glycan is transferred to substrates. Therefore, monoglucosylated substrates may only occur through reglucosylation by UGGT1/2. DNJ (500 μM) was added to block the trimming of monoglucosylated substrates by GlsII. *ALG6*$^{-/-}$ cells were then lysed and split equally between affinity purifications with either GST-CRT or GST-CRT-Y109A bound to glutathione beads. Affinity-purified samples were then reduced, alkylated, trypsinzed, and labeled with TMT labels. Samples were then deglycosylated with PNGaseF, pooled, and analyzed by mass spectrometry.

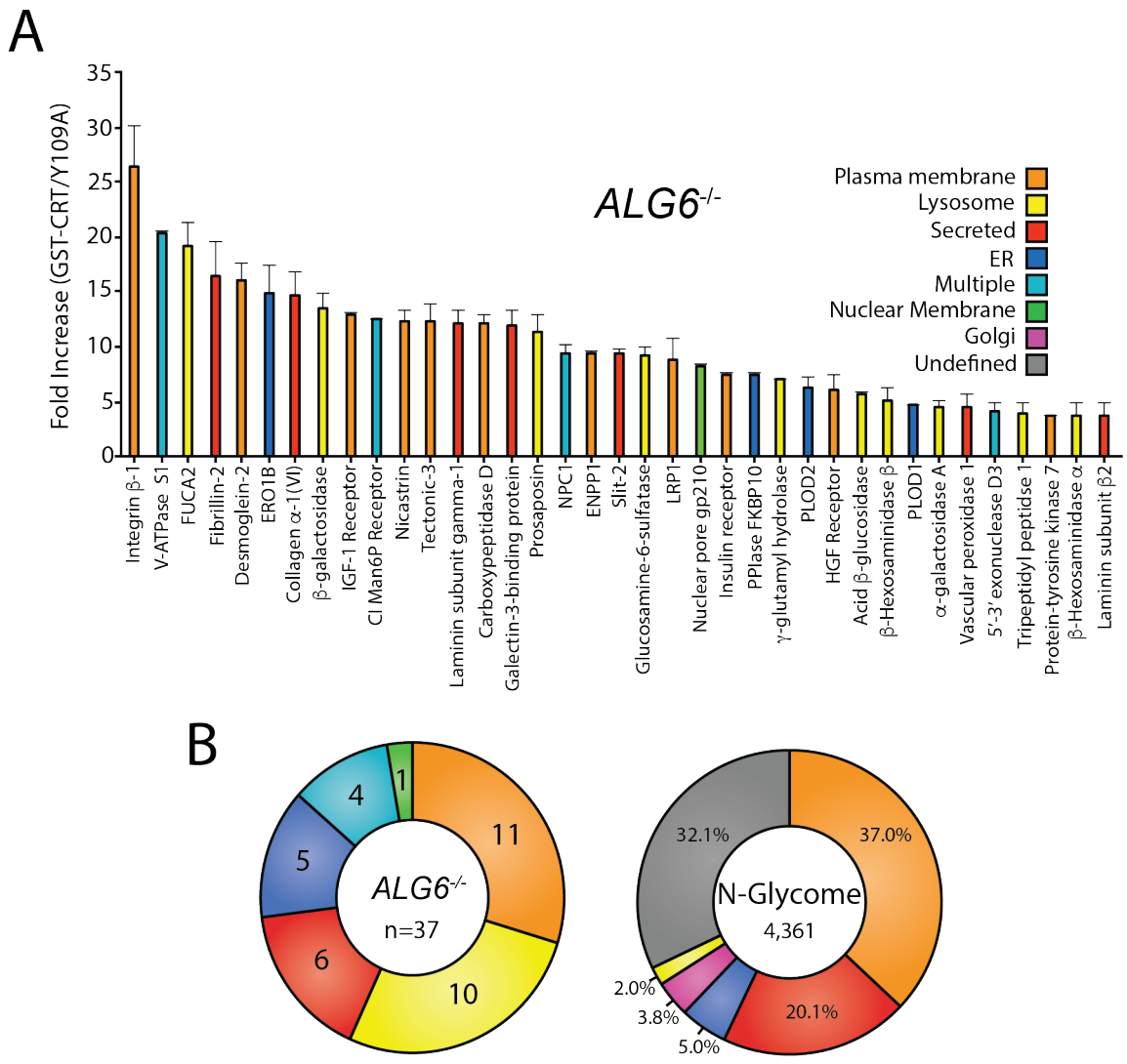

Figure 3.2 Identification of the substrates of the UGGTs

(A) Substrates were identified by dividing the quantification of the TMT label in the GST-CRT condition for each protein by that of the associated GST-CRT-Y109A condition, yielding the fold increase. Localization as predicted by Uniprot annotation is depicted. A cutoff of three-fold increase was applied. Data is representative of two independent experiments. Error bars represent standard error of the mean (SEM). (B) The N-glycome was computationally determined by collecting all proteins annotated to contain an N-glycome by Uniprot. Annotated localization information was then used to computationally determine the localization distribution of the N-glycome as well as the identified UGGT substrates.

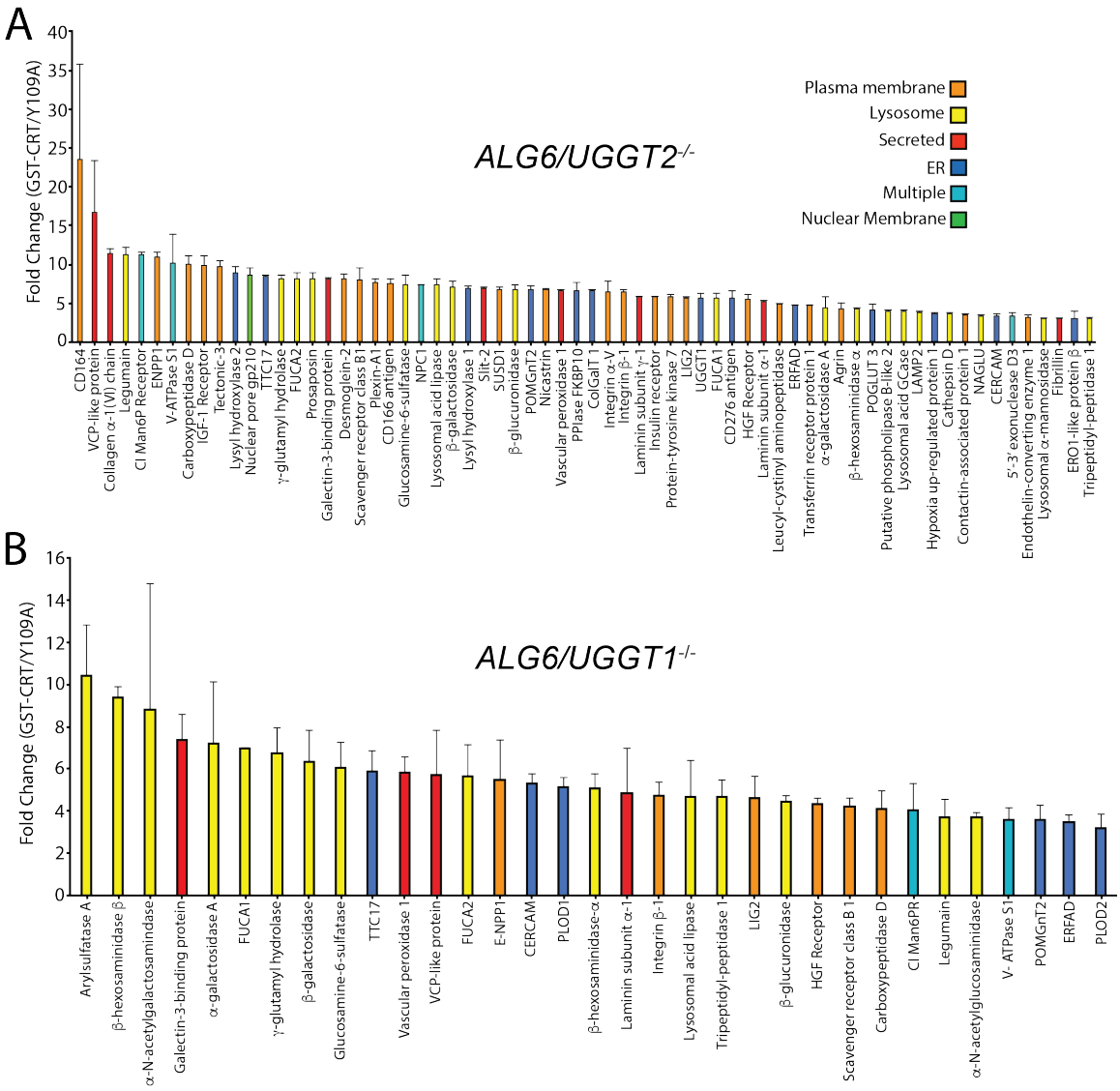

Figure 3.3. Identification of UGGT1 and UGGT2 specific substrates

(A) Reglucosylation substrates in *ALG6/UGGT2*$^{-/-}$ cells were identified and quantified as previously described in Figure 3.1. Localizations as annotated by Uniprot are depicted. Data are representative of two independent experiments. Error bars represent SEM. (B) Reglucosylation substrates in *ALG6/UGGT1*$^{-/-}$ cells were identified and quantified as previously described above.

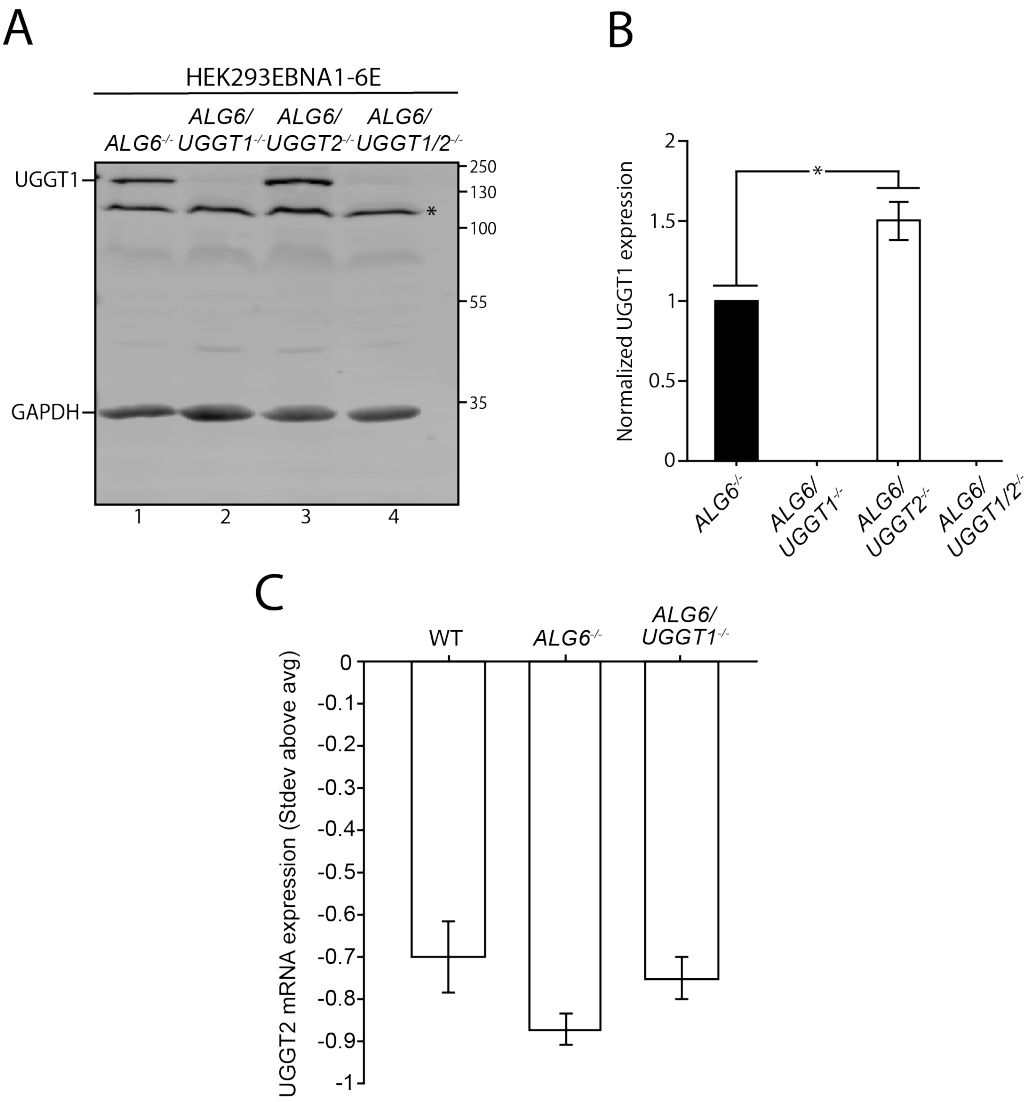

Figure 3.4 UGGT1 and UGGT2 expression

(A) The indicated cells were lysed and whole cell lysates were resolved by SDS-PAGE and imaged by immunoblotting against UGGT1 and GAPDH. Asterisk denotes background band. Data are representative of three independent experiments with quantification shown in B. UGGT1 expression was normalized to that of *ALG6*$^{-/-}$ cells. Error bars represent standard deviation. Asterisk denotes a p-value of less than 0.05. (C) Counts per million of UGGT2 mRNA generated by RNAseq from Supplemental Table 4 was analyzed for the level of *UGGT2* mRNA expression in the indicated cell lines. Counts per million of all genes were averaged and the standard deviation from the average for *UGGT2* mRNA was determined. Error bars represent the standard deviation. Data are representative of three independent experiments.

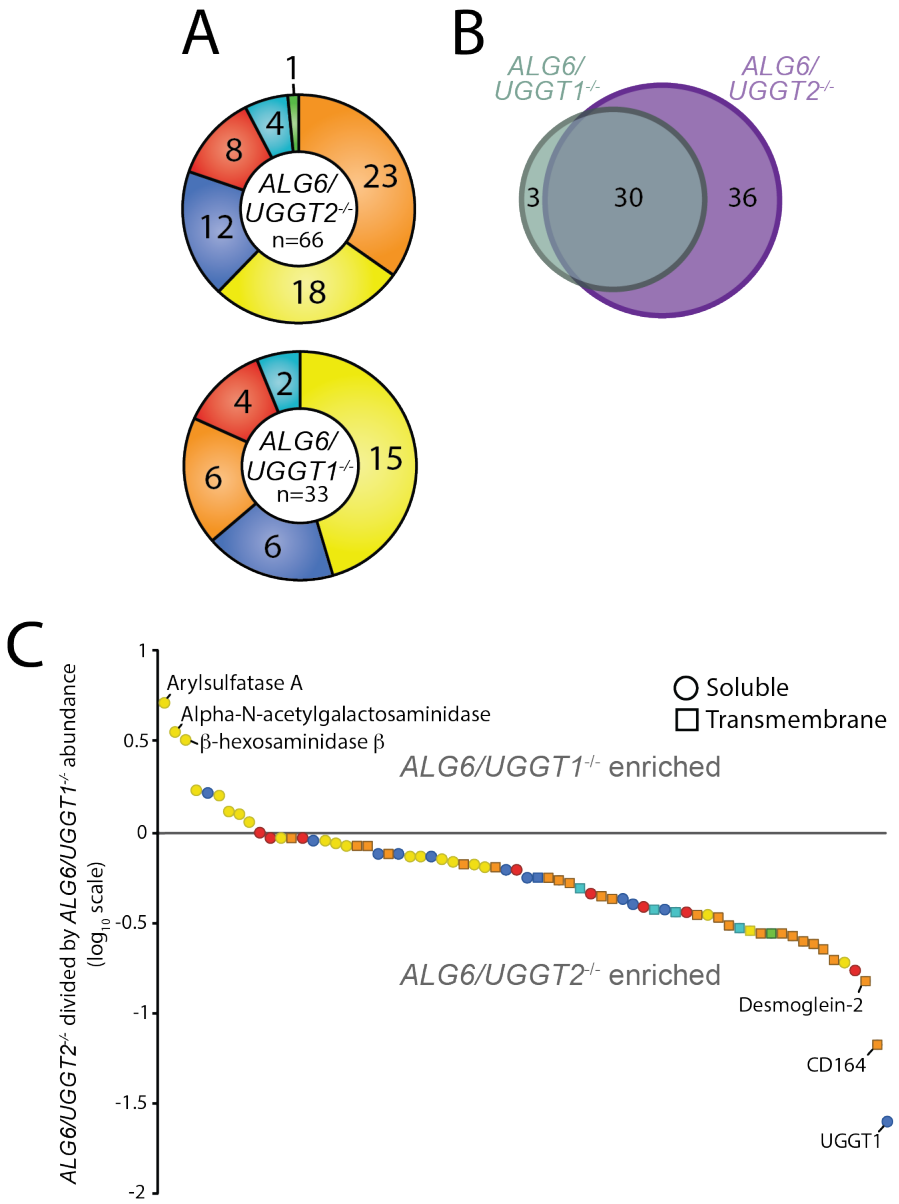

Figure 3.5 UGGT1 and UGGT2 substrates comparison

(A) The distribution of localizations as annotated by Uniprot for reglucosylation substrates identified in both *ALG6/UGGT2*$^{-/-}$ and *ALG6/UGGT1*$^{-/-}$ cells is depicted. (B) The overlap of reglucosylation substrates identified in both *ALG6/UGGT2*$^{-/-}$ cells (purple) and *ALG6/UGGT1*$^{-/-}$ cells (grey) is visualized by a Venn diagram. (C) Reglucosylation substrate enrichment in either *ALG6/UGGT1*$^{-/-}$ or *ALG6/UGGT2*$^{-/-}$ cells is depicted by dividing the TMT quantification for each protein in *ALG6/UGGT2*$^{-/-}$ cells by the associated value in *ALG6/UGGT1*$^{-/-}$ cells on a $\log_{10}$ scale. Positive and negative values represent enrichment in *ALG6/UGGT1*$^{-/-}$ and *ALG6/UGGT2*$^{-/-}$ cells, respectively. Localization based on Uniprot annotation is depicted. Proteins are depicted as either circles (soluble) or squares (transmembrane), as annotated by Uniprot.

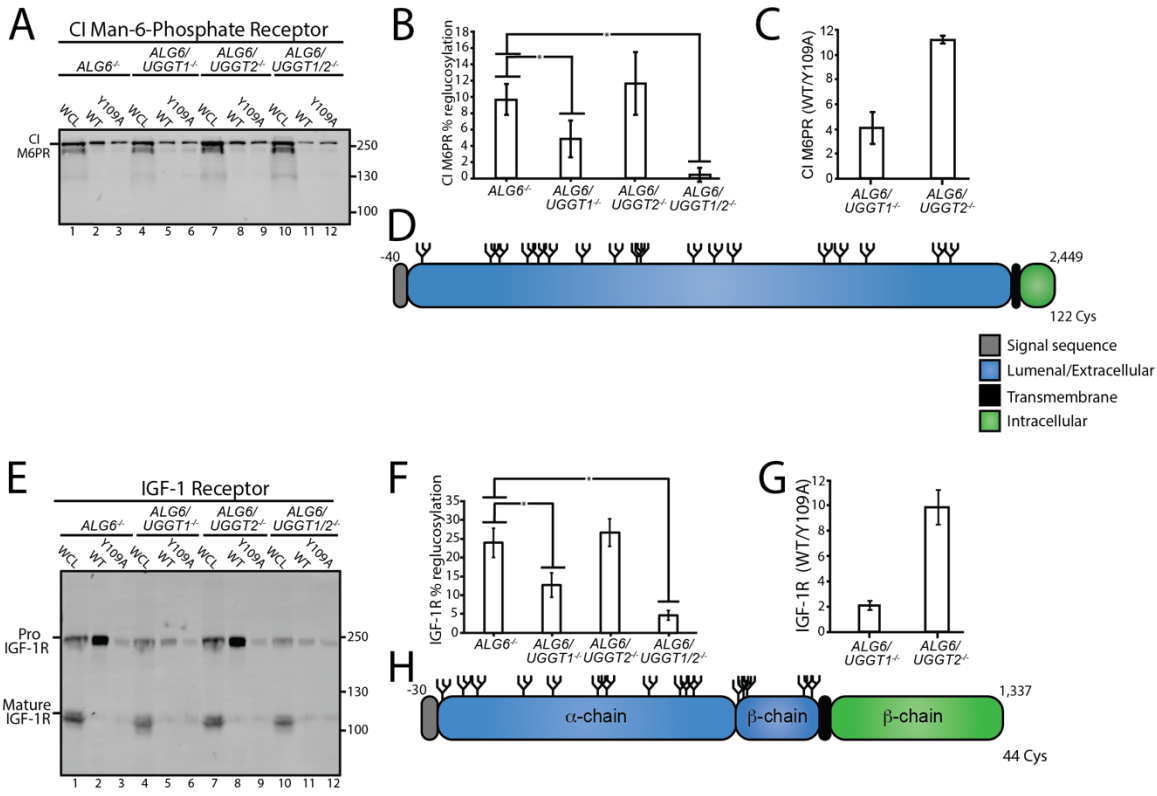

Figure 3.6 Validation of CI Man-6-Phosphate receptor and IGF-1R reglucosylation

(A) The designated cell lines were lysed and split into whole cell lysate (WCL, 10%) or affinity purification by GST-CRT-WT or GST-CRT-Y109A and imaged by immunoblotting against the CI Man-6-Phosphate receptor. Data is representative of three independent experiments with quantification shown in panel B. Quantifications were calculated by subtracting the value of protein in the Y109A lane from the value of protein in the associated WT lane, divided by the value of protein in the associated WCL lane. Error bars represent the standard deviation. Asterisks denote a p-value of less than 0.05 (C) TMT mass spectrometry quantification of CI Man-6-Phosphate receptor reglucosylation from *ALG6/UGGT1*$^{-/-}$ cells (Figure 3.3B) and *ALG6/UGGT2*$^{-/-}$ cells (Figure 3.3A). (D) Cartoon representation of CI Man-6-Phosphate receptor with N-glycans (branched structures), the signal sequence (grey), luminal/extracellular domain (blue), transmembrane domain (black) and intracellular domain (green) depicted. Number of amino acids and Cys residues are indicated. (E) Reglucosylation of IGF-1R, conducted as previously described above. Pro IGF-1R and mature IGF-1R are both observed due to proteolytic processing. Data are representative of three independent experiments with quantification displayed in F. (G) TMT mass spectrometry quantification of IGF-1R from Figure 3.3A and B, as previously described. (H) Cartoon depiction of IGF-1R.

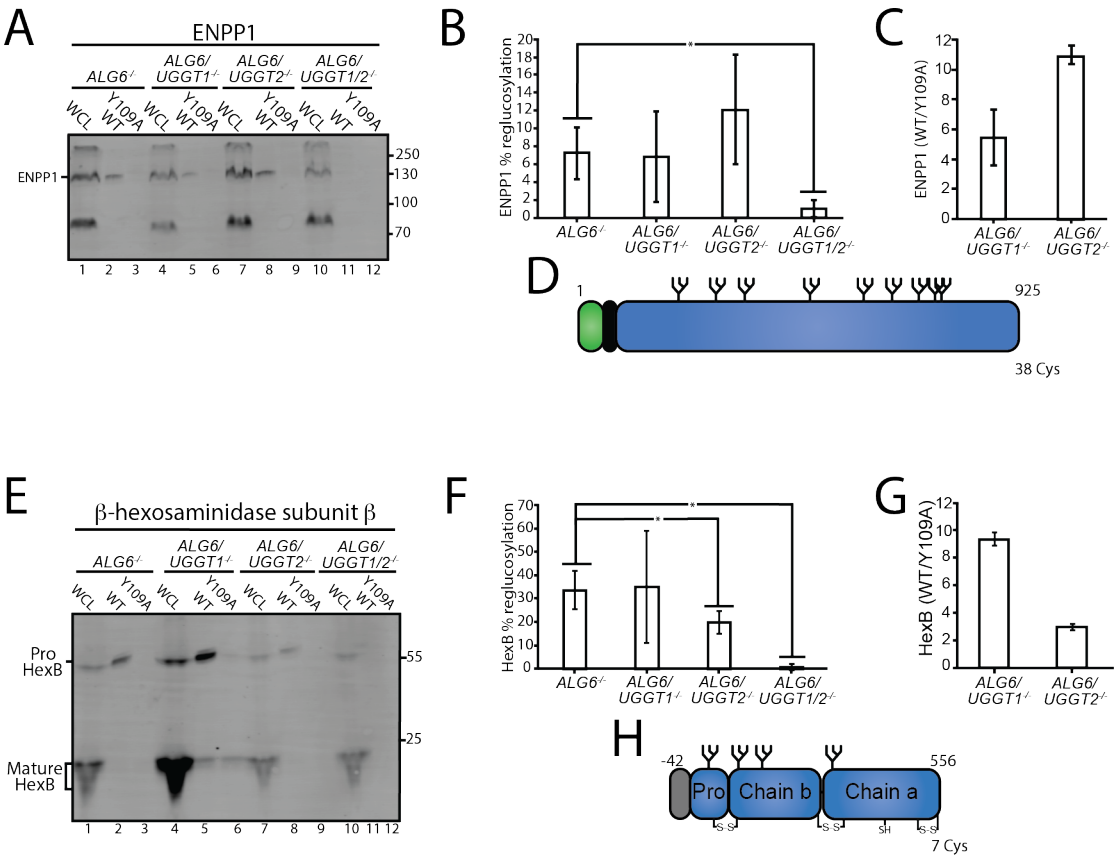

Figure 3.7 Validation of ENPP1 and β-hexosaminidase subunit β reglucosylation

(A) The reglucosylation of ENPP1 shown with quantification displayed in B, as described in Figure 3.6 (C) TMT mass spectrometry quantification of ENPP1 from Figure 3.3A and B with cartoon depiction of ENPP1 in D as described in Figure 3.6 (E) Reglucosylation of β-hexosaminidase subunit β with quantification displayed in F and TMT mass spectrometry quantification of β-hexosaminidase subunit β from Figure 3.3A and B in G with a cartoon depicting β-hexosaminidase subunit β in H, as described in Figure 3.6

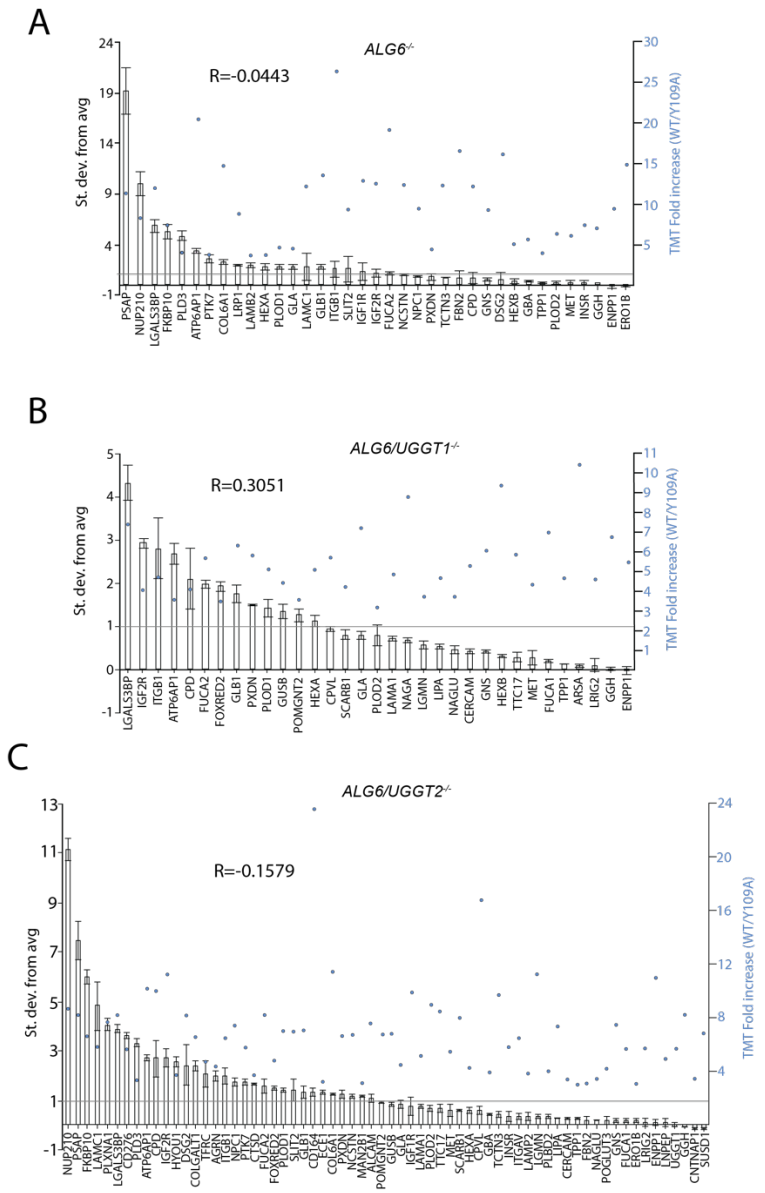

Figure 3.8 mRNA expression analysis of UGGT1 and UGGT2 substrates

(A) Reglucosylated substrates identified in *ALG6*$^{-/-}$ (A), *ALG6/UGGT1*$^{-/-}$ (B) and *ALG6/UGGT2*$^{-/-}$ (C) cells were compared to the average expression for the N-glycome in counts per million. The standard deviation from the average is plotted, with the error bars representing the standard deviation. Blue dots above each gene represent the level of fold increase (GST-CRT/GST-CRT-Y109A) found by TMT mass spectrometry (Figure 3.2 and 3.3). The Pearson's correlation coefficient (R) between the mRNA expression and TMT mass spectrometry fold increase is shown. Data is representative of three independent experiments.

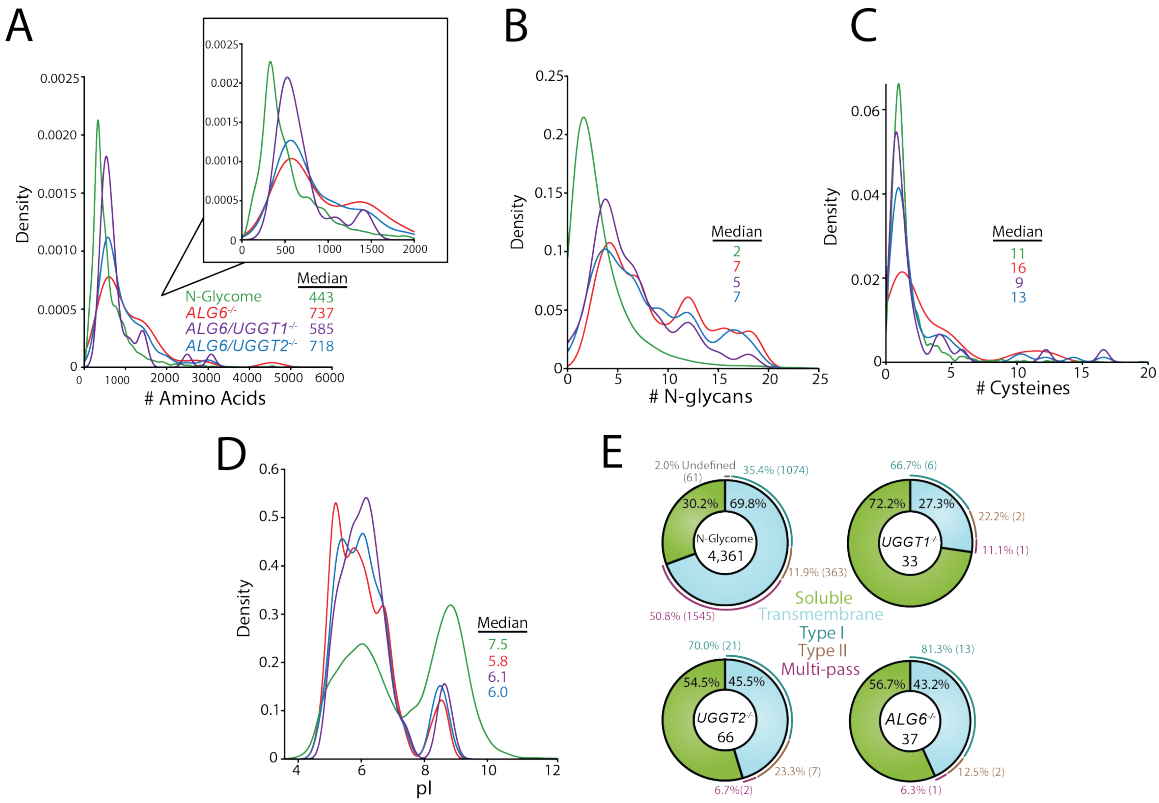

Figure 3.9 Analysis of substrates of the UGGTs and the N-glycome

(A) Amino acid lengths of each protein in the indicated datasets was visualized by density plot, with the total area under the curve integrated to 1. Amino acid number was obtained via Uniprot annotation. All density plots were generated using R and the ggplot package. (B) The number of N-glycans (B) or Cys residues (C) for each protein in the indicated datasets was visualized by density plot with the numbers determined using their Uniprot annotation. (D) The isoelectric point (pI) values for each protein in the indicated datasets was visualized by density plot. The pI values were obtained via ExPASy theoretical pI prediction. (E) The computationally predicted N-glycome and the indicated reglucosylation substrates were determined as either soluble or transmembrane using Uniprot annotations. The transmembrane portion of each dataset was then analyzed for type I, type II, or multi-pass topology using the associated Uniprot annotation. Proteins which were annotated by Uniprot as transmembrane but lacked topology information were labelled as undefined.

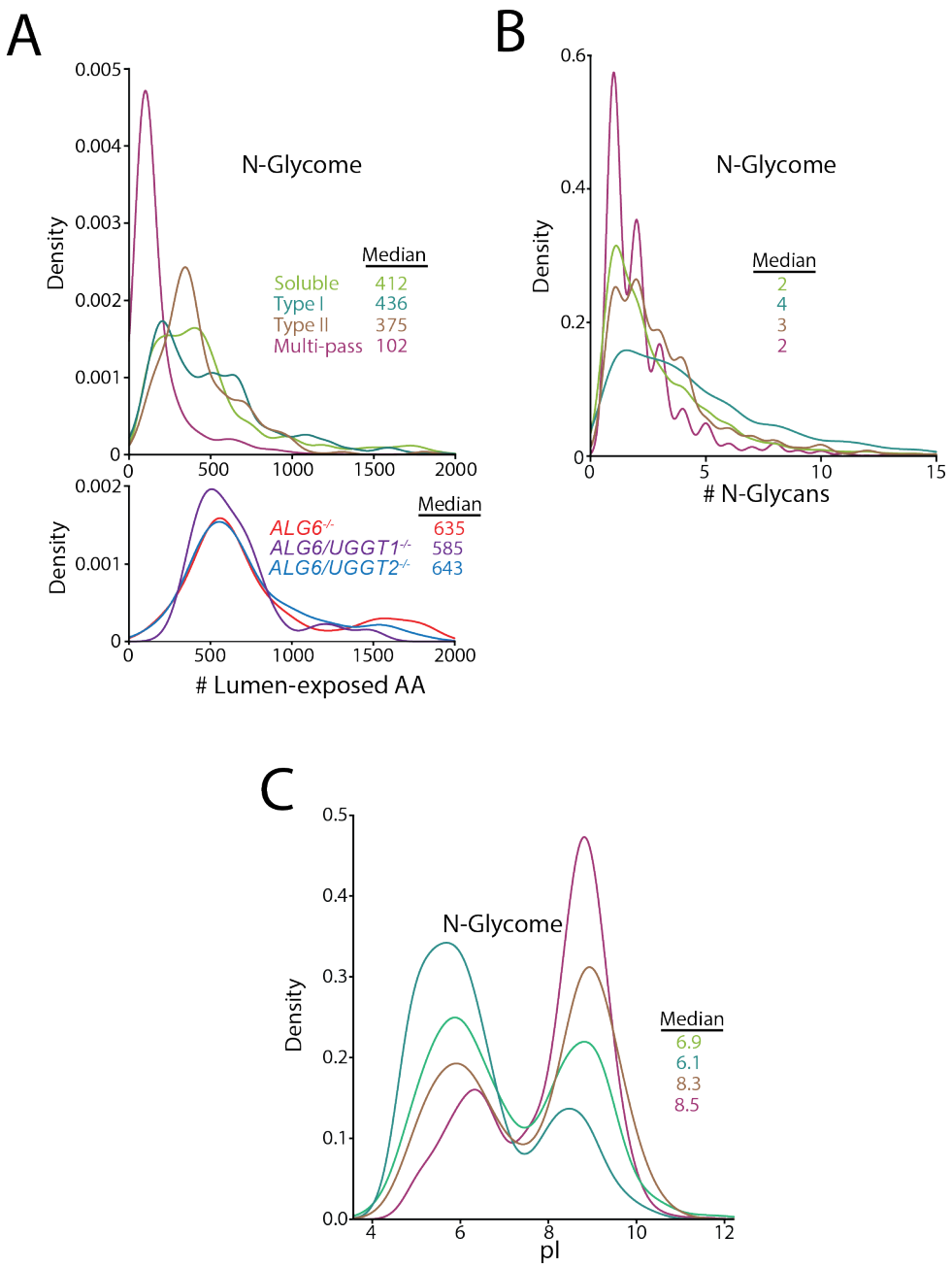

Figure 3.10 Analysis of N-glycome transmembrane proteins

(A) The computationally determined N-glycome was separated into soluble, type I, type II, and multi-pass transmembrane proteins using Uniprot annotations. Luminally exposed amino acids were computationally determined using Uniprot annotations for each subset of the N-glycome and each indicated reglucosylation substrate dataset. The resulting data was visualized by density plot. (B) The indicated N-glycome subsets were analyzed for N-glycan content using Uniprot annotation and visualized by density plot, as described. (C) The indicated N-glycome subsets were analyzed for predicted pI using ExPASy theoretical pI prediction and visualized by density plot.

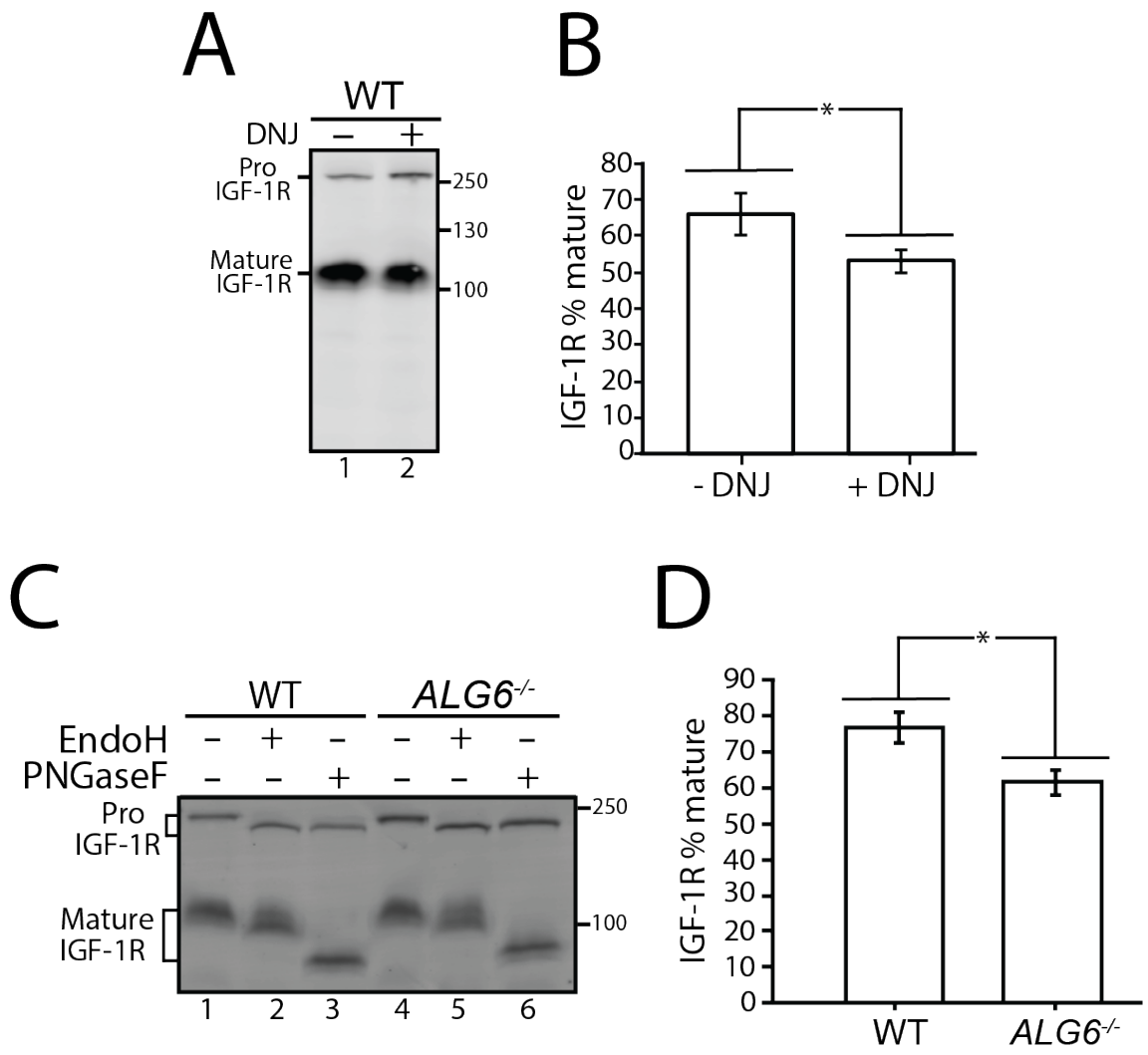

Figure 3.11 Calnexin/calreticulin cycle role for IGF-1R trafficking at steady state

(A) WT HEK293-EBNA1-6E cells treated without or with DNJ (500 μM) for 12-hr were lysed and WCL samples were resolved by reducing 9% SDS-PAGE and imaged by immunoblotting against IGF-1R. Data are representative of three independent experiments with quantification shown in B. Percent of IGF-1R mature was calculated by dividing the amount of mature protein by the total protein in each lane. Errors bars represent standard deviation. Asterisk denotes a p-value of less than 0.05 (C) The indicated cell lines were lysed in RIPA buffer. Samples were split evenly between non-treated and PNGaseF or EndoH treated. Samples were visualized by immunoblotting against IGF-1R and data are representative of three independent experiments with quantification displayed in D.

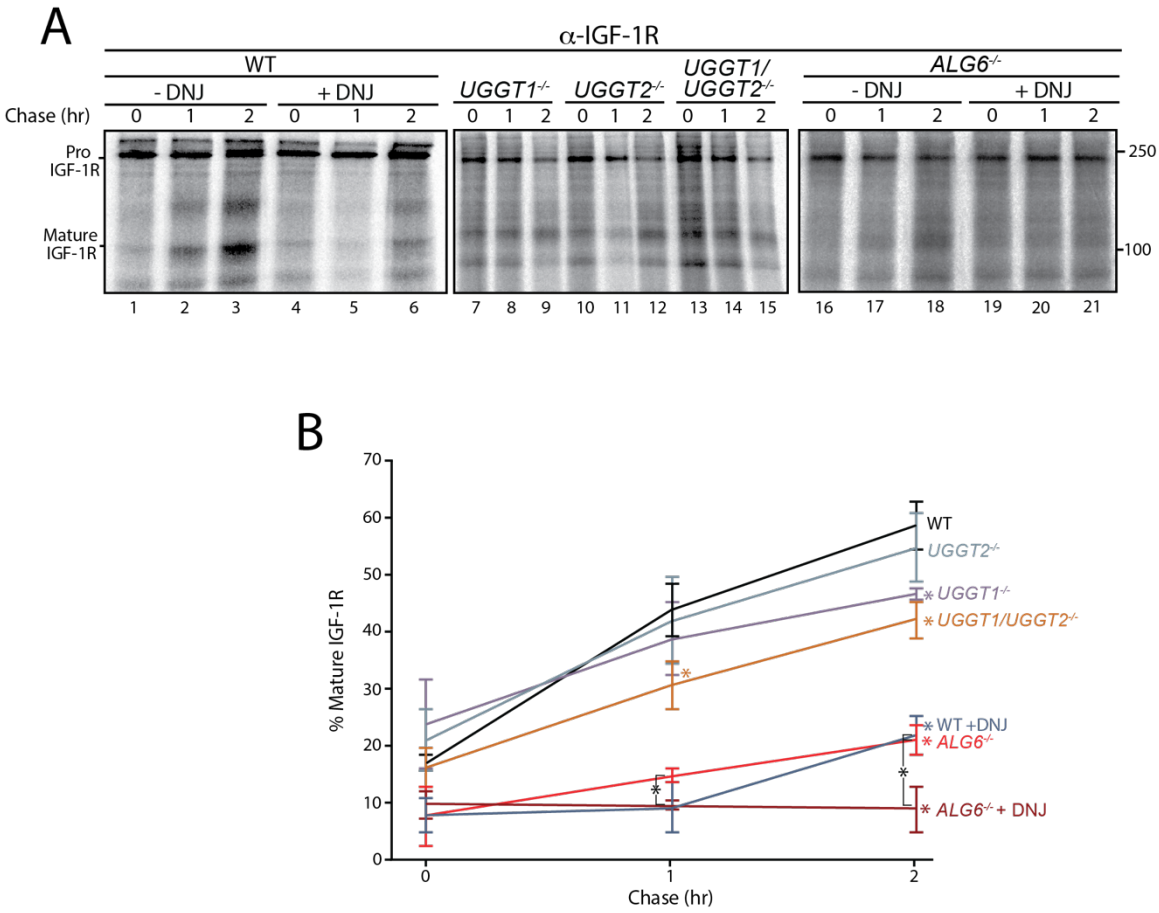

Figure 3.12 Mechanistic calnexin/calreticulin cycle role for IGF-1R trafficking

(A) Indicated cells were treated without or with DNJ, pulsed with [^{35}S]-Met/Cys for 1-hr and chased for the indicated times. Cells were lysed and samples were immunoprecipitated using anti-β IGF-1R antibody and resolved by reducing SDS-PAGE and imaged by autoradiography. Data are representative of three independent experiments with quantification shown in B.

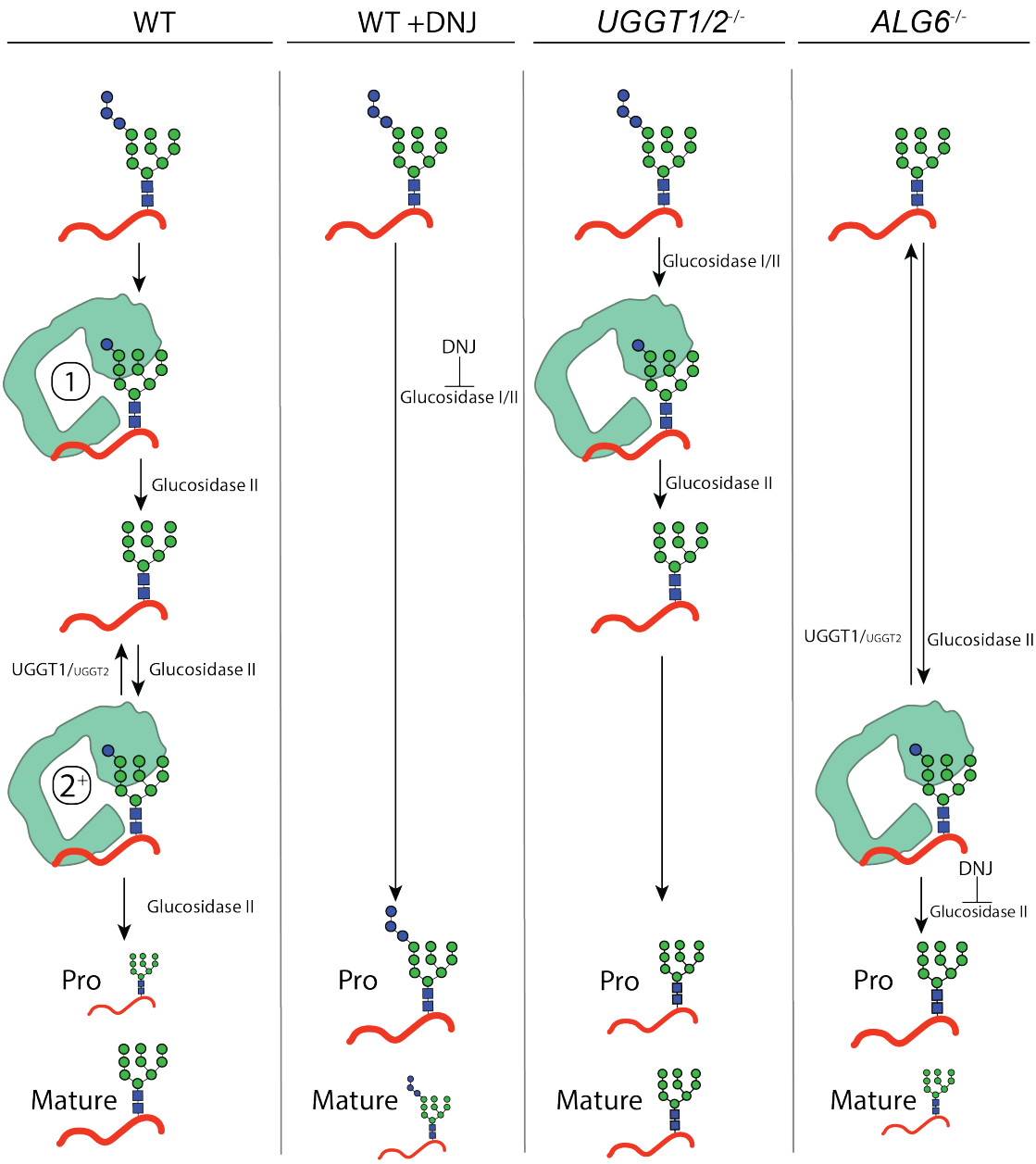

Figure 3.13 Model for IGF-1R engagement by the lectin chaperone cycle

In WT cells, N-glycans with three terminal glucoses are appended to IGF-1R. Trimming of two terminal glucoses by glucosidases I/II generates a monoglucosylated protein which supports an initial round of interaction with calreticulin (calnexin not shown, denoted by a 1). Trimming of the final glucose by glucosidase II yields a non-glucosylated N-glycan. If recognized as non-native primarily by UGGT1, and to a lesser extent UGGT2, IGF-1R may then be reglucosylated, supporting a second round of interaction with calreticulin (denoted by a 2^{+}). Multiple rounds of trimming, reglucosylation and binding to calnexin or calreticulin can occur until proper folding and trafficking. Under this system, IGF-1R is efficiently trafficked from the ER and mature

IGF-1R accumulates. When glucosidase I/II activity is inhibited by treatment with DNJ in WT cells, all rounds of binding to the lectin chaperones are ablated and IGF-1R is retained in the ER, yielding primarily pro IGF-1R. In *UGGT1/2*$^{-/-}$ cells, initial binding to calnexin or calreticulin directed by glucosidases I/II trimming is maintained but rebinding via reglucosylation does not occur. Under this system, IGF-1R is inefficiently trafficked from the ER. In *ALG6*$^{-/-}$ cells, N-glycans are transferred without glucoses, eliminating the initial round of binding to calnexin or calreticulin by glucosidases trimming. Only the second round of binding is supported by UGGT1, and to a lesser extend UGGT2, mediated reglucosylation. Upon treatment with DNJ, reglucosylated IGF-1R may persistently interact with the lectin chaperones resulting in ER retention.

CHAPTER 4

FUTURE DIRECTIONS

As the ER is the site of folding and maturation for a large percentage of the proteome, it is critical to understand the quality control processes these proteins are subjected to. Errors in ER quality control are implicated in numerous diseases (Hartl, 2017; Hebert & Molinari, 2007), and targeted approaches to improve these and other disease states with connections to ER quality control are of great interest. As demonstrated by work presented here (Adams, Ke, et al., 2019), it is highly difficult to predict what branch of ER quality control a given protein will engage, and as such must be experimentally determined. An understanding of what quality control branch a protein of interest is engaged by allows for the identification of potential sites of intervention regarding the maturation and trafficking pathway associated.

A large percentage of proteins targeted to the ER are N-glycosylated (Quellhorst et al., 1999; Zielinska et al., 2010) and therefore are potentially engaged by the N-glycan dependent calnexin/calreticulin cycle. This cycle is a major branch of ER quality control known to be involved in the folding and maturation pathway of multiple glycosylated substrates (Hebert et al., 1996; Molinari et al., 2005; Ferris et al., 2013; Adams, Oster, et al., 2019). UGGT1, and the lesser studied UGGT2, are key drivers of the calnexin/calreticulin cycle and essential for cellular proteostasis (Hebert et al., 1995; Molinari et al., 2005; Pearse et al., 2010; Tannous et al., 2015), but are poorly studied with regards to their endogenous roles as most work uses overexpressed proteins or *in vitro* approaches. The general goal of this work was therefore to experimentally identify and investigate the substrates which are engaged by UGGT1 and UGGT2 and are

therefore most dependent on the calnexin/calreticulin cycle. To this end, this work laid out an experimental workflow which identified a combined total of 71 substrates of UGGT1 and UGGT2, as well as a number of exciting discoveries regarding the substrate specificity of each glucosyltransferase and the role of the calnexin/calreticulin cycle for one UGGT1 substrate, IGF-1R (Chapter 3). This work has helped to open a number of possibilities for future research which will be described here.

Cell type and condition specific reglucosylation substrates

The work presented in Chapter 3 was conducted in HEK293-EBNA1-6E cells. The profile of substrates for UGGT1 and UGGT2 are likely, to an extent, specific to this cell line as each cell line exhibits unique protein expression profiles which are therefore available to be queried by ER quality control (Wilhelm et al., 2014; Ding et al., 2016). It would be of great interest to conduct the previously described workflow of UGGT1 and UGGT2 substrate identification in other cells lines including cell types with a high secretory load, such as liver or pancreatic cells. These cells express secretory pathway proteins at a high level and may therefore require increased chaperone engagement to maintain proteostasis in the crowded environment. Also, identifying UGGT1 and UGGT2 substrates in neuronal-derived cell lines would be of interest as congenital disorders of glycosylation involving *ALG6*, an early component of the calnexin/calreticulin cycle, often manifest as neurological issues (Morava et al., 2016). It is therefore possible that in some neuronal cells, specific proteins require high levels of engagement by UGGT1 or UGGT2 and the calnexin/calreticulin cycle, and therefore ablation of this cycle through an *ALG6* mutation could lead to disease onset due to misfolding of these proteins, similar

to the severe loss of trafficking of IGF-1R observed under *ALG6*$^{-/-}$ (Figure 3.12). Identifying these potential substrates may allow for improved explanation of disease phenotype and potential therapeutic approaches. The identification of substrates highly engaged by UGGT1 and UGGT2 in different cell types may therefore highlight glycoproteins which are most sensitive to off-pathway folding and therefore require high chaperone engagement for each cell type. This information may be valuable for understanding tissue-specific phenotypes of protein misfolding diseases and highlighting the potential benefit of targeting the calnexin/calreticulin cycle for treatment.

As UGGT1 and UGGT2 engage non-native proteins and promote chaperone interaction, their role may be especially important under conditions of proteostatic stress. Under such stress conditions, ER proteins may misfold at a high rate and therefore UGGT1 and UGGT2 would engage these substrates at an increased level. This would function to promote increased chaperone engagement, therefore retaining non-native proteins while decreasing potential aggregation and secretion of non-functional proteins. Numerous stress treatments, including oxidative stress, heat stress, and nutrient starvation could be investigated in order to observe both levels of reglucosylation of substrates as well as identify proteins which were not found to be substrates under homeostatic conditions but become substrates upon stress-induced misfolding. Oxidative stress may be of special interest as UGGT1 and UGGT2 tightly bind Sep15, a selenocysteine-containing protein which has a high redox potential but no well-defined role (Kasaikina et al., 2011; Y. Takeda et al., 2014). The entire pool of Sep15 is engaged by either UGGT1 or UGGT2 and may be involved in binding unpaired or mispaired Cys residues on UGGT1 and UGGT2 substrates, thereby directing substrate recognition. In addition to

oxidative stress, reglucosylation substrate identification could be conducted under Sep15 knockout, which may also help identify substrates directly engaged by this intriguing oxidoreductase.

The profile of engaged substrates may also be altered by the overexpression of a known substrate of UGGT1 or UGGT2, such as those identified in Chapter 3, or a known misfolding mutant of a substrate. This would mimic the environment of the ER in a disease state dominated by a misfolded protein which is heavily engaged by ER quality control. Under this condition, UGGT1 or UGGT2 may be titrated away from other substrates, resulting in decreased chaperone engagement and potentially leading to secondary effects. Potential substrates include β-hexosaminidase subunit β, arylsulfatase A, or α-N-acetylgalactosaminidase, all of which are UGGT2 substrates with mutations leading to lysosomal storage diseases (Mahuran, 1999; Cesani et al., 2016; Ferreira & Gahl, 2017), or collagen α-1(VI) chain, for which multiple known mutations exist which lead to decreased secretion and disease onset (Giusti et al., 2005; Lucioli et al., 2005). Observation of such an effect would demonstrate that the levels of UGGT1 or UGGT2 can become exhausted, and therefore under disease states due to a misfolded mutant, general ER proteostasis may be improved through increased expression of UGGT1 or UGGT2.

Glycan-specific reglucosylation

While the work presented here identified proteins which are substrates of UGGT1 and UGGT2 (Chapter 3), this work did not identify which specific glycan or glycans on the protein was reglucosylated as the entire protein was affinity purified by GST-CRT

prior to mass spectrometry preparation. As such, details regarding the sites of interaction for UGGT1 or UGGT2 and calnexin/calreticulin could not be elucidated. The identification of the specific glycans which are reglucosylated would be of great value for multiple reasons. As previous work has demonstrated, an understanding of the glycans which are engaged by calnexin/calreticulin can allow for an understanding of the folding trajectory and chaperone engagement pathway for a given substrate (Daniels et al., 2003). Such a detailed picture could be gained for numerous substrates by the identification of glycan-specific reglucosylation. This would allow for the identification of potential sites of alteration in order to modify the maturation pathway of a given substrate of interest, potentially to increase or decrease the level of ER quality control engagement. Glycan-specific information may also allow for an understanding of preferred binding sites for UGGT1 and UGGT2. Those sites could then be analyzed for sequence similarity and biophysical characteristics such as hydrophobicity, α-helix or β-sheet propensity, or the presence of unique amino acids such as Cys and Pro. This information would allow for a greatly improved ability to predict the substrates UGGT1 or UGGT2 engages as well as specific sites of engagement. Such prediction would be valuable as the chaperone engagement pathway of a substrate may be of interest though empirical data has not determined if UGGT1 or UGGT2 engage that substrate, due to potential experimental issues such as expression level in a given cell line or a disease relevant mutation.

In order to gain glycan-specific information, the existing protocol laid out in this work must be adapted (Figure 4.1). First, the steps of calreticulin affinity purification and trypsin-mediated peptide generation could be altered. The initial GST-CRT affinity purification post-lysis could be kept in order to enrich the pool of monoglucosylated

substrates. These proteins would then need to be eluted from GST-CRT using heat and a buffer which forms very small micelles, such as sodium deoxycholate, before alkylation and trypsinization. The sample would then be heated to denature trypsin and digested peptides would then be transferred to a filter-based approach as has been used for other applications regarding glycoprotein purifications (Zielinska et al., 2010). Peptides would then be placed into a second round of GST-CRT affinity purification, without beads, before being transferred to a 30 kDa filter. On this filter, GST-CRT bound peptides would be retained while non-bound peptides would flow through. The addition of the endoglycosidase PNGaseF would deglycosylated peptides, leading to a loss of interaction with GST-CRT and allowing those peptides to flow through the filter and be collected. As PNGaseF deglycosylates peptides via deamidation of the N-glycosylated Asn residue, sites of deamidated Asn could be identified by mass spectrometry, allowing for specific identification of N-glycosylated peptides. This would increase the confidence in identified peptides, as non-deamidated peptides, as well as peptides without an N-glycosylation motif, would have been non-specifically purified. The resulting peptides could be TMT-labeled, allowing for quantitative mass-spectrometry peptide identification. Conducting this work using the described *ALG6/UGGT1*$^{-/-}$ or *ALG6/UGGT2*$^{-/-}$ cell lines would allow for identification of UGGT1 or UGGT2 glycan-specific substrates. The described approach has been undertaken in the lab by Kevin Guay with promising initial results.

Live-cell reglucosylation imaging

While the work presented in Chapter 3, as well as the potential addition of glycan-specific reglucosylation data, lends great insight into lectin chaperone-mediated protein folding in the ER, there is still much spatio-temporal detail which remains unknown. Many questions could be addressed by gaining real-time and spatial data regarding reglucosylation. While the ER is composed of a singular lumen, much recent work has demonstrated that the ER is organized into subdomains through concentration of proteins with similar functions (Graham et al., 2019; Saito & Maeda, 2019; Nishimura & Stefan, 2020), as well as morphologically distinct regions (Borgese et al., 2006; Westrate et al., 2015). Therefore, a primary question is does reglucosylation in general occur within distinct subdomains of the ER, and do UGGT1 and UGGT2 specifically function in different subdomains of the ER. Data presented in Chapter 3 indicates UGGT2 more heavily engages lysosomal proteins than does UGGT1, suggesting the possibility that UGGT1 and UGGT2 function in distinct regions of the ER which are accumulated for different sets of substrates. Another unanswered question is does UGGT1 or UGGT2 function with temporal specificity, such as high activity at distinct phases during the cell cycle or peaks of activity under proteotoxic stress. In order to answer these questions, it is necessary to develop a live-cell reporter of reglucosylation. A FRET-based reporter has been used for live-cell imaging of O-glycosylation (Carrillo et al., 2011), suggesting this approach may be functional for imaging reglucosylation.

To this end, we have been working to generate a fluorescence lifetime imaging microscopy (FLIM)-FRET-based reporter for reglucosylation. The FLIM-FRET approach allows for improved quantification and noise reduction as compared to standard FRET (Becker, 2012). From N- to C-termini, the reporter will contain a signal sequence for ER

targeting, the fluorescent protein mTFP1 (the donor in the FLIM-FRET pair), a flexible linker of Gly and Ser residues, the lectin region of calreticulin, a second flexible linker, a glycosylated sequence which is recognized by either UGGT1 or UGGT2, the dark fluorescent protein ShadowG (the acceptor in the FLIM-FRET pair) (Murakoshi et al., 2015), and a KDEL sequence for ER-retention. Flexible linkers may be added or modified in order to achieve a function reporter. The reporter would function through reglucosylation of the glycosylated substrate, leading to the lectin domain of calreticulin binding to the monoglucosylated glycan. This would lead to a conformation shift of the reporter and bring the FLIM-FRET pairs into close contact. Under excitation, the fluorescence lifetime of mTFP1 would be measured. Increased reglucosylation of the reporter would lead to increased FLIM-FRET, yielding a decreased fluorescence lifetime of mTFP1 (Figure 4.2).

The described construct has been generated except for the glycosylated substrate region. The ideal substrate for either UGGT1 or UGGT2 may be identified through data generated by glycan-specific reglucosylation substrates, but may also be found through substrates presented in Chapter 3. For example, the UGGT2 substrate β-hexosaminidase subunit β contains four glycans (Figure 3.7). Using overexpression assays, these glycans could be individually mutated to determine the specific glycan, or glycans, that are reglucosylated by UGGT2. The region surrounding the glycan of interest could then be used as the substrate.

In addition to the described uses of this reporter in addressing spatio-temporal questions regarding reglucosylation, this reporter could also be useful as a high-throughput assay for the development of inhibitors or activators of UGGT1 or UGGT2. A

library of compounds could be investigated for the ability to decrease FLIM-FRET in a multi-well plate format before further investigation of compounds of interest using more direct techniques such as those used in Figure 3.6 and 3.7. These compounds could then be used both for targeted cell biology studies regarding the function of UGGT1 and UGGT2 as well as investigating potential therapeutic function for diseases relating to ER quality control, such as cystic fibrosis or lysosomal storage diseases (Younger et al., 2006; Dersh et al., 2016). As described in Chapter 3, IGF-1R is overexpressed in many cancers and, under UGGT1 and UGGT2 knockout, is trafficked to the plasma membrane at a decreased rate (Figure 3.12). Therefore, a potential inhibitor of UGGT1 and UGGT2 may function as a treatment against cancer. As many substrates are engaged by UGGT1 and UGGT2, such therapeutic intervention would be of interest in a number of related diseases. Additionally, an inhibitor of either UGGT1 or UGGT2 would be very useful for basic cell biology questions related to the calnexin/calreticulin process, as the function of UGGT1 or UGGT2 could be specifically inhibited without the requirement of generating a knockout in each cell line of interest.

Role of UGGT2 in lysosomal function

A significant result presented in Chapter 3 is that UGGT2 has reglucosylation activity and is preferential towards lysosomal proteins (Figure 3.5). The majority of the proteins are enzymes which function in the degradation process of carbohydrates or lipids and, as previously discussed, can lead to lysosomal storage diseases when mutated. Therefore, it would be of great interest to understand the effect of UGGT2 on lysosomal function and the potential role it could play in the suppression of disease states through

promotion of ER to lysosome trafficking. As some mutations in such enzymes which manifest as lysosomal storage diseases do in fact possess partial activity (Leinekugel et al., 1992; Wens et al., 2012), it may be therapeutic to promote increased trafficking of these mutants to the lysosome. Additionally, impaired UGGT2 activity may lead to a substrate specific effect on trafficking to the lysosome. As presented in Figure 4.3, ablation of the ability of β-hexosaminidase subunit β to enter the calnexin/calreticulin pathway through treatment with the glucosidase inhibitor DNJ lead to a decrease in mature β-hexosaminidase subunit β, suggesting UGGT2 is necessary for efficient trafficking to the lysosome. Therefore, UGGT2 may play a role in lysosome function through promotion of productive folding and trafficking of lysosomal proteins. However, much work is needed to clarify the function of UGGT2 for the many functions the lysosome possesses, especially with regards to the substrates highlighted in Chapter 3.

One approach to elucidate the role of UGGT2 in lysosome function would be to study the metabolomic profile of the lysosome under *UGGT2*$^{-/-}$. If a lysosomal enzyme does not fold properly due to a loss of UGGT2 mediated reglucosylation, poor trafficking to the lysosome may result. As such, the metabolic pathway catalyzed by the enzyme of interest would be altered and metabolites may accumulate. The identification of altered levels of metabolites would highlight pathways which are impaired and therefore require UGGT2 function. This approach would require the purification of lysosomes from cells before lysis and sample preparation. Previous work has demonstrated the ability to purify lysosomes using an overexpressed triple HA-tagged TMEM192 construct (Abu-Remaileh et al., 2017). TMEM192 is a multi-pass lysosomal membrane protein, and the overexpression of this protein with a cytosolically exposed triple HA-tag allows for the

purification of lysosomes from cell lysate. Samples can then be sent to mass spectrometry facilities with capabilities to conduct metabolomics. Amount of total lysosomal enzymes present in both the lysosomal enriched and de-enriched fractions could be analyzed by quantitative mass spectrometry or targeted western blots to control for total number of cells and expression of lysosomal enzymes.

However, many lysosomal enzymes at significantly reduced concentrations compared to WT levels can still generate similar concentrations of metabolites. As such, metabolomics may not effectively identify alterations in enzyme concentrations due to trafficking defects in *UGGT2*$^{-/-}$ cells, though the depletion of multiple enzymes may lead to additive effects which could improve the utility of a metabolomics approach. If a metabolomics approach is not successful, a more targeted approach using activity assays designed against single proteins may be used, as are available for proteins such as β-hexosaminidase subunit β and arylsulfatase A (Matzner et al., 2007; Thelen et al., 2017). In this case, purified lysosomes would also be necessary in order to concentrate the sample for the enzymes of interest. While the described Tmem192-HA approach may be used, an alternative approach would be to isolate lysosomes using a magnetic purification approach which exposes cells for 24 hr to dextran-coated 10-nm iron oxide particles. These particles are then endocytosed and delivered to the lysosome, where they accumulate and, post cell lysis, allow lysosomes to be purified magnetically (Thelen et al., 2017).

The role of UGGT2 regarding lysosome function may also be investigated through microscopy, as a decrease in enzyme localization at the lysosome may lead to accumulation of substrates and lead to lysosomal swelling. Lysosome size can then be

analyzed via a lysosome targeted fluorescent antibody. Such an approach has been used to investigate the effect of chaperone overexpression in primary fibroblasts from patients with multiple types of lysosomal storage diseases (Kirkegaard et al., 2016). These approaches may also be combined to investigate possible therapeutic treatments involving UGGT2 for cell lines expressing mutant lysosomal enzymes.

Role of the calnexin/calreticulin cycle in proteome trafficking

Pulse-chase data presented in Figure 3.12 demonstrated that endogenous IGF-1R exhibits decreased trafficking under multiple conditions which alter the calnexin/calreticulin cycle. These data gave a detailed picture of the role of the calnexin/calreticulin cycle for IGF-1R, but a proteomic level of detail would be a great improvement over attempting numerous targeted pulse-chases for single proteins, especially as many proteins will not have antibodies which function for immunoprecipitations. While a steady-state approach would allow for more functional antibodies as immunoprecipitations would not be necessary, a pulse-chase type approach would give much more information regarding trafficking kinetics than a steady-state approach. The level of detail between the two general approaches was demonstrated in Chapter 3, where IGF-1R was found to display a 19% decrease in trafficking from the ER under DNJ treatment at steady-state (Figure 3.11), while a 63% reduction after 2 hr was observed under pulse-chase (Figure 3.12). As such, we have begun developing an L-Azidohomoalanine (AHA) based approach to investigate the role of the calnexin/calreticulin cycle for the trafficking of clients of the secretory pathway.

AHA is a methionine analog that is metabolically incorporated in the place of methionine and possesses an azido moiety. This azido moiety can then be used in copper-catalyzed alkyne click chemistry reactions to specifically label AHA with desired groups such as biotin which can be used to purify labeled proteins. AHA labelling can occur for different amounts of time, commonly 1-2 hours, then chased with complete media in order to generate a pulse-chase under which the trafficking of the labeled population can be tracked. Cell membrane impermeable click-reactive biotin may be used in order to label only proteins the region of interest, such as secreted, plasma membrane, ER, or lysosome. Purified proteins from a given region can then be prepared for quantitative mass spectrometry. By comparing the amount of proteins in the different localizations, the effect of a given knock-out or inhibitor on proteomic-scale trafficking can be examined. Such conditions may include *ALG6*$^{-/-}$, *UGGT1*$^{-/-}$, *UGGT2*$^{-/-}$, *CANX*$^{-/-}$, *CALR*$^{-/-}$, and combinations, as well as DNJ or potential UGGT1 or UGGT2 inhibitors. The ability to understand the role of the calnexin/calreticulin cycle in the trafficking of proteins at a proteomic level would allow for an unprecedented amount of information regarding the role of this cycle and how it can be modified to alter the folding and trafficking pathway of any identified protein of interest.

Summary

While ER quality control, and specifically the calnexin/calreticulin cycle, has been studied for many years which has greatly expanded the understanding of the quality control and chaperone systems which engage nascent secretory pathway proteins, this field has not been well examined using proteomic and computational approaches as

presented here. By applying a proteomics-based approach, many new avenues of research are opening which may allow for a much improved understanding of the role and importance of this process at a systematic level. Some of these avenues, including identifying reglucosylation substrates in various cell lines and stress conditions as well as glycan-specific reglucosylation, are likely to be productive in a short timeline and will be very exciting expansions of the current field. With sufficient study, the ER quality control processes being discovered may be of great use for both basic cell biology and the development of treatments for diseases with which they are involved.

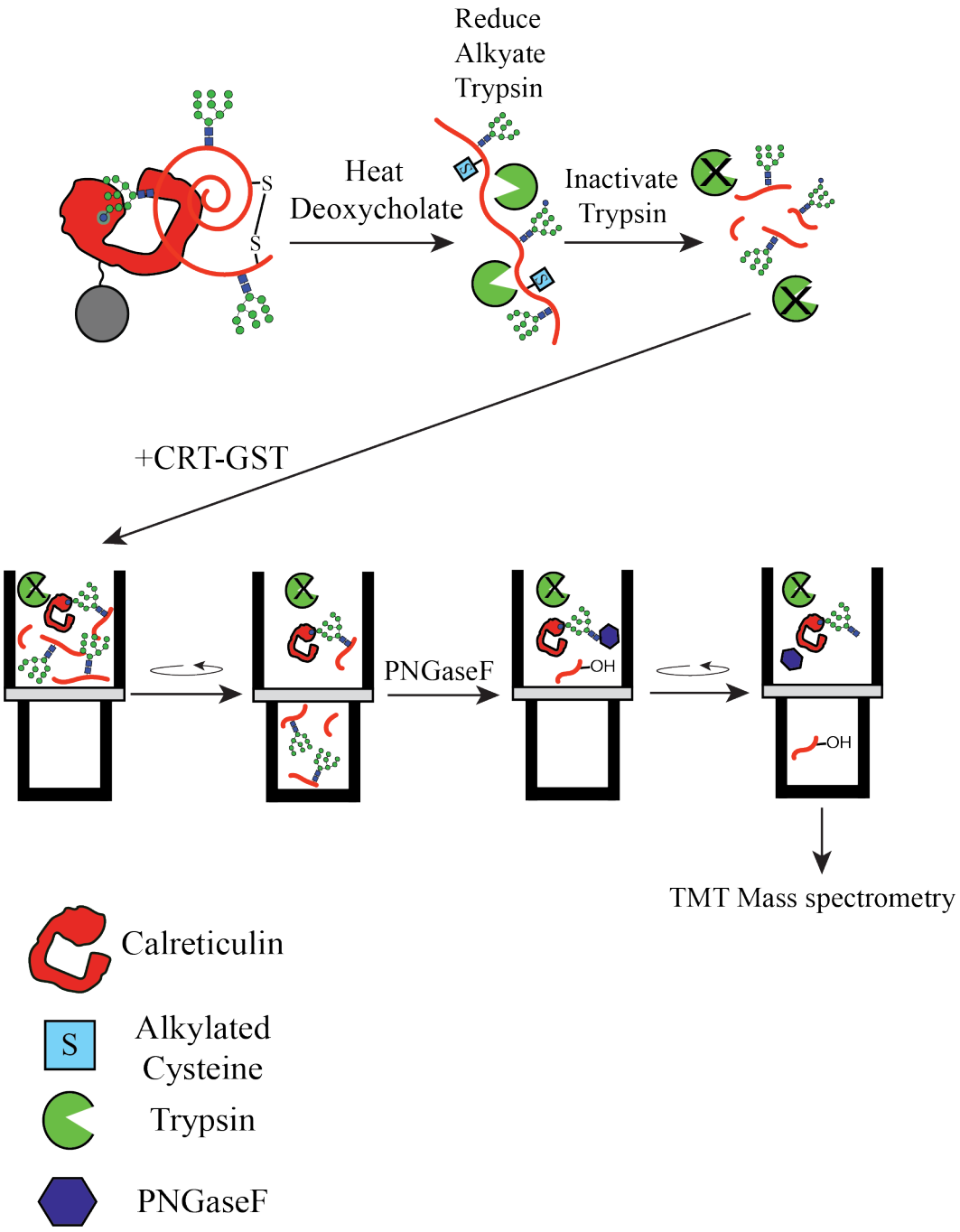

Figure 4.1 Glycan-specific reglucosylation substrate identification workflow
Reglucosylated proteins from cell lysates would first be enriched using the previously described on-bead GST-CRT affinity purification approach (Figure 3.1B). Protein would then be eluted from GST-CRT using heat and deoxycholate. The sample would then be reduced with DTT, alkylated using iodoacetamide, and trypsinized. Trypsin would then be heat inactivated and peptides incubated with GST-CRT in solution to bind monoglucosylated peptides. The sample would then be placed in a 30 kDa spin filter and unbound peptides would flow through the filter while GST-CRT bound peptides would be retained. GST-CRT bound peptides would then be eluted through deglycosylation

using PNGaseF. Asn residues modified with N-glycans would be deamidated (denoted by OH group) after PNGaseF mediated deglycosylation. Samples would then be labeled with TMT labels and analyzed by LC-MS/MS. Figure provided by Kevin Guay.

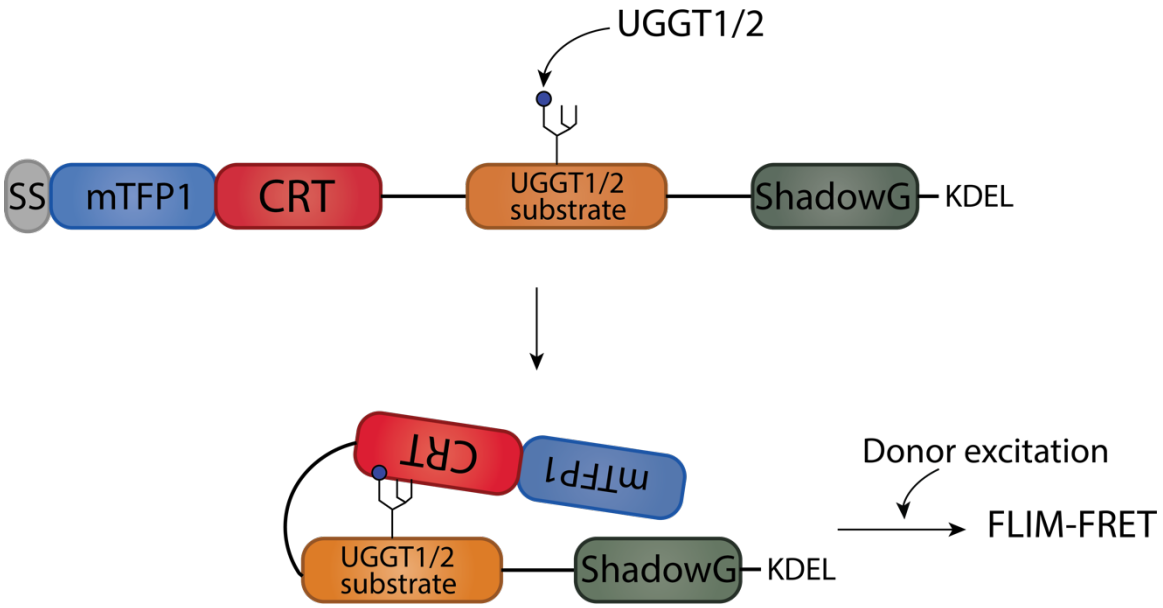

Figure 4.2 Live-cell reglucosylation FLIM-FRET reporter
The current reporter design and proposed method of action is depicted. The reporter would be targeted to the ER via a signal sequence (grey), which would be cleaved after translocation. The FLIM-FRET pairs mTFP1 (blue) and ShadowG (dark green) are placed on the N- and C-termini, respectively. The lectin domain of calreticulin (CRT) (red) immediately C-terminal to mTFP1 followed by the glycosylated UGGT1/2 substrate (orange). Flexible linker regions are shown by black lines between domains. Upon reglucosylation of the glycosylated substrate by UGGT1/2, the lectin domain of CRT would bind to the monoglucosylated glycan, bringing mTFP1 and ShadowG in sufficiently close proximity to undergo FRET upon donor excitation. FRET between mTFP1 and ShadowG would be measured by the fluorescence lifetime microscopy (FLIM) of mTFP1.

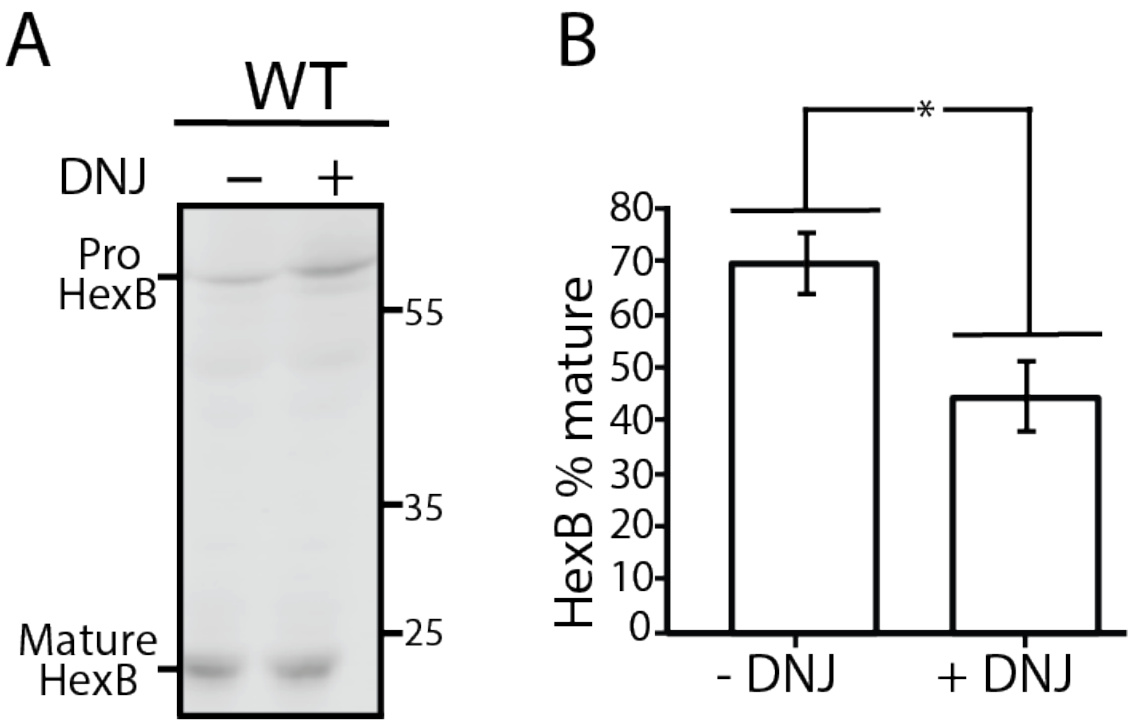

Figure 4.3 Role of the calnexin/calreticulin pathway for β-hexosaminidase subunit β trafficking

WT HEK293-EBNA1-6E cells treated without or with DNJ (500 μM) for 12-hr were lysed and WCL samples were resolved by reducing 9% SDS-PAGE and imaged by immunoblotting against IGF-1R. Data are representative of three independent experiments with quantification shown in B.

BIBLIOGRAPHY

Abu-Remaileh, M., Wyant, G. A., Kim, C., Laqtom, N. N., Abbasi, M., Chan, S. H., Freinkman, E., & Sabatini, D. M. (2017). Lysosomal metabolomics reveals V-ATPase- and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science (New York, N.Y.)*, *358*(6364), 807–813. PubMed. https://doi.org/10.1126/science.aan6298

Adams, B. M., Ke, H., Gierash, L. M., Gershenson, A., & Hebert, D. N. (2019). Proper secretion of the serpin antithrombin relies strictly on thiol-dependent quality control. *Journal of Biological Chemistry*, *294*(50), 18992–19011. https://doi.org/10.1074/jbc.RA119.010450

Adams, B. M., Oster, M. E., & Hebert, D. N. (2019). Protein Quality Control in the Endoplasmic Reticulum. *The Protein Journal*, *38*(3). https://doi.org/10.1007/s10930-019-09831-w.

Aebi, M. (2013). N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta - Molecular Cell Reasarch*, *1833*(11), 2430–2437. https://doi.org/10.1016/j.bbamcr.2013.04.001

Aikawa, J., Matsuo, I., & Ito, Y. (2012). In vitro mannose trimming property of human ER α-1,2 mannosidase I. *Glycoconjugate Journal*, *29*(1), 35–45. https://doi.org/10.1007/s10719-011-9362-1

Amin-Wetzel, N., Saunders, R. A., Kamphius, M. J., Rato, C., Preissler, S., Harding, H. P., & Ron, D. (2017). A J-Protein Co-chaperone Recruits BiP to Monomerize IRE1 and Repress the Unfolded Protein Response. *Cell*, *171*(7), 1625–1637.

Andrews, S. (n.d.). *FastQC: a quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Anelli, T., Ceppi, S., Bergamelli, L., Cortini, M., Masciarelli, S., Valetti, C., & Sitia, R. (2007). Sequential steps and checkpoints in the early exocytic compartment during secretory IgM biogenesis. *The European Molecular Biology Organization Journal*, *26*(19), 4177–4188.

Anelli, T., Sannino, S., & Sitia, R. (2015). Proteostasis and "redoxtasis" in the secretory pathway: Tales of tails from ERp44 and immunoglobulins. *Free Radical Biology and Medicine*, *83*, 323–330.

Anwesha, S., Chen, A. J., Nakayasu, E. S., Lazar, C. S., Zbornik, E. A., Worby, C. A., Koller, A., & Mattoo, S. (2015). A novel link between Fic mediated adenylation and the unfolded protein response. *Journal of Biological Chemistry*, *290*(13), 8482–8499.

Appenzeller-Herzog, C., Riemer, J., Christensen, B., Sørensen, E. S., & Ellgaard, L. (2008). A novel disulphide switch mechanism in Ero1α balances ER oxidation in human cells. *The European Molecular Biology Organization Journal*, *27*(22), 2977–2987.

Apweiler, R., Hermjakob, H., & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta*, *1473*(1), 4–8.

Araki, K., & Nagata, K. (2011). Protein Folding and Quality Control in the ER. *Cold Spring Harbor Perspectives in Biology*, *3*(11), a007526.

Arnold, S. M., Fessler, L. I., Fessler, J. H., & Kaufman, R. J. (2000). Two Homologues Encoding Human UDP-Glucose:Glycoprotein Glucosyltransferase Differ in mRNA

Expression and Enzymatic Activity. *Biochemistry*, *39*(9), 2149–2163. https://doi.org/10.1021/bi9916473

Arnold, S. M., & Kaufman, R. J. (2003). The Noncatalytic Portion of Human UDP-glucose: Glycoprotein Glucosyltransferase I Confers UDP-glucose Binding and Transferase Function to the Catalytic Domain. *Journal of Biological Chemistry*, *278*(44), 43320–43328. https://doi.org/10.1074/jbc.M305800200

Bajaj, L., Sharma, J., di Ronza, A., Zhang, P., Eblimit, A., Pal, R., Roman, D., Collette, J. R., Booth, C., Chang, K. T., Sifers, R. N., Jung, S. Y., Weimer, J. M., Chen, R., Schekman, R. W., & Sardiello, M. (2020). A CLN6-CLN8 complex recruits lysosomal enzymes at the ER for Golgi transfer. *The Journal of Clinical Investigation*, *130*(8), 4118–4132. https://doi.org/10.1172/JCI130955

Baksh, S., & Michalak, M. (1991). Expression of calreticulin in Escherichia coli and identification of its Ca2+ binding domains. *The Journal of Biological Chemistry*, *266*(32), 21458–21465.

Balchin, D., Hayer-Hartl, M., & Hartl, F. U. (2020). Recent advances in understanding catalysis of protein folding by molecular chaperones. *FEBS Letters*, *594*(17), 2770–2781. https://doi.org/10.1002/1873-3468.13844

Balchin, D., Hayer-Hartl, M., & Hartl, U. F. (2016). In vivo aspects of protein folding and quality control. *Science*, *353*(6294), aac4354. https://doi.org/10.1126/science.aac4354

Barlowe, C., & Helenius, A. (2016). Cargo Capture and Bulk Flow in the Early Secretory Pathway. *Annual Review of Cell and Developmental Biology*, *32*, 197–222.

Barlowe, C. K., & Miller, E. A. (2013). Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. *Genetics*, *193*(2), 383–410.

Beck, M., & Hurt, E. (2016). The nuclear pore complex: Understanding its function through structural insight. *Nature Reviews Molecular Cell Biology*, *18*(2), 73–89. https://doi.org/10.1038/nrm.2016.147

Becker, W. (2012). Fluorescence lifetime imaging – techniques and applications. *Journal of Microscopy*, *247*(2), 119–136.

Behnke, J., Feige, M. J., & Hendershot, L. M. (2015). BiP and its nucleotide exchange factors Grp170 and Sil1: Mechanisms of action and biological functions. *Journal of Molecular Biology*, *427*(7), 1589–1608.

Behnke, J., Mann, M. J., Scruggs, F.-L., Feige, M. J., & Hendershot, L. M. (2016). Members of the Hsp70 family recognize distinct types of sequences to execute ER quality control. *Molecular Cell*, *63*(5), 739–752.

Benyair, R., Ogen-Shtern, N., Mazkereth, N., Shai, B., Ehrlich, M., & Lederkremer, G. Z. (2015). Mammalian ER mannosidase I resides in quality control vesicles where it encounters its glycoprotein substrates. *Molecular Biology of the Cell*, *26*(2), 172–184.

Bernasconi, R., Pertel, T., Luban, J., & Molinari, M. (2008). A dual task for the XBP1-responsive OS-9 variants in the mammalian endoplasmic reticulum. *The Journal of Biological Chemistry*, *283*(24), 16446–16454.

Bhamidipati, A., Denic, V., Quan, E. M., & Weissman, J. S. (2005). Exploration of the Topological Requirements of ERAD Identifies Yos9p as a Lectin Sensor of Misfolded Glycoproteins in the ER Lumen. *Molecular Cell*, *19*(6), 741–751.

Blobel, G., & Dobberstein, B. (1975a). Transfer of protein across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *Journal of Cell Biology*, *67*(3), 852–862.

Blobel, G., & Dobberstein, B. (1975b). Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *Journal of Cell Biology*, *67*(3), 835–851.

Blobel, G., & Sabatini, D. D. (1971). Ribosome-membrane interaction in eukaryotic cells. *Biomembranes*, *2*, 193–195.

Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F., & Gething, M.-J. (1993). Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell*, *75*(4), 717–728. https://doi.org/10.1016/0092-8674(93)90492-9

Boisramé, A., Kabani, M., Beckerich, J.-M., Hartmann, E., & Claude, G. (1998). Interaction of Kar2p and Sls1p is required for efficient co-translational translocation of secreted proteins in the yeast Yarrowia lipolytica. *Journal of Biological Chemistry*, *273*(47), 30903–30908.

Borgese, N., Francolini, M., & Snapp, E. (2006). Endoplasmic reticulum architecture: Structures in flux. *Current Opinion in Cell Biology*, *18*(4), 358–364. PubMed. https://doi.org/10.1016/j.ceb.2006.06.008

Breitling, J., & Aebi, M. (2013). N-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, *5*(8), a013359.

Brodsky, J. L. (2012). Cleaning Up: Endoplasmic Reticulum Associated Degradation to the Rescue. *Cell*, *151*(6), 1163–1167.

Brodsky, J. L., Werner, E. D., Dubas, M. E., Goeckeler, J. L., Kruse, K. B., & Ardythe, M. A. (1999). The requirement for molecular chaperones during endoplasmic reticulum-associated protein degradation demonstrates that protein export and import are mechanistically distinct. *Journal of Biological Chemistry*, *274*(6), 3453–3460.

Bulleid, N. J., & Ellgaard, L. (2011). Multiple ways to make disulfides. *Trends in Biochemical Sciences*, *36*(9), 485–492.

Bulleid, N. J., & Freedman, R. B. (1988). Defective co-translational formation of disulphide bonds in protein disulphide-isomerase-deficient microsomes. *Nature*, *335*(6191), 649–651.

Cacan, R., Duvet, S., Labiau, O., Verbert, A., & Krag, S. S. (2001). Monoglucosylated oligomannosides are released during the degradation process of newly synthesized glycoproteins. *Journal of Biological Chemistry*, *276*(25), 22307–22312. https://doi.org/10.1074/jbc.M101077200

Cai, G., Salonikidis, P. S., Fei, J., Schwarz, W., Schülein, R., Reutter, W., & Fan, H. (2005). The role of N-glycosylation in the stability, trafficking and GABA-uptake of GABA-transporter 1. *The Federation of European Biochemical Sciences Journal*, *272*(7), 1625–1638.

Calfon, M., Zeng, H., Urano, F., Till, J. H., Hubbard, S. R., Harding, H. P., Clark, S. G., & Ron, D. (2002). IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP1 mRNA. *Nature*, *415*(6867), 92–96.

Calì, T., Galli, C., Olivari, S., & Molinari, M. (2008). Segregation and rapid turnover of EDEM1 by an autophagy-like mechanism modulates standard ERAD and folding activities. *Biochemical and Biophysical Research Communications*, *371*(3), 405–410.

Cannon, K. S., Hebert, D. N., & Helenius, A. (1996). Glycan dependent and independent association of Vesicular Stomatitis Virus G Protein with Calnexin. *Journal of Biological Chemistry*, *271*(24), 14280–14284.

Caramelo, J. J., Castro, O. A., Prat-Gay, G. de, & Parodi, A. J. (2004). The endoplasmic reticulum glucosyltransferase recognizes nearly native glycoprotein folding intermediates. *Journal of Biological Chemistry*, *279*(44), 46280–46285. https://doi.org/10.1074/jbc.M408404200

Caramelo, J. J., & Parodi, A. J. (2015). A sweet code for glycoprotein folding. *FEBS Letters*, *589*(22), 3379–3387. https://doi.org/10.1016/j.febslet.2015.07.021

Carrillo, L. D., Froemming, J. A., & Mahal, L. K. (2011). Targeted in vivo O-GlcNAc sensors reveal discrete compartment-specific dynamics during signal transduction. *The*

Journal of Biological Chemistry, *286*(8), 6650–6658. PubMed. https://doi.org/10.1074/jbc.M110.191627

Cesani, M., Lorioli, L., Grossi, S., Amico, G., Fumagalli, F., Spiga, I., Filocamo, M., & Biffi, A. (2016). Mutation Update of ARSA and PSAP Genes Causing Metachromatic Leukodystrophy. *Human Mutation*, *37*(1), 16–27. https://doi.org/10.1002/humu.22919

Chandrasekhar, K., Ke, H., Wang, N., Goodwin, T., Gierash, L. M., Gershenson, A., & Hebert, D. N. (2016). Cellular folding pathway of a metastable serpin. *Proceedings of the National Acadamy of Sciences of the United States of America*, *113*(23), 6484–6489.

Chen, W., Helenius, J., Braakman, I., & Helenius, A. (1995). Cotranslational folding and calnexin binding during glycoprotein synthesis. *Proceedings of the National Acadamy of Sciences of the United States of America*, *92*(14), 6229–6233.

Chen, X., Easton, D., Oh, H.-J., Lee-Yoon, D.-S., Liu, X., & Subjeck, J. (1996). The 170 kDa glucose regulated stress protein is a large HSP70- HSP110-like protein of the endoplasmic reticulum. *Federation of European Biochemical Sciences Letters*, *380*(1–2), 68–72.

Cheng, M. Y., Hartl, U. F., Martin, J., Pollock, R. A., Kalousek, F., Neupert, W., Hallberg, E. M., Hallberg, R. L., & Horwich, A. L. (1989). Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature*, *337*(6208), 620–625.

Cherepanova, N. A., Shrimal, S., & Gilmore, R. (2016). N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Current Opinion in Cell Biology*, *41*, 57–65.

Cherepanova, N. A., Venev, S. V., Leszyk, J. D., Shaffer, S. A., & Gilmore, R. (2019). Quantitative proteomics reveals new classes of STT3A- and STT3B-dependent N-glycosylation sites. *Journal of Cell Biology*, *218*(8), 2782–2796. https://doi.org/10.1083/jcb.201904004

Chng, W. J., Gualberto, A., & Fonseca, R. (2006). IGF-1R is overexpressed in poor-prognostic subtypes of multiple myeloma. *Leukemia*, *20*(1), 174–176. https://doi.org/10.1038/sj.leu.2403997

Christianson, J. C., Shaler, T. A., Tyler, R. E., & Kopito, R. R. (2009). OS-9 and GRP94 deliver mutant alpha1-antitrypsin to the Hrd1-SEL1L ubiquitin ligase complex for ERAD. *Nature Cell Biology*, *10*(3), 272–282.

Chung, K. T., Shen, Y., & Hendershot, L. M. (2002). BAP, a mammalian BiP-associated protein, is a nucleotide exchange factor that regulates the ATPase activity of BiP. *Journal of Biological Chemistry*, *277*(49), 47557–47563.

Coe, H., Jung, J., Groenendyk, J., Prins, D., & Michalak, M. (2010). ERp57 modulates STAT3 signaling from the lumen of the endoplasmic reticulum. *Journal of Biological Chemistry*, *285*(9), 6725–6738.

Cormier, J. H., Tamura, T., Sunryd, J. C., & Hebert, D. N. (2010). EDEM1 recognition and delivery of misfolded proteins to the SEL1L-containing ERAD complex. *Molecular Cell*, *34*(5), 627–633.

Corral, J., de la Morena-Barrio, M. E., & Vicente, V. (2018). The genetics of antithrombin. *Thrombosis Research*, *169*, 23–29.

Culyba, E. K., Price, J. L., Hanson, S. R., Dhar, A., Wong, C.-H., Gruebele, M., Powers, E. T., & Kelly, J. W. (2011). Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science*, *331*(6017), 571–575.

Cunnea, P. M., Miranda-Vizuete, A., Bertoli, G., Simmen, T., Damdimopoulos, A. E., Hermann, S., Leinonen, S., Huikko, M. P., Gustafsson, J.-Å., Sitia, R., & Spyrou, G. (2003). ERdj5, an Endoplasmic Reticulum (ER)-resident Protein Containing DnaJ and Thioredoxin Domains, Is Expressed in Secretory Cells or following ER Stress. *Journal of Biological Chemistry*, *278*(2), 1059–1066.

Daniels, R., Kurowski, B., Johnson, A. E., & Hebert, D. N. (2003). N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Molecular Cell*, *11*(1), 79–90. https://doi.org/10.1016/s1097-2765(02)00821-3

Davies, M. J., & Lomas, D. A. (2008). The molecular aetiology of the serpinopathies. *The International Journal of Biochemistry and Cell Biology*, *40*(6–7), 1273–1286.

Dell'Angelica, E. C., & Payne, G. S. (2001). Intracellular Cycling of Lysosomal Enzyme Receptors: Cytoplasmic Tails' Tales. *Cell*, *106*(4), 395–398. https://doi.org/10.1016/S0092-8674(01)00470-6

Dementiev, A., Dobó, J., & Gettins, P. G. (2006). Active site distortion is sufficient for proteinase inhibition by serpins: Structure of the covalent complex of alpha1-proteinase inhibitor with porcine pancreatic elastase. *Journal of Biological Chemistry*, *281*(6), 3452–3457.

Denic, V., Quan, E. M., & Weissman, J. S. (2006). A luminal surveillance complex that selects misfolded glycoproteins for ER-associated degradation. *Cell*, *126*(2), 349–359.

Denisov, A. Y., Määttänen, P., Dabrowski, C., Kozlov, G., Thomas, D. Y., & Gehring, K. (2009). Solution structure of the bb' domains of human protein disulfide isomerase. *The Federation of European Biochemical Sciences Journal*, *276*(5), 1440–1449.

Denzel, A., Molinari, M., Trigueros, C., Martin, J. E., Velmurgan, S., Brown, S., Stamp, G., & Owen, M. J. (2002). Early postnatal death and motor disorders in mice congenitally deficient in calnexin expression. *Molecular and Cellular Biology*, *22*(21), 7398–7404.

Dersh, D., Iwamoto, Y., & Argon, Y. (2016). Tay-Sachs disease mutations in HEXA target the α chain of hexosaminidase A to endoplasmic reticulum-associated degradation. *Molecular Biology of the Cell*, *27*(24), 3813–3827. PubMed. https://doi.org/10.1091/mbc.E16-01-0012

Dersh, D., Jones, S. M., Eletto, D., Christianson, J. C., & Argon, Y. (2014). OS-9 facilitates turnover of nonnative GRP94 marked by hyperglycosylation. *Molecular Biology of the Cell*, *25*(15), 2220–2234.

Desbois-Mouthon, C., Wendum, D., Cadoret, A., Rey, C., Leneuve, P., Blaise, A., Housset, C., Tronche, F., Le Bouc, Y., & Holzenberger, M. (2006). Hepatocyte proliferation during liver regeneration is impaired in mice with liver-specific IGF-1R knockout. *The FASEB Journal*, *20*(6), 773–775. https://doi.org/10.1096/fj.05-4704fje

Di, X.-J., Wang, Y.-J., Han, D.-Y., Fu, Y.-L., Adam, D. S., Blagg, B. S., & Mu, T.-W. (2016). Grp94 Protein Delivers γ-Aminobutyric Acid Type A (GABAA) Receptors to Hrd1 Protein-mediated Endoplasmic Reticulum-associated Degradation. *Journal of Biological Chemistry*, *291*(18), 9526–9539.

Dima, R. I., & Thirumalai, D. (2004). Asymmetry in the Shapes of Folded and Denatured States of Proteins. *Journal of Physical Chemistry B*, *108*(21), 6564–6570.

Ding, C., Li, Y., Guo, F., Jiang, Y., Ying, W., Li, D., Yang, D., Xia, X., Liu, W., Zhao, Y., He, Y., Li, X., Sun, W., Liu, Q., Song, L., Zhen, B., Zhang, P., Qian, X., Qin, J., & He, F. (2016). A Cell-type-resolved Liver Proteome. *Molecular & Cellular Proteomics : MCP*, *15*(10), 3190–3202. PubMed. https://doi.org/10.1074/mcp.M116.060145

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. PubMed. https://doi.org/10.1093/bioinformatics/bts635

Dong, M., Bridges, J. P., Apsley, K., Xu, Y., & Weaver, T. E. (2008). ERdj4 and ERdj5 are required for endoplasmic reticulum-associated protein degradation of misfolded surfactant protein C. *Molecular Biology of the Cell*, *19*(6), 2620–2630.

Elkabetz, Y., Argon, Y., & Bar-Nun, S. (2005). Cysteines in CH1 Underlie Retention of Unassembled Ig Heavy Chains. *Journal of Biological Chemistry*, *280*(15), 14402–14412.

Ellgaard, L., Bettendorff, P., Braun, D., Herrmann, T., Fiorito, F., Jelesarov, I., Güntert, P., Helenius, A., & Wüthrich, K. (2002). NMR structures of 36 and 73-residue

fragments of the calreticulin P-domain. *Journal of Molecular Biology*, *322*(4), 773–784.

Ellgaard, L., & Helenius, A. (2003). Quality control in the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, *4*(3), 181–191.

Ellis, J. R. (1996). Discovery of molecular chaperones. *Cell Stress and Chaperones*, *1*(3), 155–160.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, *32*(19), 3047–3048. PubMed. https://doi.org/10.1093/bioinformatics/btw354

Farinha, C. M., & Amaral, M. D. (2005). Most F508del-CFTR Is Targeted to Degradation at an Early Folding Checkpoint and Independently of Calnexin. *Molecular and Cellular Biology*, *25*(12), 5242–5252.

Fatal, N., Suntio, T., & Makarow, M. (2002). Selective protein exit from yeast endoplasmic reticulum in absence of functional COPII coat component sec13p. *Molecular Biology of the Cell*, *13*(12), 4130–4140.

Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Araç, D., Chelliah, Y., Amezcua, C. A., Rizo, J., Gladyshev, V. N., & Deisenhofer, J. (2006). NMR Structures of the Selenoproteins Sep15 and SelM Reveal Redox Activity of a New Thioredoxin-like Family. *Journal of Biological Chemistry*, *281*(6), 3536–3543.

Ferreira, C. R., & Gahl, W. A. (2017). Lysosomal storage diseases. *Translational Science of Rare Diseases*, *2*(1–2), 1–71. https://doi.org/10.3233/TRD-160005

Ferris, S. P., Jaber, N. S., Molinari, M., Arvan, P., & Kaufman, R. J. (2013). UDP-glucose:glycoprotein Glucosyltransferase (UGGT1) Promotes Substrate Solubility in the Endoplasmic Reticulum. *Molecular Biology of the Cell*, *24*(17), 2597–2608.

Flynn, G. C., Pohl, J., Flocco, M. T., & Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature*, *353*(6346), 726–730.

Forster, M. L., Sivick, K., Park, Y., Arvan, P., Lencer, W. I., & Tsai, B. (2006). Protein disulfide isomerase–like proteins play opposing roles during retrotranslocation. *Journal of Cell Biology*, *173*(6), 853–859.

Fraldi, A., Zito, E., Annunziata, F., Lombardi, A., Cozzolino, M., Monti, M., Spampanato, C., Ballabio, A., Pucci, P., Sitia, R., & Cosma, M. P. (2008). Multistep, sequential control of the trafficking and function of the multiple sulfatase deficiency gene product, SUMF1 by PDI, ERGIC-53 and ERp44. *Human Molecular Genetics*, *17*(17), 2610–2621.

Frickel, E.-M., Frei, P., Bouvier, M., Stafford, W. F., Helenius, A., & Ellgaard, L. (2004). ERp57 is a multifunctional thiol-disulfide oxidoreductase. *Journal of Biological Chemistry*, *279*(18), 18277–18287. https://doi.org/10.1074/jbc.M314089200

Galli, C., Bernasconi, R., Soldà, T., Calanca, V., & Molinari, M. (2011). Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER. *Public Library of Science One*, *6*(1), e16304.

Gao, B., Adhikari, R., Howarth, M., Nakamura, K., Gold, M. C., Hill, A. B., Knee, R., Michalak, M., & Elliott, T. (2002). Assembly and antigen-presenting function of

MHC class I molecules in cells lacking the ER chaperone calreticulin. *Immunity*, *16*(1), 99–109.

Gauss, R., Kanehara, K., Carvalho, P., Ng, D. T., & Aebi, M. (2011). A Complex of Pdi1p and the Mannosidase Htm1p Initiates Clearance of Unfolded Glycoproteins from the Endoplasmic Reticulum. *Molecular Cell*, *42*(6), 782–793.

Giusti, B., Lucarini, L., Pietroni, V., Lucioli, S., Bandinelli, B., Sabatelli, P., Squarzoni, S., Petrini, S., Gartioux, C., Talim, B., Roelens, F., Merlini, L., Topaloglu, H., Bertini, E., Guicheney, P., & Pepe, G. (2005). Dominant and recessive COL6A1 mutations in Ullrich scleroatonic muscular dystrophy. *Annals of Neurology*, *58*(3), 400–410. https://doi.org/10.1002/ana.20586

Goloubinoff, P., Christeller, J. T., Gatenby, A. A., & Lorimer, G. H. (1989). Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfoleded state depends on two chaperonin proteins and Mg-ATP. *Nature*, *342*(6252), 884–889.

Gooptu, B., & Lomas, D. A. (2009). Conformational pathology of the serpins: Themes, variations, and therapeutic strategies. *Annual Review of Biochemistry*, *78*, 147–176.

Gopalakrishnapai, J., Gupta, G., Karthikeyan, T., Sinha, S., Kandiah, E., Gemma, E., Oscarson, S., & Surolia, A. (2006). Isothermal titration calorimetric study defines the substrate binding residues of calreticulin. *Biochemical and Biophysical Research Communications*, *351*(1), 14–20.

Graham, J. B., Canniff, N. P., & Hebert, D. N. (2019). TPR-containing proteins control protein organization and homeostasis for the endoplasmic reticulum. *Critical*

Reviews in Biochemistry and Molecular Biology, *54*(2), 103–118. PubMed. https://doi.org/10.1080/10409238.2019.1590305

Griffiths, S. W., King, J., & Cooney, C. L. (2002). The Reactivity and Oxidation Pathway of Cysteine 232 in Recombinant Human α1-Antitrypsin. *Journal of Biological Chemistry*, *277*(28), 25486–25492.

Ham, H., Woolery, A. R., Tracy, C., Stenesen, D., Krämer, H., & Orth, K. (2014). Unfolded protein response-regulated Drosophila Fic protein reversibly AMPylates BiP chaperone during endoplasmic reticulum homeostasis. *Journal of Biological Chemistry*, *289*(52), 36059–36069.

Hammond, C., Braakman, I., & Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proceedings of the National Acadamy of Sciences of the United States of America*, *91*(3), 913–917.

Haraguchi, M., Yamashiro, S., Furukawa, K., Takamiya, K., Shiku, H., & Furukawa, K. (1995). The effects of the site-directed removal of N-glycosylation sites from β-1,4-N-acetylgalactosaminyltransferase on its function. *Biochemical Journal*, *312*, 273–280.

Hartl, U. F. (2017). Protein Misfolding Diseases. *Annual Review of Biochemistry*, *86*, 21–26. https://doi.org/10.1146/annurev-biochem-061516-044518

Hatahet, F., & Ruddock, L. W. (2007). Substrate recognition by the protein disulfide isomerases. *The Federation of European Biochemical Sciences Journal*, *274*(20), 5223–5234.

Haze, K., Yoshida, H., Yanagi, H., Yura, T., & Mori, K. (1999). Mammalian Transcription Factor ATF6 Is Synthesized as a Transmembrane Protein and Activated by Proteolysis in Response to Endoplasmic Reticulum Stress. *Molecular Biology of the Cell*, *10*(11), 3787–3799.

Hebert, D. N., Foellmer, B., & Helenius, A. (1995). Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, *81*(3), 425–433.

Hebert, D. N., Foellmer, B., & Helenius, A. (1996). Calnexin and calreticulin promote folding, delay oligomerization and suppress degradation of influenza hemagglutinin in microsomes. *The European Molecular Biology Organization*, *15*(12), 2961–2968.

Hebert, D. N., Garman, S. C., & Molinari, M. (2005). The glycan code of the endoplasmic reticulum: Asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends in Cell Biology*, *15*(7), 364–370.

Hebert, D. N., Lamriben, L., Powers, E. T., & Kelly, J. W. (2014). The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature Chemical Biology*, *10*(11), 902–910. https://doi.org/10.1038/nchembio.1651

Hebert, D. N., & Molinari, M. (2007). In and out of the ER: protein folding, quality control, degradation, and related human diseases. *Physiological Reviews*, *87*(4), 1377–1408.

Hebert, D. N., & Molinari, M. (2012). Flagging and docking: Dual roles for N-glycans in protein quality control and cellular proteostasis. *Trends in Biochemical Sciences*, *37*(10), 404–410.

Hebert, D. N., Zhang, J.-X., Chen, W., Foellmer, B., & Helenius, A. (1997). The number and location of glycans on influenza hemagglutinin determine folding and association with calnexin and calreticulin. *Journal of Cell Biology*, *139*(3), 613–623.

Hebert, D. N., Zhang, J.-X., & Helenius, A. (1998). Protein folding and maturation in a cell-free system. *Biochemistry and Cell Biology*, *76*(5), 867–873.

Helenius, A. (1994). How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Molecular Biology of the Cell*, *5*(3), 253–265.

Helenius, A., & Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry*, *73*, 1019–1049.

Helenius, A., & Hammond, C. (1994). Folding of VSV G protein: Sequential interaction with BiP and calnexin. *Science*, *266*(5184), 456–458.

Hendershot, L. M. (2004). The ER function BiP is a master regulator of ER function. *Mount Sinai Journal of Medicine*, *71*(5), 289–297.

Hetz, C., & Papa, F. R. (2018). The unfolded protein response and cell fate control. *Molecular Cell*, *69*(2), 169–181.

Hirao, K., Natsuka, Y., Tamura, T., Wada, I., Morito, D., Natsuka, S., Romero, P., Sleno, B., Tremblay, L. O., Herscovics, A., Nagata, K., & Hosokawa, N. (2006). EDEM3, a soluble EDEM homolog, enhances glycoprotein endoplasmic reticulum-

associated degradation and mannose trimming. *Journal of Biological Chemistry*, *281*(14), 9650–9658.

Holmgren, A., Söderberg, B.-O., Eklund, H., & Brändén, C.-I. (1975). Three-dimensional structure of Escherichia coli thioredoxin-S2 to 2.8 A resolution. *Proceedings of the National Acadamey of Sciences of the United States of America*, *72*(6), 2305–2309.

Horwich, A., Neupert, W., & Hartl, U. F. (1990). Protein-catalysed protein folding. *Trends in Biotechnology*, *8*(5), 126–131.

Hosokawa, N., Tremblay, L. O., Sleno, B., Kamiya, Y., Wada, I., Nagata, K., Kato, K., & Herscovics, A. (2010). EDEM1 accelerates the trimming of α1,2-linked mannose on the C branch of N-glycans. *Glycobiology*, *20*(5), 567–575.

Hosokawa, N., Wada, I., Nagasawa, K., Moriyama, T., Okawa, K., & Nagata, K. (2008). Human XTP3-B forms an endoplasmic reticulum quality control scaffold with the HRD1-SEL1L ubiquitin ligase complex and BiP. *Journal of Biological Chemistry*, *283*(30), 20914–20924.

Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., & Hartl, F. U. (1999). Identification of in vivo substrates of the chaperonin GroEL. *Nature*, *402*(6758), 147–154. https://doi.org/10.1038/45977

Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, *425*(6959), 686–691.

Huntington, J. A., Read, R. J., & Carrell, R. W. (2000). Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, *407*(6806), 923–926.

Hurtley, S. M., & Helenius, A. (1989). Protein oligomerization in the endoplasmic reticulum. *Annual Review of Cell Biology*, *5*, 277–307.

Hwang, J., & Qi, L. (2018). Quality Control in the Endoplasmic Reticulum: Crosstalk between ERAD and UPR pathways. *Trends in Biochemical Sciences*, *43*(8), 593–605.

Isidoro, C., Maggioni, C., Demoz, M., Pizzagalli, A., Fra, A. M., & Sitia, R. (1996). Exposed thiols confer localization in the endoplasmic reticulum by retention rather than retrieval. *Journal of Biological Chemistry*, *271*(42), 26138–26142.

Itzhak, D. N., Tyanova, S., Cox, J., & Borner, G. H. (2016). Global, Quantitative and Dynamic Mapping of Protein Subcellular Localization. *ELife*, *5*(e16950). https://doi.org/10.7554/eLife.16950

Jakob, C. A., Burda, P., Roth, J., & Aebi, M. (1998). Degradation of Misfolded Endoplasmic Reticulum Glycoproteins in Saccharomyces cerevisiae Is Determined by a Specific Oligosaccharide Structure. *The Journal of Cell Biology*, *142*(5), 1223–1233.

Jansen, G., Määttänen, P., Denisov, A. Y., Scarffe, L., Schade, B., Balghi, H., Dejaarrd, K., Chen, L. Y., Muller, W. J., Gehring, K., & Thomas, D. Y. (2012). An Interaction Map of Endoplasmic Reticulum Chaperones and Foldases. *Molecular and Cellular Proteomics*, *11*(9), 710–723.

Kabani, M., Beckerich, J.-M., & Brodsky, J. L. (2002). Nucleotide exchange factor for the yeast Hsp70 molecular chaperone Ssa1p. *Molecular and Cellular Biology*, *22*(13), 4677–4689.

Kampinga, H. H., Andreasson, C., Barducci, A., Cheetham, M. E., Cyr, D., Emanuelsson, C., Geneveaux, P., Gestwicki, J. E., Goloubinoff, P., Huerta-Cepas, J., Kirstein, J., Liberek, K., Mayer, M. P., Nagata, K., Nillegoda, N. B., Pulido, P., Ramos, C., de los Rios, P., Rospert, S., … Marszalek, J. (2018). Function, evolution, and structure of J-domain proteins. *Cell Stress and Chaperones*.

Kapoor, M., Ellgaard, L., Gopalakrishnapai, J., Schirra, C., Gemma, E., Oscarson, S., Helenius, A., & Surolia, A. (2004). Mutational analysis provides molecular insight into the carbohydrate-binding region of calreticulin: Pivotal roles of tyrosine-109 and aspartate-135 in carbohydrate recognition. *Biochemistry*, *43*(1), 97–106. https://doi.org/10.1021/bi0355286

Karaivanova, V. K., Luan, P., & Spiro, R. G. (1998). Processing of viral envelope glycoprotein by the endomannosidase pathway: Evaluation of host cell specificity. *Glycobiology*, *8*(7), 725–730.

Kasaikina, M. V., Fomenko, D. E., Labunskyy, V. M., Lachke, S. A., Qiu, W., Moncaster, J. A., Zhang, J., Wojnarowicz, M. W. J., Natarajan, S. K., Malinouski, M., Schweizer, U., Tsuji, P. A., Carlson, B. A., Maas, R. L., Lou, M. F., Goldstein, L. E., Hatfield, D. L., & Gladyshev, V. N. (2011). Roles of the 15-kDa Selenoprotein (Sep15) in Redox Homeostasis and Cataract Development Revealed by the Analysis of Sep 15

Knockout Mice. *Journal of Biological Chemistry*, *286*(38), 33203–33212. https://doi.org/10.1074/jbc.M111.259218

Katta, S. S., Smoyer, C. J., & Jaspersen, S. L. (2014). Destination: Inner nuclear membrane. *Trends in Cell Biology*, *24*(4), 221–229. https://doi.org/10.1016/j.tcb.2013.10.006

Kerem, A., Kronman, C., Bar-Nun, S., Shafferman, A., & Velan, B. (1993). Interrelations between assembly and secretion of recombinant human acetylcholinesterase. *Journal of Biological Chemistry*, *268*(1), 180–184.

Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H.-C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., & Hartl, F. U. (2005). Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell*, *122*(2), 209–220. https://doi.org/10.1016/j.cell.2005.05.028

Khodayari, N., Marek, G., Lu, Y., Krotova, K., Wang, R. L., & Brantly, M. (2017). Erdj3 Has an Essential Role for Z Variant Alpha-1-Antitrypsin Degradation. *Journal of Cellular Biochemistry*, *118*(10), 3090–3101.

Kim, W., Spear, E. D., & Ng, D. T. (2005). Yos9p detects and targets misfolded glycoproteins for ER-associated degradation. *Molecular Cell*, *19*(6), Yos9p detects and targets misfolded glycoproteins for ER-associated degradation.

Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Hartl, U. F. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry*, *82*, 323–355.

King, H., Aleksic, T., Haluska, P., & Macaulay, V. M. (2014). Can we unlock the potential of IGF-1R inhibition in cancer therapy? *Cancer Treatment Reviews*, *40*(9), 1096–1105. https://doi.org/10.1016/j.ctrv.2014.07.004

Kirkegaard, T., Gray, J., Priestman, D. A., Wallom, K.-L., Atkins, J., Olsen, O. D., Klein, A., Drndarski, S., Petersen, N. H. T., Ingemann, L., Smith, D. A., Morris, L., Bornæs, C., Jørgensen, S. H., Williams, I., Hinsby, A., Arenz, C., Begley, D., Jäättelä, M., & Platt, F. M. (2016). Heat shock protein–based therapy as a potential candidate for treating the sphingolipidoses. *Science Translational Medicine*, *8*(355), 355ra118. https://doi.org/10.1126/scitranslmed.aad9823

Korotkov, K. V., Kuramaswamy, E., Zhou, Y., Hatfield, D. L., & Gladyshev, V. N. (2001). Association between the 15-kDa Selenoprotein and UDP-glucose:Glycoprotein Glucosyltransferase in the Endoplasmic Reticulum of Mammalian Cells. *Journal of Biological Chemistry*, *276*(18), 15330–15336.

Kozlov, G., Bastos-Aristizabal, S., Määttänen, P., Rosenauer, A., Zheng, F., Killikelly, A., Trempe, J.-F., Thomas, D. Y., & Gehring, K. (2010). Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. *Journal of Biological Chemistry*, *285*(46), 35551–35557. https://doi.org/10.1074/jbc.M110.160101

Kozlov, G., & Gehring, K. (2020). Calnexin Cycle—Structural Features of the ER Chaperone System. *Federation of European Biochemical Sciences Journal*, *doi: 10.1111/febs.15330. Online ahead of print*.

Kozlov, G., Määttänen, P., Schrag, J. D., Hura, G. L., Gabriel, L., Cygler, M., Thomas, D. Y., & Gehring, K. (2009). Structure of the Noncatalytic Domains and Global Fold of

the Protein Disulfide Isomerase ERp72. *Structure*, *17*(5), 651-659 Journal home page for Structure.

Kozlov, G., Muñoz-Escobar, J., Castro, K., & Gehring, K. (2017a). Mapping the ER Interactome: The P Domains of Calnexin and Calreticulin as Plurivalent Adapters for Foldases and Chaperones. *Structure*, *25*(9), 1415–1422.

Kozlov, G., Muñoz-Escobar, J., Castro, K., & Gehring, K. (2017b). Mapping the ER interactome: The P domains of calnexin and calreticulin as plurivalent adapters for foldases and chaperones. *Structure*, *25*(9), 1415–1422. https://doi.org/10.1016/j.str.2017.07.010

Kozlov, G., Pocanschi, C. L., Rosenauer, A., Bastos-Aristizabal, S., Gorelik, A., Williams, D. B., & Gehring, K. (2010). Structural basis of carbohydrate recognition by calreticulin. *Journal of Biological Chemistry*, *285*(49), 38612–38620. https://doi.org/10.1074/jbc.M110.168294

Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J., & Sambrook, J. (1988). The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. *Nature*, *332*(6163), 462–464.

Lamriben, L., Graham, J. B., Adams, B. M., & Hebert, D. N. (2016). N-glycan based ER molecular chaperone and protein quality control system: The calnexin binding cycle. *Traffic*, *17*(4), 308–326.

Lamriben, L., Oster, M. E., Tamura, T., Tian, W., Yang, Z., Clausen, H., & Hebert, D. N. (2018). EDEM1's mannosidase-like domain binds ERAD client proteins in a redox-

sensitive manner and possesses catalytic activity. *Journal of Biological Chemistry*, *293*(36), 13932–13945.

Laskey, R. A., Honda, B., Mills, A. D., & Finch, J. T. (1978). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Nature*, *275*, 416–420.

Lederkremer, G. Z. (2009). Glycoprotein folding, quality control and ER-associated degradation. *Current Opinion in Structural Biology*, *19*(5), 515–523.

Lee, K., Tirasophon, W., Shen, X., Michalak, M., Prywes, R., Okada, T., Yoshida, H., Mori, K., & Kaufman, R. J. (2002). IRE1-mediated unconventional mRNA splicing and S2P-mediated ATF6 cleavage merge to regulate XBP1 in signaling the unfolded protein response. *Genes and Development*, *16*(4), 452–466.

Lehmann, M., André, F., Bellan, C., Remacle-Bonnet, M., Garrouste, F., Parat, F., Lissitsky, J.-C., Marvaldi, J., & Pommier, G. (1998). Deficient Processing and Activity of Type I Insulin-Like Growth Factor Receptor in the Furin-Deficient LoVo-C5 Cells*. *Endocrinology*, *139*(9), 3763–3771. https://doi.org/10.1210/endo.139.9.6184

Leinekugel, P., Michel, S., Conzelmann, E., & Sandhoff, K. (1992). Quantitative correlation between the residual activity of β-hexosaminidase A and arylsulfatase A and the severity of the resulting lysosomal storage disease. *Human Genetics*, *88*(5), 513–523. https://doi.org/10.1007/BF00219337

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323–323. PubMed. https://doi.org/10.1186/1471-2105-12-323

Li, Z., Stafford, W. F., & Bouvier, M. (2001). The metal ion binding properties of calreticulin modulate its conformational flexibility and thermal stability. *Biochemistry*, *40*(37), 11193–11201.

Liu, C.-Y., Fujimori, D. G., & Weissman, J. S. (2016). Htm1p–Pdi1p is a folding-sensitive mannosidase that marks N-glycoproteins for ER-associated protein degradation. *Proceedings of the National Acadamy of Sciences of the United States of America*, *113*(28), E4015–E4024.

Liu, L., Werner, M., & Gershenson, A. (2014). Collapse of a long axis: Single-molecule Förster resonance energy transfer and serpin equilibrium unfolding. *Biochemistry*, *53*(18), 2903–2914.

Liu, Y., Choudhury, P., Cabral, C. M., & Sifers, R. N. (1997). Intracellular Disposal of Incompletely Folded Human α1-Antitrypsin Involves Release from Calnexin and Post-translational Trimming of Asparagine-linked Oligosaccharides. *Journal of Biological Chemistry*, *272*(12), 7946–7951.

Lonowski, L. A., Narimatsu, Y., Riaz, A., Delay, C. E., Yang, Z., Niola, F., Duda, K., Ober, E. A., Clausen, H., Wandall, H. H., Hansen, S. H., Bennett, E. P., & Frödin, M. (2017). Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nature Protocols*, *12*(3), 581–603. PubMed. https://doi.org/10.1038/nprot.2016.165

Lucioli, S., Giusti, B., Mercuri, E., Vanegas, O. C., Lucarini, L., Pietroni, V., Urtizberea, A., Yaou, R. B., de Visser, M., van der Kooi, A. J., Bönnemann, C., Iannaccone, S. T., Merlini, L., Bushby, K., Muntoni, F., Bertini, E., Chu, M.-L., & Pepe, G. (2005).

Detection of common and private mutations in the COL6A1 gene of patients with Bethlem myopathy. *Neurology*, *64*(11), 1931. https://doi.org/10.1212/01.WNL.0000163990.00057.66

Määttänen, P., Gehring, K., Bergeron, J. J., & Thomas, D. Y. (2010). Protein quality control in the ER: the recognition of misfolded proteins. *Seminars in Cell and Developmental Biology*, *21*(5), 500–511.

Maegawa, K., Watanabe, S., Noi, K., Okumura, M., Amagai, Y., Michio, I., Ushioda, R., Nagata, K., Ogura, T., & Inaba, K. (2017). The Highly Dynamic Nature of ERdj5 Is Key to Efficient Elimination of Aberrant Protein Oligomers through ER-Associated Degradation. *Structure*, *25*(6), 846–857.

Mahuran, D. J. (1999). Biochemical consequences of mutations causing the GM2 gangliosidoses. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, *1455*(2–3), 105–138. https://doi.org/10.1016/s0925-4439(99)00074-5

Mahuran, D. J., Neote, K., Klavins, M. H., Leung, A., & Gravel, R. A. (1988). Proteolytic processing of pro-alpha and pro-beta precursors from human beta-hexosaminidase. Generation of the mature alpha and beta a beta b subunits. *Journal of Biological Chemistry*, *263*(10), 4612–4618.

Margittai, É., & Sitia, R. (2011). Oxidative Protein Folding in the Secretory Pathway and Redox Signaling Across Compartments and Cells. *Traffic*, *12*(1), 1–8. https://doi.org/10.1111/j.1600-0854.2010.01108.x

Mariappan, M., Radhakrishnan, K., Dierks, T., Schmidt, B., & von Figura, K. (2008). ERp44 Mediates a Thiol-independent Retention of Formylglycine-generating Enzyme in the Endoplasmic Reticulum. *Journal of Biological Chemistry*, *283*(10), 6375–6383.

Marin, J. L. (1995). Thioredoxin—A fold for all reasons. *Structure*, *3*(3), 245–250.

Matlack, K. E., Misselwitz, B., Plath, K., & Rapoport, T. A. (1999). BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro-α Factor across the ER Membrane Author links open overlay panel. *Cell*, *97*(5), 553–564.

Matzner, U., Matthes, F., Herbst, E., Lüllmann-Rauch, R., Callaerts-Vegh, Z., D'Hooge, R., Weigelt, C., Eistrup, C., Fogh, J., & Gieselmann, V. (2007). Induction of tolerance to human arylsulfatase A in a mouse model of metachromatic leukodystrophy. *Molecular Medicine (Cambridge, Mass.)*, *13*(9–10), 471–479. PubMed. https://doi.org/10.2119/2007-00063.Matzner

Mazzarella, R. A., Srinivasan, M., Haugejordan, S. M., & Green, M. (1990). ERp72, an abundant luminal endoplasmic reticulum protein, contains three copies of the active site sequences of protein disulfide isomerase. *Journal of Biological Chemistry*, *265*(2), 1094–1101.

McCracken, A. A., & Brodsky, J. L. (1996). Assembly of ER-associated protein degradation in vivo: Dependence on cytosol, calnexin and ATP. *Journal of Cell Biology*, *132*(3), 291–298.

McLaughlin, S. H., & Bulleid, N. J. (1998). Thiol-independent interaction of protein disulphide isomerase with type X collagen during intra-cellular folding and assembly. *Biochemical Journal*, *331*, 793–800.

Menon, S., Lee, J., Abplanalp, W. A., Yoo, S.-E., Agui, T., Furudata, S., Kim, P. S., & Arvan, P. (2007). Oxidoreductase Interactions Include a Role for ERp72 Engagement with Mutant Thyroglobulin from the rdw/rdw Rat Dwarf. *Journal of Biological Chemistry*, *282*(9), 6183–6191.

Mesaeli, N., Nakamura, K., Zvaritch, E., Dickie, P., Dziak, E., Krause, K.-H., Opas, M., MacLennan, D. H., & Michalak, M. (1999). Calreticulin is essential for cardiac development. *Journal of Cell Biology*, *144*(5), 857–868.

Molinari, M., Calanca, V., Galli, C., Lucca, P., & Paganetti, P. (2003). Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science*, *299*(5611), 1397–1400.

Molinari, M., Galli, C., Vanoni, O., Arnold, S. M., & Kaufman, R. J. (2005). Persistent glycoprotein misfolding activates the glucosidase II/UGT1-driven calnexin cycle to delay aggregation and loss of folding competence. *Molecular Cell*, *20*(4), 503–512.

Morava, E., Tiemes, V., Thiel, C., Seta, N., de Lonlay, P., de Klerk, H., Mulder, M., Rubio-Gozalbo, E., Visser, G., van Hasselt, P., Horovitz, D. D. G., de Souza, C. F. M., Schwartz, I. V. D., Green, A., Al-Owain, M., Uziel, G., Sigaudy, S., Chabrol, B., van Spronsen, F.-J., … Wevers, R. A. (2016). ALG6-CDG: a recognizable phenotype with epilepsy, proximal muscle weakness, ataxia and behavioral and limb anomalies. *Journal of Inherited Metabolic Disease*, *39*(5), 713–723. https://doi.org/10.1007/s10545-016-9945-x

Murakoshi, H., Shibata, A. C. E., Nakahata, Y., & Nabekura, J. (2015). A dark green fluorescent protein as an acceptor for measurement of Förster resonance energy transfer. *Scientific Reports*, *5*, 15334–15334. PubMed. https://doi.org/10.1038/srep15334

Mutgan, A. C., Besikcioglu, H. E., Wang, S., Friess, H., Ceyhan, G. O., & Demir, I. E. (2018). Insulin/IGF-driven cancer cell-stroma crosstalk as a novel therapeutic target in pancreatic cancer. *Molecular Cancer*, *17*(1), 66–66. PubMed. https://doi.org/10.1186/s12943-018-0806-0

Narimatsu, Y., Joshi, H. J., Yang, Z., Gomes, C., Chen, Y.-H., Lorenzetti, F. C., Furukawa, S., Schjoldager, K. T., Hansen, L., Clausen, H., Bennett, E. P., & Wandall, H. H. (2018). A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology*, *28*(5), 295–305.

Nicchitta, C. V., & Blobel, G. (1993). Lumenal proteins of the mammalian endoplasmic reticulum are required to complete protein translocation. *Cell*, *73*(5), 989–998.

Ninagawa, S., Okada, T., Sumitomo, Y., Kamiya, Y., Kato, K., Horimoto, S., Ishikawa, T., Takeda, S., Sakuma, T., Yamamoto, T., & Mori, K. (2014). EDEM2 initiates mammalian glycoprotein ERAD by catalyzing the first mannose trimming step. *Journal of Cell Biology*, *206*(3), 347–356.

Nishimura, T., & Stefan, C. J. (2020). Specialized ER membrane domains for lipid metabolism and transport. *Endoplasmic Reticulum Platforms for Lipid Dynamics*, *1865*(1), 158492. https://doi.org/10.1016/j.bbalip.2019.07.001

Oda, Y., Hosokawa, N., Wada, I., & Nagata, K. (2003). EDEM as an acceptor of terminally misfolded glycoproteins released from calnexin. *Science*, *299*(5611), 1394–1397.

Oka, O. B., & Bulleid, N. J. (2013). Forming disulfides in the endoplasmic reticulum. *Biochimica et Biophysica Acta*, *1833*(11), 2425–2429.

Okumura, M., Kadokura, H., & Inaba, K. (2015). Structures and functions of protein disulfide isomerase family members involved in proteostasis in the endoplasmic reticulum. *Free Radical Biology and Medicine*, *83*, 314–322.

Olivari, S., Galli, C., Alanen, H., Ruddock, L., & Molinari, M. (2005). A novel stress-induced EDEM variant regulating endoplasmic reticulum associated glycoprotein degradation. *The Journal of Biological Chemistry*, *280*(4), 2424–2428.

Olivari, S., & Molinari, M. (2007). Glycoprotein folding and the role of EDEM1, EDEM2 and EDEM3 in degradation of folding-defective glycoproteins. *Federation of European Biochemical Sciences Letters*, *581*(19), 3658–3664.

Oliver, J. D., van der Wal, F. J., Bulleid, N. J., & High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. *Science*, *275*(5296), 86–88.

Olzmann, J. A., Kopito, R. R., & Christianson, J. C. (2013). The mammalian endoplasmic reticulum-associated degradation system. *Cold Spring Harbor Perspectives in Biology*, *5*(9), a013185.

Otero, J. H., Lizák, B., & Hendershot, L. M. (2010). Life and Death of a BiP substrate. *Seminars in Cell and Developmental Biology*, *21*(5), 472–478.

Parodi, A. J., & Caramelo, J. J. (2015). A sweet code for glycoprotein folding. *Federation of European Biochemical Sciences Letters*, *589*(22), 3379–3387.

Parodi, A. J., & Cazzulo, J. J. (1982). Protein glycosylation in Trypanosoma cruzi. II. Partial characterization of protein-bound oligosaccharides labeled “in vivo.” *Journal of Biological Chemistry*, *257*(13), 7641–7645.

Patschull, A. O., Segu, L., Nyon, M. P., Lomas, D. A., Nobeli, I., Barrett, T. E., & Gooptu, B. (2011). Therapeutic target-site variability in α1-antitrypsin characterized at high resolution. *Acta Crysallographica Section F Structural Biology and Crystallization Communications*, *67 (Pt 12)*, 1492–1497.

Pearse, B. R., Gabriel, L., Wang, N., & Hebert, D. N. (2008). A cell-based reglucosylation assay demonstrates the role of GT1 in the quality control of a maturing glycoprotein. *Journal of Cell Biology*, *181*(2), 309–320.

Pearse, B. R., Tamura, T., Sunryd, J. C., Grabowski, G. A., Kaufman, R. J., & Hebert, D. N. (2010). The role of UDP-Glc:glycoprotein glucosyltransferase 1 in the maturation of an obligate substrate prosaposin. *Journal of Cell Biology*, *189*(5), 829–841.

Pelham, H. R. (1986). Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell*, *46*(7), 959–961.

Perry, D., & Carrell, R. W. (1996). Molecular genetics of human antithrombin deficiency. *Human Mutation*, *7*(1), 7–22.

Perry, D., Daly, M., Colvin, B., Brown, K., & Carrell, R. W. (1995). Two antithrombin mutations in a compound heterozygote: Met20Thr and Tyr166Cys. *American Journal of Hematology*, *50*(3), 215–216.

Peterson, J. R., Ora, A., Van, P. N., & Helenius, A. (1995). Transient, lectin-like association of calreticulin with folding intermediates of cellular and viral glycoproteins. *Molecular Biology of the Cell*, *6*(9), 1173–1184.

Picard, V., Ersdal-Badju, E., & Bock, S. (1995). Partial glycosylation of antithrombin III asparagine-135 is caused by the serine in the third position of its N-glycosylation consensus sequence and is responsible for production of the beta-antithrombin III isoform with enhanced heparin affinity. *Biochemistry*, *34*(26), 8433–8440.

Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R. A., Byrne, L., Alanen, H. I., Salo, K. E., Kivirikko, K. I., Freedman, R. B., & Ruddock, L. W. (2004). Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *Journal of Biological Chemistry*, *279*(11), 10374–10381.

Plate, L., Cooley, C. B., Chen, J. J., Paxman, R. J., Gallagher, C. M., Madoux, F., Genereux, J. C., Dobbs, W., Garza, D., Spicer, T. P., Scampavia, L., Brown, S. J., Rosen, H., Powers, E. T., Walter, P., Hodder, P., Wiseman, L. R., & Kelly, J. W. (2016). Small molecule proteostasis regulators that reprogram the ER to reduce extracellular protein aggregation. *ELife*, *5*, e15550.

Plemper, R. K., Böhmler, S., Bordallo, J., Sommer, T., & Wolf, D. H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. *Nature*, *388*(6645), 891–895.

Pobre, K. F. R., Poet, G. J., & Hendershot, L. M. (2018). The endoplasmic reticulum (ER) chaperone BiP is a master regulator of ER functions: Getting by with a little help from ERdj friends. *Journal of Biological Chemistry*, *294*(6), 2098–2108.

Preissler, S., Rato, C., Chen, R., Antrobus, R., Ding, S., Fearnley, I. M., & Ron, D. (2015). AMPylation matches BiP activity to client protein load in the endoplasmic reticulum. *ELife*, *4*, e12621.

Preissler, S., Rato, C., Perera, L., Saudek, V., & Ron, D. (2017). FICD acts bi-functionally to AMPylate and de-AMPylate the endoplasmic reticulum chaperone BiP. *Nature Structural and Molecular Biology*, *24*(1), 23–29.

Preissler, S., Rohland, L., Yan, Y., Chen, R., Read, R. J., & Ron, D. (2017). AMPylation targets the rate-limiting step of BiP's ATPase cycle for its functional inactivation. *ELife*, *6*, e29428.

Preissler, S., & Ron, D. (2018). Early Events in the Endoplasmic Reticulum Unfolded Protein Response. *Cold Spring Harbor Perspectives in Biology*, a033894.

Qiang, L., Wang, H., & Farmer, S. R. (2007). Adiponectin Secretion Is Regulated by SIRT1 and the Endoplasmic Reticulum Oxidoreductase Ero1-Lα. *Molecular and Cellular Biology*, *27*(13), 4698–4707.

Qin, S.-Y., Hu, D., Matsumoto, K., Takeda, K., Matsumoto, N., Yamaguchi, Y., & Yamamoto, K. (2012). Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins. *Journal of Biological Chemistry*, *287*(45), 38080–38089.

Quan, E. M., Kamiya, Y., Kamiya, D., Denic, V., Weibezhan, J., Kato, K., & Weissman, J. S. (2010). Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Molecular Cell*, *32*(6), 870–877.

Quellhorst, G. J., Jr., O'Rear, J. L., Cacan, R., Verbert, A., & Krag, S. S. (1999). Nonglucosylated oligosaccharides are transferred to protein in MI8-5 Chinese hamster ovary cells. *Glycobiology*, *9*(1), 65–72.

Rabouille, C. (2017). Pathways of unconventional protein secretion. *Trends in Cell Biology*, *27*(3), 230–240.

Rajagopalan, S., & Brenner, M. B. (1994). Calnexin retains unassembled major histocompatibility complex class I free heavy chains in the endoplasmic reticulum. *Journal of Experimental Medicine*, *180*(1), 407–412.

Rajagopalan, S., Xu, Y., & Brenner, M. B. (1994). Retention of Unassembled Components of Integral Membrane Proteins by Calnexin. *Science*, *263*(5145), 387–390. https://doi.org/10.1126/science.8278814

Rauniyar, N., & Yates, J. R. I. (2014). Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of Proteome Research*, *13*(12), 5293–5309. https://doi.org/10.1021/pr500880b

Rebeaud, M. E., Mallik, S., Goloubinoff, P., & Tawfik, D. S. (2020). On the evolution of chaperones and co-chaperones and the expansion of proteomes across the Tree of Life. *BioRxiv*, 2020.06.08.140319. https://doi.org/10.1101/2020.06.08.140319

Ritter, C., & Helenius, A. (2000). Recognition of local glycoprotein misfolding by the ER folding sensor UDP-glucose:glycoprotein glucosyltransferase. *Nature Structural Biology*, *7*(4), 278–280. https://doi.org/10.1038/74035

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. PubMed. https://doi.org/10.1093/bioinformatics/btp616

Ronzoni, R., Berardelli, R., Medicina, D., Sitia, R., Gooptu, B., & Fra, A. M. (2016). Aberrant disulphide bonding contributes to the ER retention of alpha1-antitrypsin deficiency variants. *Human Molecular Genetics*, *25*(4), 642–650.

Roversi, P., Marti, L., Caputo, A. T., Alonzi, D. S., Hill, J. C., Dent, K. C., Kumar, A., Levasseur, M. D., Lia, A., Waksman, T., Basu, S., Albrecht, Y. S., Qian, K., McIvor, J. P., Lipp, C. B., Siliqi, D., Vasiljević, S., Mohammed, S., Lukacik, P., … Zitzmann, N. (2017). Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint. *Proceedings of the National Acadamy of Sciences of the United States of America*, *114*(32), 8544–8549. https://doi.org/10.1073/pnas.1703682114

Ruiz-Canada, C., Kelleher, D. J., & Gilmore, R. (2009). Cotranslational and Posttranslational N-glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. *Cell*, *136*(2), 272–283.

Rutkowski, T. D., Kang, S.-W., Goodman, A. G., Garrison, J. L., Taunton, J., Katze, M. G., Kaufman, R. J., & Hedge, R. S. (2007). The role of p58IPK in protecting the stressed endoplasmic reticulum. *Molecular Biology of the Cell*, *18*(9), 3681–3691.

Saeed, M., Suzuki, R., Watanabe, N., Masaki, T., Tomonaga, M., Muhammad, A., Kato, T., Matsuura, Y., Watanabe, H., Wakita, T., & Suzuki, T. (2011). Role of the endoplasmic reticulum-associated degradation (ERAD) pathway in degradation of hepatitis C virus envelope proteins and production of virus particles. *Journal of Biological Chemistry*, *286*(43), 37264–37273.

Saito, K., & Maeda, M. (2019). Not just a cargo receptor for large cargoes; an emerging role of TANGO1 as an organizer of ER exit sites. *The Journal of Biochemistry*, *166*(2), 115–119. https://doi.org/10.1093/jb/mvz036

Sato, Y., Kojima, R., Okumura, M., Hagiwara, M., Masui, S., Maegawa, K., Saiki, M., Horibe, T., Suzuki, M., & Inaba, K. (2013). Synergistic cooperation of PDI family members in peroxiredoxin 4-driven oxidative protein folding. *Scientific Reports*, *3*, 2456.

Satoh, T., Song, C., Zhu, T., Toshimori, T., Murata, K., Hayashi, Y., Kamikubo, H., Uchihashi, T., & Kato, K. (2017). Visualization of a flexible modular structure of the ER folding-sensor enzyme UGGT. *Visualisation of a Flexible Modular Structure of the ER Folding-Sensor Enzyme UGGT*, *7*(1), 12142. https://doi.org/10.1038/s41598-017-12283-w

Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., Mackeen, M., Stier, G., Gibson, T. J., Feizi, T., Pieler, T., & Muhle-Goll, C. (2008). Malectin: A novel

carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Molecular Biology of the Cell*, *19*(8), 3404–3414.

Schlecht, R., Erbse, A. H., Bukau, B., & Mayer, M. P. (2011). Mechanics of Hsp70 chaperones enables differential interaction with client proteins. *Nature Structural and Molecular Biology*, *18*(3), 345–351.

Schneider, M., Rosam, M., Glaser, M., Patronov, A., Shah, H., Back, K. C., Daake, M. A., Buchner, J., & Antes, I. (2016). BiPPred: Combined sequence- and structure-based prediction of peptide binding to the Hsp70 chaperone BiP. *Proteins*, *84*(10), 1390–1407.

Schrag, J. D., Bergeron, J. J., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y., & Cygler, M. (2001). The Structure of calnexin, an ER chaperone involved in quality control of protein folding. *Molecular Cell*, *8*(3), 633–644. https://doi.org/10.1016/s1097-2765(01)00318-5

Sell, C., Dumenil, G., Deveaud, C., Miura, M., Coppola, D., DeAngelis, T., Rubin, R., Efstratiadis, A., & Baserga, R. (1994). Effect of a null mutation of the insulin-like growth factor I receptor gene on growth and transformation of mouse embryo fibroblasts. *Molecular and Cellular Biology*, *14*(6), 3604. https://doi.org/10.1128/MCB.14.6.3604

Sevier, C. S., & Kaiser, C. A. (2008). Ero1 and redox homeostasis in the endoplasmic reticulum. *Biochimica et Biophysica Acta*, *1783*(4), 549–556.

Shenkman, M., Ron, E., Yehuda, R., Benyair, R., Khalaila, I., & Lederkremer, G. Z. (2018). Mannosidase activity of EDEM1 and EDEM2 depends on an unfolded state of their glycoprotein substrates. *Communications Biology*, *1*(1), 172. https://doi.org/10.1038/s42003-018-0174-8

Shi, Y., Vattem, K. M., Sood, R., An, J., Liang, J., Stramm, L., & Wek, R. C. (1998). Identification and characterization of pancreatic eukaryotic initiation factor 2 alpha-subunit kinase, PEK, involved in translational control. *Molecular and Cellular Biology*, *18*(12), 7499–7509.

Shrimal, S., Cherepanova, N. A., & Gilmore, R. (2015). Cotranslational and posttranslocational N-glycosylation of proteins in the endoplasmic reticulum. *Seminars in Cell and Developmental Biology*, *41*, 71–78.

Shrimal, S., Cherepanova, N. A., Mandon, E. S., Venev, S. V., & Gilmore, R. (2019). Asparagine-linked Glycosylation Is Not Directly Coupled to Protein Translocation Across the Endoplasmic Reticulum in Saccharomyces cerevisiae. *Molecular Biology of the Cell*, *30*(21), 2626–2638. https://doi.org/10.1091/mbc.E19-06-0330

Shrimal, S., & Gilmore, R. (2015). Reduced expression of the oligosaccharyltransferase exacerbates protein hypoglycosylation in cells lacking the fully assembled oligosaccharide donor. *Glycobiology*, *25*(7), 774–783. PubMed. https://doi.org/10.1093/glycob/cwv018

Sitia, R., Neuberger, M., Alberini, C., Bet, P., Fra, A., Valetti, C., Williams, G., & Milstein, C. (1990). Development regulation of IgM secretion: The role of the carboxy-terminal cysteine. *Cell*, *60*(5), 781–790.

Skowronek, M. H., Hendershot, L. M., & Haas, I. G. (1998). The variable domain of nonassembled Ig light chains determines both their half-life and binding to the chaperone BiP. *Proceedings of the National Acadamy of Sciences of the United States of America*, *95*(4), 1574–1578.

Skropeta, D. (2009). The effect of individual N-glycans on enzyme activity. *Bioorganic and Medicinal Chemistry*, *17*(7), 2545–2653.

Słomińska-Wojewódzka, M., & Sandvig, K. (2015). The Role of Lectin-Carbohydrate Interactions in the Regulation of ER-Associated Protein Degradation. *Molecules*, *20*(6), 9816–9846. https://doi.org/10.3390/molecules20069816

Soldà, T., Galli, C., Kaufman, R. J., & Molinari, M. (2007). Substrate-specific requirements for UGT1-dependent release from calnexin. *Molecular Cell*, *27*(2), 238–249.

Sörgjerd, K., Ghafouri, B., Jonsson, B.-H., Kelly, J. W., Blond, S. Y., & Hammarström, P. (2006). Retention of misfolded mutant transthyretin by the chaperone BiP/GRP78 mitigates amyloidogenesis. *Journal of Molecular Biology*, *356*(12), 469–482.

Sousa, M., & Parodi, A. J. (1995a). The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *The European Molecular Biology Organization Journal*, *14*(17), 4196–4203.

Sousa, M., & Parodi, A. J. (1995b). The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *The European Molecular Biology Organization*, *14*(17), 4196–4203.

Steel, G. J., Fullerton, D. M., Tyson, J. R., & Stirling, C. J. (2004). Coordinated activation of Hsp70 chaperones. *Science*, *303*(5654), 98–101.

Stein, P. E., & Carrell, R. W. (1995). What do dysfunctional serpins tell us about molecular mobility and disease? *Nature Structural Biology*, *2*(2), 96–113.

Su, K., Stoller, T., Rocco, J., Zemsky, J., & Green, R. (1993). Pre-Golgi degradation of yeast prepro-alpha-factor expressed in a mammalian cell. Influence of cell type-specific oligosaccharide processing on intracellular fate. *Journal of Biological Chemistry*, *268*(19), 14301–14309.

Sunryd, J. C., Tannous, A., Lamriben, L., & Hebert, D. N. (2014). Chaperones of the Endoplasmic Reticulum Associated Degradation (ERAD) Pathway. In *The Molecular Chaperones Interaction Networks in Protein Folding and Degradation*. Springer Science.

Szathmary, R., Bielmann, R., Nita-Lazar, M., Burda, P., & Jakob, C. A. (2005). Yos9 protein is essential for degradation of misfolded glycoproteins and may function as lectin in ERAD. *Molecular Cell*, *19*(6), 765–775.

Takeda, K., Qin, S.-Y., Matsumoto, N., & Yamamoto, K. (2014). Association of malectin with ribophorin I is crucial for attenuation of misfolded glycoprotein secretion. *Biochemical and Biophysical Research Communications*, *454*(3), 436–440.

Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., Fujikawa, K., Kajihara, Y., & Ito, Y. (2014). Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology*, *24*(4), 344–350. https://doi.org/10.1093/glycob/cwt163

Tamura, T., Sunryd, J. C., & Hebert, D. N. (2010). Sorting things out through endoplasmic reticulum quality control. *Molecular Membrane Biology*, *27*(8), 412–427.

Tan, Y. L., Genereux, J. C., Pankow, S., Aerts, J. M., Yates, J. R. I., & Kelly, J. W. (2014). ERdj3 is an endoplasmic reticulum degradation factor for mutant glucocerebrosidase variants linked to Gaucher's disease. *Chemistry and Biology*, *21*(8), 967–976.

Tang, H.-Y., Huang, C.-H., Zhuang, Y.-H., Christianson, J. C., & Chen, X. (2014). EDEM2 and OS-9 are required for ER-associated degradation of non-glycosylated sonic hedgehog. *Public Library of Science One*, *9*(6), e92164.

Tannous, A., Patel, N., Tamura, T., & Hebert, D. N. (2015). Reglucosylation by UDP-glucose:glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation. *Molecular Biology of the Cell*, *26*(3), 390–405.

Taylor, S. C., Thibault, P., Tessier, D. C., Bergeron, J. J., & Thomas, D. Y. (2003). Glycopeptide specificity of the secretory protein folding sensor UDP-glucose glycoprotein:glucosyltransferase. *European Molecular Biology Organization Reports*, *4*(4), 405–411.

Thelen, M., Winter, D., Braulke, T., & Gieselmann, V. (2017). SILAC-Based Comparative Proteomic Analysis of Lysosomes from Mammalian Cells Using LC-MS/MS. In K.

Öllinger & H. Appelqvist (Eds.), *Lysosomes: Methods and Protocols* (pp. 1–18). Springer New York. https://doi.org/10.1007/978-1-4939-6934-0_1

Tiziana, A., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T., & Sitia, R. (2003). Thiol-mediated protein retention in the endoplasmic reticulum: The role of ERp44. *The European Molecular Biology Organization Journal*, *22*(19), 5015–5022.

Tom, R., Bisson, L., & Durocher, Y. (2008). Culture of HEK293-EBNA1 Cells for Production of Recombinant Proteins. *Cold Spring Harbor Protocols*, *pdb.prot4976*. https://doi.org/10.1101/pdb.prot4976

Tsunoda, S., Avezov, E., Zyryanova, A., Konno, T., Mendes-Silva, L., Pinho Melo, E., Harding, H. P., & Ron, D. (2014). Intact protein folding in the glutathione-depleted endoplasmic reticulum implicates alternative protein thiol reductants. *ELife*, *3:e03421*.

Tyson, J. R., & Stirling, C. J. (2000). LHS1 and SIL1 provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. *The European Molecular Biology Organization Journal*, *19*(23), 6440–6452.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., … Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419. https://doi.org/10.1126/science.1260419

Ushioda, R., Hoseki, J., Araki, K., Jansen, G., Thomas, D. Y., & Nagata, K. (2008). ERdj5 is required as a disulfide reductase for degradation of misfolded proteins in the ER. *Science*, *321*(5888), 569–572.

Ushioda, R., Hoseki, J., & Nagata, K. (2013). Glycosylation-independent ERAD pathway serves as a backup system under ER stress. *Molecular Biology of the Cell*, *24*(20), 3155–3163.

van der Goot, A. T., Pearce, M. M., Leto, D. E., Shaler, T. A., & Kopito, R. R. (2018). Redundant and Antagonistic Roles of XTP3B and OS9 in Decoding Glycan and Non-glycan Degrons in ER-Associated Degradation. *Molecular Cell*, *70*(3), 516–530.

Vassilakos, A., Cohen-Doyle, M. F., Peterson, P. A., Jackson, M. R., & Williams, D. B. (1996). The molecular chaperone calnexin facilitates folding and assembly of class I histocompatibility molecules. *The European Molecular Biology Organization*, *15*(7), 1495–1506.

Vassilakos, A., Michalak, M., Lehrman, M. A., & Williams, D. B. (1998). Oligosaccharide binding characteristics of the molecular chaperones calnexin and calreticulin. *Biochemistry*, *37*(10), 3480–3490. https://doi.org/10.1021/bi972465g

Vassilakos, A., Myrna, C.-D. F., Peterson, P. A., Jackson, M. R., & Williams, D. B. (1996). The molecular chaperone calnexin facilitates folding and assembly of class I histocompatibility molecules. *The European Molecular Biology Organization Journal*, *15*(7), 1495–1506.

Vavassori, S., Cortini, M., Masui, S., Sannino, S., Anelli, T., Caserta, I. R., Fagioli, C., Mossuto, M. F., Fornili, A., van Anken, E., Degano, M., Inaba, K., & Sitia, R. (2013). A pH-Regulated Quality Control Cycle for Surveillance of Secretory Protein Assembly. *Molecular Cell*, *50*(6), 783–792.

Walczak, C. P., & Tsai, B. (2011). A PDI family network acts distinctly and coordinately with ERp29 to facilitate polyomavirus infection. *Journal of Virology*, *85*(5), 2386–2396.

Walter, P., & Ron, D. (2011). The Unfolded Protein Response: From Stress Pathway to Homeostatic Response. *Science*, *334*(6059), 1081–1086.

Wang, H., Li, S., Wang, J., Chen, S., Sun, X.-L., & Wu, Q. (2018). N-glycosylation in the protease domain of trypsin-like serine proteases mediates calnexin-assisted protein folding. *ELife*, *7*, e35672.

Wang, L., Wang, L., Vavassori, S., Li, S., Ke, H., Anelli, T., Degano, M., Ronzoni, R., Sitia, R., Sun, F., & Wang, C. (2008). Crystal structure of human ERp44 shows a dynamic functional modulation by its carboxy-terminal tail. *European Molecular Biology Organization Reports*, *9*(7), 642–647.

Wang, N., Glidden, E. J., Murphy, S. R., Pearse, B. R., & Hebert, D. N. (2008). The cotranslational maturation program for the type II membrane glycoprotein influenza neuraminidase. *Journal of Biological Chemistry*, *283*(49), 33826–33837. https://doi.org/10.1074/jbc.M806897200

Wang, Z. V., Schraw, T. D., Kim, J.-Y., Khan, T., Rajala, M. W., Follenzi, A., & Scherer, P. E. (2007). Secretion of the adipocyte-specific secretory protein adiponectin

critically depends on thiol-mediated protein retention. *Molecular and Cellular Biology*, *27*(10), 3716–3731.

Watanabe, S., Harayama, M., Kanemura, S., Sitia, R., & Inaba, K. (2017). Structural basis of pH-dependent client binding by ERp44, a key regulator of protein secretion at the ER-Golgi interface. *Proceedings of the National Acadamy of Sciences of the United States of America*, *114*(16), E3224–E3232.

Weerapana, E., Wang, C., Simon, G. M., Richter, F., Khare, S., Dillon, M. B., Bachovchin, D. A., Kerri Mowen, Baker, D., & Cravatt, B. F. (2010). Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature*, *468*(7325), 790–795.

Wens, S. C. A., Kroos, M. A., de Vries, J. M., Hoogeveen-Westerveld, M., Wijgerde, M. G. J. M., van Doorn, P. A., van der Ploeg, A. T., & Reuser, A. J. J. (2012). Remarkably low fibroblast acid α-glucosidase activity in three adults with Pompe disease. *Molecular Genetics and Metabolism*, *107*(3), 485–489. https://doi.org/10.1016/j.ymgme.2012.09.003

Westrate, L. M., Lee, J. E., Prinz, W. A., & Voeltz, G. K. (2015). Form Follows Function: The Importance of Endoplasmic Reticulum Shape. *Annual Review of Biochemistry*, *84*(1), 791–811. https://doi.org/10.1146/annurev-biochem-072711-163501

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina,

J., Boese, J.-H., Bantscheff, M., … Kuster, B. (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, *509*(7502), 582–587. https://doi.org/10.1038/nature13319

Yim, S. H., Everley, R. A., Schildberg, F. A., Lee, S.-G., Orsi, A., Barbati, Z. R., Karatepe, K., Fomenko, D. E., Tsuji, P. A., Luo, H. R., Gygi, S. P., Sitia, R., Sharpe, A. H., Hatfield, D. L., & Gladyshev, V. N. (2018). Role of Selenof as a Gatekeeper of Secreted Disulfide-Rich Glycoproteins. *Cell Reports*, *23*(5), 1387–1398.

Yoshida, H., Matsui, T., Hosokawa, N., Kaufman, R. J., Nagata, K., & Mori, K. (2003). A time-dependent phase shift in the mammalian unfolded protein response. *Developmental Cell*, *4*(2), 265–271.

Yoshida, H., Matsui, T., Yamamoto, A., Okada, T., & Mori, K. (2001). XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*, *107*(7), 881–891.

Younger, J. M., Chen, L., Ren, H.-Y., Rosser, M. F. N., Turnbull, E. L., Fan, C.-Y., Patterson, C., & Cyr, D. M. (2006). Sequential Quality-Control Checkpoints Triage Misfolded Cystic Fibrosis Transmembrane Conductance Regulator. *Cell*, *126*(3), 571–582. https://doi.org/10.1016/j.cell.2006.06.041

Zapun, A., Darby, N. J., Tessier, D. C., Michalak, M., Bergeron, J. J., & Thomas, D. Y. (1998). Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *Journal of Biological Chemistry*, *273*(11), 6009–6012. https://doi.org/10.1074/jbc.273.11.6009

Zhong, Y., Shen, H., Wang, Y., Yang, Y., Yang, P., & Fang, S. (2015). Identification of ERAD components essential for dislocation of the null Hong Kong variant of α-1-antitrypsin (NHK). *Biochemical and Biophysical Research Communications*, *458*(2), 424–428.

Zhou, Z., & Smith, D. L. (1990). Location of disulfide bonds in antithrombin III. *Biomedical and Environmental Mass Spectrometry*, *19*(12), 782–786.

Zielinska, D. F., Gnad, F., Wiśniewski, J. R., & Mann, M. (2010). Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell*, *141*(5), 897–907. https://doi.org/10.1016/j.cell.2010.04.012