The N-end Rule Pathway and Ubr1 enforce the compartmentalization of mitochodrial and secretory proteins through the recognition of P2-encoded cellular location signals

Anthony Tran*

Department of Biological Sciences
National University of Singapore, Singapore 117604

*Corresponding author:
Anthony Tran
anthonytran17@gmail.com

Keywords: Ubr1, N-end rule pathway, CytoQC, protein quality control

ABSTRACT

Cells have evolved intricate systems of protein quality control (PQC) that serve to monitor the cell for misfolded proteins and target them for degradation, ameloriating the effects of the majority of folding failures. In recent years, Ubr1, an E3 ligase conserved from yeast to humans, has been demonstrated to be involved in the degradation of several misfolded substrates within the cytosol through the N-end rule pathway, but the significance of this activity was not completely understood. Through a systematic examination of P2-residue variants of model misfolded substrates, followed by global and targeted analyses of yeast proteins, we have determined that Ubr1 is specifically programmed to preferentially target mistranslocated secretory and mitochondrial proteins in the cytosol through the recognition of P2-encoded cellular location signals. Recognition is partially dependent on whether a protein is misfolded, and hence more likely to disrupt homeostasis in the cytosol. This newfound P2-location signaling dependent mechanism for maintaining cytosolic proteomic integrity may shed light on how Ubr1 and the N-end rule pathway are involved in the progression of neurodegenerative diseases, such as Alzheimer's and Parkinson's, and other maladies linked to the accumulation of misfolded proteins. A corollary to this discovery is that the N-end rule pathway enforces the compartmentalization of secretory and mitochondria-bound proteins by degrading those that fail to reach their intended sub-cellular locations. Thus, the N-end rule pathway is likely to have also played a critical role in the evolution of endosymbiotic relationships which paved the way for advanced eukaryotic cellular life.

INTRODUCTION

Protein quality control (PQC) is an essential protein quality surveillance and degradation system through which cells ensure the integrity of the proteome and maintain cellular homeostasis (Kim et al., 2013). Uncontrolled aggregation of misfolded proteins leads to the formation of insoluble protein deposits that are detrimental to the cell and are widely associated with numerous human diseases such as Alzheimer's, Parkinson's, and Huntington's disease (Currais et al., 2017; Pearce and Kopito, 2017; Dubnikov et al., 2016). Cytosolic quality control (CytoQC) pathways, a subset of PQC, specifically mediates the clearance of cytosolically-localized aberrant proteins. In yeast, San1p and Ubr1p are the primary E3s in CytoQC, having been shown to be involved in the degradation of a diverse range of substrates under stress and non-stress conditions (Szoradi et al., 2018; Fredrickson and Gardner, 2012; Prasad et al., 2010; Heck et al., 2010; Buchberger et al., 2010). Doa10p, Hul5p, Ubr2p, have thus far been demonstrated to have more targeted substrate specificities: Doa10p recognizes C-terminal hydrophobic motifs of pre-inserted secretory proteins which have delayed entry into the ER; Hul5 aids the degradation of cytosolic proteins that misfold due to heat stress and functions primarily as an E4; Ubr2 exhibits partial redundancy with Ubr1 in the geldanamycin-dependent degradation of newly synthesized protein kinases (Ast et al., 2014; Buchberger et al., 2010; Nillegoda et al., 2010; Metzger et al., 2008; Fang et al., 2011).

San1p targets substrates with exposed hydrophobicity consisting of contiguous sequences of at least five or more hydrophobic residues, and is also involved in nuclear quality control (Fredrickson et al., 2011, 2013; Fredrickson and Gardner 2012). Ubr1p is best known for its role in the N-end rule pathway, which relates the half-life of a protein to its N-terminal residues (Varshavsky, 2011). While there have been conflicting reports on the role of the N-end rule pathway in Ubr1-mediated CytoQC (Caplan et al., 2010; Heck et al, 2010), it was previously demonstrated that $\Delta ssC^{22\text{-}519}Leu2_{myc}$, a model misfolded cytosolic substrate, was recognized by Ubr1 through its Met-Φ N-degron, defined as an N-terminal Met followed by a Leu, Phe, Tyr, Trp, or Ile residue at the P2 position (Kim et al., 2014). The finding definitively linked Ubr1-mediated CytoQC with N-terminally encoded N-degrons for at least a subset of cytosolic quality control substrates. Two other model misfolded substrates, Ste6*C and Δ2GFP, are degraded through a combination of the San1 and Ubr1 pathways. Efficient degradation of each substrate requires both E3 ligases with differing dependencies: the degradation of Ste6*C is biased towards the Ubr1 pathway, whereas the degradation of Δ2GFP is biased towards the San1 pathway, which we confirmed here (**Figures 1A-B**; Prasad et al., 2012; Prasad et al., 2010).

As mentioned, previous studies linked the Met-Φ branch of the N-end rule pathway to CytoQC, while the same has been suggested for the Ac/N-end branch (Kim et al., 2014; Kim et al., 2013; Shemorry et al., 2013; Eisele and Wolf 2008). However, the full scope and role of the N-end rule pathway's participation in CytoQC is not well understood. Here we successfully characterize the

full set of P2 amino acid residues which produce N-degrons recognized by Ubr1 by conducting a systematic biochemical analyses of P2-residue variants of a model misfolded CytoQC substrate, Ste6*C (Prasad et al., 2010). This approach allowed us to identify, *in vivo*, N-degrons relevant to misfolded proteins that were not previously detected using peptide arrays on membrane support (SPOT), which only utilizes short peptides *in vitro* (Kim et al., 2014) Furthermore, we performed detailed global bioinformatic analysis of P2 amino acid usage frequency in the *S. cerevisiae* proteome, taking into account the expanded P2-dependent N-degron set, and found that Ubr1's specificity is designed to preferentially degrade secretory and mitochondrial proteins that fail to translocate and thus become mislocalized to the cytosol. Precise *in vivo* analyses of turnover rates of endogenous translocation-inhibited secretory and mitochondrial proteins in the presence and absence of Ubr1 confirm this model of CytoQC actiivty. Finally, an in-depth analysis of a representative set of signal-sequence bearing proteins revealed that ~93% of soluble ER proteins are encoded with Ubr1-QC-compatible P2-residues, versus only ~26% of natively localized cytosolic proteins, indicating that P2-residues encode signals of cellular location that also facilitate the degradation of displaced proteins in the cytosol. Taken together, the findings strongly suggest that the P2-encoding of cellular location, and consequently, the N-end rule pathway, both serve essential roles in enforcing the fidelity of protein compartmentalization in eukaryotic cells.

RESULTS

Ste6*C is a truncated form of Ste6p, plasma membrane ATP-binding cassette transporter, containing only the cytosolic tail of the protein, and harbors a deletion of residues 1249-1290. Δ2GFP is a derivative of wild-type GFP wherein residues 25-36 are deleted. The deletions in each substrate cause misfolding, resulting in their degradation by CytoQC pathways (Prasad et al., 2012; Prasad et al., 2010). Efficient degradation of each substrate requires both Ubr1 and San1 E3 ligases with differing dependencies: the degradation of Ste6*C is biased towards the Ubr1 pathway, whereas the degradation of Δ2GFP is biased towards the San1 pathway **(Figures 1A-B)**. Based on the rules of N-terminal methionine excision, Ste6*C is predicted to possess a Met-Ile sequence at the N-terminus, while Δ2GFP would harbor a -Ser upon methionine excision (Moerschell et al., 1990; Giglione et al., 2004). N-terminal Met-Ile, an N-degron of the Arg/N-end rule pathway, should help direct Ste6*C towards the Ubr1 pathway, while Δ2GFP, with an N-terminal -Ser, is shielded from Ubr1 N-degron recognition (Kim et al., 2014). Since the N-terminal sequence of a misfolded protein impacts its recognition by the Ubr1 pathway, we surmised that by altering the N-terminal sequences of Ste6*C and Δ2GFP, it would be possible to increase or decrease the amount of degradation which proceeds via Ubr1. We thus generated P2 mutants of Ste6*C and Δ2GFP that matched P2-residue of the other, and observed the impact on Ubr1-mediated degradation for each. We modified the N-terminus of Ste6*C to mimic that of Δ2GFP by substituting the P2 position Ile of Ste6*C (Ste6*C-I2I) with a Ser residue, generating Ste6*C-I2S, which carries an N-terminal -Ser. This was confirmed through Edman's degradation

sequencing of immunoprecipitated protein **(Figure S1A).** In contrast to Ste6*C-I2I, the presence of Ubr1p did not increase degradation rates of Ste6*C-I2S in either Δ*san1* or +SAN1 backgrounds (Δ*san1*Δ*ubr1* vs Δ*san1*+UBR1; +SAN1Δ*ubr1* vs +SAN1+UBR1) **(Figure 1C).** Thus, by substituting the P2 Ile of Ste6*C with Ser, Ubr1-mediated degradation of Ste6*C was effectively inhibited. On the other hand, degradation of Ste6*C-I2S in the presence of only San1p remained similar to Ste6*C-I2I **(Figures 1A and 1C**; +SAN1Δ*ubr1***)**.

The inverse experiment was performed with Δ2GFP (Δ2GFP-S2S) by substituting the P2 Ser with Ile, generating Δ2GFP-S2I. Degradation of Δ2GFP-S2I was far more efficient in Δ*san1*+UBR1 cells than Δ*san1*Δ*ubr1* cells, in contrast to Δ2GFP, for which the presence of Ubr1p (Δ*san1*+UBR1) conferred only a minimal increase of degradation efficiency **(Figures 1B and 1D)**. As the case with Ste6*C-I2S, degradation efficiency of Δ2GFP via the San1-mediated pathway remained relatively unaffected by the S2I substitution **(Figures 1B and 1D**; +SAN1Δ*ubr1***).** For both substrates, possessing a P2 Ile as opposed to a P2 Ser increased the overall degradation rate of the substrates in wild-type cells, suggesting that the enhancement of Ubr1-mediated degradation conferred by a P2 Ile occurs under normal physiological conditions **(Figure 1E)**.

We next addressed whether the initiating methionine of Ste6*C-I2I was being cleaved through an unidentified mechanism specific to the processing of misfolded proteins, possibly presenting an exception to the known rules of N-terminal methionine excision (Moerschell et al., 1990; Giglione et al., 2004). In the Arg/N-end rule pathway, isoleucine is a Type-2 destabilizing residue (Varshavksy, 2011; Sriram et al., 2011; Tasaki et al., 2012)**.** Cleavage of the N-terminal methionine of Ste6*C-I2I would expose the isoleucine residue at the P2 position, triggering its recognition by Ubr1 and subsequent degradation via the Arg/N-end rule pathway. To determine if this was the case, we performed N-terminal sequencing on Ste6*C-I2I isolated from Δ*san1*Δ*ubr1* cells. This analysis confirmed that the leading methionine is retained by Ste6*C-I2I, as predicted **(Figure S1B)**. Therefore, recognition of Ste6*C and Δ2GFP-S2I by Ubr1p is through an N-terminal Met-Ile as opposed to an N-terminal Ile, both of which are Arg/N-end rule pathway N-degrons.

We observed that nine amino acids at the P2 position resulted in significant inhibition of Ubr1-mediated degradation of Ste6*C in Δ*san1*+UBR1 cells: Ala, Asn, Ser, Cys, Glu, Pro, Asp, Thr, and Met, which we refer to as Ubr1-QC-*incompatible* P2 amino acids. His, Trp, Phe, Tyr, Ile, Val, Gln, Gly, Lys, Leu, and Arg, allowed rapid degradation by Ubr1, which we refer to as Ubr1-QC-*compatible* P2 amino acids **(Figure 2A, S2A-S2C)**. The same effect was not seen in +SAN1Δ*ubr1* cells, for which no statistically significant change to turnover rate relative to the original Ste6*C substrate was observed **(Figure 2A, S2D-S2F)**. The Ubr1-dependent degradation of the 9 most rapidly degraded Ste6*C P2 mutants was confirmed in Δ*san1*Δ*ubr1* cells **(Figure S3)**. Ste6*C P2 mutants inhibited for Ubr1 degradation was not a result of Ubr1

blockage by N-terminal acetylation since Ste6*C is encoded with a P3 Pro, which is known to prevent N-terminal acetylation (Polevoda and Sherman, 2003). Leading methionine retention is not a prerequisite for compatibility with the Ubr1-QC degradation pathway, as Ste6*C-I2V was demonstrated to be degraded efficiently by Ubr1 and is predicted to undergo methionine cleavage, also confirmed via sequencing **(Figure S1C)**.

Interestingly, P2 His, Lys, Gln, and Arg, all of which cause retention of initiating methionines based on the rules of methionine excision (Sherman et al., 1985), are not predicted to produce N-termini with known N-degrons of the N-end rule pathway (Kim et al, 2014). Likewise, P2 Gly and Val residues, which lead to the cleavage of leading methionines and become exposed have also not been identified as N-degrons. Thus, our experiments demonstrate that P2 His, Lys, Gln, Arg, Gly, and Val residues generate novel N-terminal N-degrons recognized by Ubr1 when present on misfolded substrates. That the Ubr1-compatibility of these P2 residues was not detected with assays involving short peptides indicates that additional allosteric interaction with Ubr1 may take place with a misfolded substrate that increases the range of Ubr1's specificity for N-degrons. Interestingly, the mutation of P2 Leu to Lys in the case of $\Delta ssC^{22\text{-}519}Leu2_{myc}$ inhibited its Ubr1-mediated degradation (Kim et al., 2014). While a concomitant change in acetylation state was not ruled out for that substrate, which would prevent N-terminal recognition by Ubr1, the differences suggest that there are indeed context and substrate specific influences on the N-end rule pathway's activity in CytoQC that need to be further dissected. Thus far, the expanded set of N-degrons discovered using Ste6*C is expected to apply generally to misfolded proteins in the cytosol, but may possibly extend to other targets and physiological states not explored here.

The majority of secretory pathway and mitochondrially-localized proteins encode Ubr1-QC-compatible P2 residues

To determine if there is a bias within specific cellular compartments for proteins to possess P2 residues that are Ubr1-QC-compatible, we utilized data from a genome-wide GFP-fusion based localization study of protein localization spanning 4156 proteins **(Table S2-S4**; Huh et al., 2003**)**. P2 amino acid usage frequency was analyzed for proteins with exclusive nuclear, cytosolic, secretory pathway, or mitochondrial localization **(Tables S5-S8; Refer to Materials and Methods)**. Frequency distribution across all residues was significantly different for secretory (n=559, $p < 1 \times 10^{-6}$, $x^2 = 64.33$, $df = 19$) and mitochondrial proteins (n=458, $p < 1 \times 10^{-55}$, $x^2 = 317.85$, $df = 19$) when compared against cytosolic proteins (n=823) **(Figures 2B-D; Tables S9 and S10)**. Pair-wise chi-square analysis of individual amino acid usage frequencies demonstrated that biases were seen for the mitochondrial and secretory proteins to be encoded with P2 residues that are Ubr1-QC-compatible, with 6 Ubr1-QC-compatible P2 amino acids (Trp, Phe, Ile, Lys, Leu, and Arg) having statistically significant higher percentage usage frequencies in at least one or both of the mitochondrial and secretory pathway protein sets when compared with cytosolic

proteins. An equally important finding is that we observed a significant usage bias *against* 7 amino acids that do *not* enable efficient Ubr1 degradation (Ubr1-QC-*incompatible* P2 residues Ala, Ser, Cys, Glu, Pro, Asp, Thr) in either one or both mitochondrial and secretory pathway protein sets when compared to cytosolic proteins. In contrast, frequency distribution between nuclear and cytosolic proteins had a much less significant difference (nuclear, n=639; cytosolic n = 823; $p > 0.001$, $x^2 = 36.33$, $df = 19$), and only a few significant differences in relative frequencies of individual amino acids. No general bias for or against Ubr1-QC P2 compatibility was observed between the two groups **(Figure 2E; Table S10).**

We determined the total relative abundance of proteins within each localization category based on abundance level data gathered from the global yeast protein GFP-fusion study conducted by Huh and colleagues (refer to Materials and Methods). The majority of nuclear and cytosolically localized protein is encoded with Ubr1-QC-*incompatible* P2 amino acids (84.7% and 70.9%, respectively), while the majority of secretory and mitochondrial protein is associated with Ubr1-QC-*compatible* P2 amino acids (59.7% and 82.3%, respectively) **(Figure 2F; Table S17-S27)**.

The much higher prevalence of Ubr1-QC-compatible proteins in the latter two categories indicated that Ubr1p may be optimized for the degradation of secretory and mitochondrial protein. That they do not natively reside in the cytosol suggested a model in which proteins in these pathways are subject to Ubr1-mediated CytoQC when they fail to translocate. Compartment-specific chaperones and enzymes are important for the native folding processes of secretory and mitochondrial proteins (Hartl et al., 2011; Haynes and Ron, 2010; Stevens and Argon, 1999). Spontaneous folding by mitochondrial proteins in the cytosol has been shown to occur, thereby preventing import (Strobel, 2002). Various cytosolic chaperones are designed to maintain pre-translocated proteins in partially-folded, import-competent states, while the folding efficiency of signal-sequence carrying precursor proteins has been to shown to be significantly lower than for that of their mature, signal-sequence cleaved counterparts (Neupert,1997; Laminet and Pluckthun, 1989). Thus, a translocation failure of mitochondrial and secretory proteins should theoretically generate conditions optimal for recognition and degradation by Ubr1p: availability of an uncleaved N-terminal sequence containing an N-degron accompanied by impaired or inhibited folding.

Mis-translocated secretory and mitochondrial proteins are degraded by the N-end rule pathway

To test the above hypothesis, we examined ATP2Δ1,2,3, a translocation-defective mutant of Atp2p that is import-deficient and degraded in the cytosol (Bedwell et al., 1987). ATP2Δ1,2,3 carries a Ubr1-QC-compatible P2 Val residue. Degradation was dependent on a combination of San1 and Ubr1 **(Figure 3A).** We substituted the P2 Val with Ser and observed an inhibition of Ubr1-mediated degradation **(Figure 3B)**. The P3 Leu residue was mutated to Pro to inhibit

acetylation (Polevoda and Sherman, 2003). Both forms of the substrate were stabilized to similar levels in *Δsan1Δubr1* cells. Degradation was also dependent on Sse1, Ydj1, and Ssa1 and Ssa2 chaperones **(Figure S4)**, hallmarks of UPS-dependent CytoQC. Degradation of wild-type Atp2p in a temperature sensitive mitochondrial-import mutant, *tom40-2*, was dependent on Ubr1 (Krimmer et al., 2001) **(Figure S5).** These results confirm that translocation-deficient Atp2p is an endogenous substrate of Ubr1-mediated CytoQC.

To determine if translocation-defective secretory pathway proteins are also subject to Ubr1 CytoQC, we generated an import-impaired mutant of vacuolar carboxypeptidase Y (CPY), CPY-12iE, in which a Glu residue is inserted at position 12 to disrupt the hydrophobic core of the native CPY signal sequence. Since any mature, vacuole-processed CPY-12iE would prevent accurate analysis of its pre-form levels, we utilized the fact that C-terminal processing of CPY-HA results in HA tag removal upon maturation **(Figure S6).** Degradation of pre-CPY-12iE was significantly reduced by the absence of Ubr1p **(Figure 3C**, *Δsan1*+UBR1 vs *Δsan1Δubr1)*. The degradation of endogenous pre-forms of GAS1 and CPY proteins in ER-import deficient strain *sec63-1* was dependent on Ubr1, demonstrating that Ubr1 is required for the efficient degradation of a subset of mis-translocated endogenous secretory proteins (personal communication with Ng Lab; unpublished data). Both CPY and GAS1 are encoded with Ubr1-QC-compatible P2 residues, Lys and Leu, respectively.

To assess the significance of a substrate's misfoldedness to Ubr1-recognition of the substrate via a Ubr1-QC-compatible P2 residue, we examined wild-type GFP, which is expected to escape CytoQC-based degradation. The P2 Ser of GFP was substituted with Ile to generate GFP-S2I. GFP-S2I was significantly less susceptible to Ubr1 compared to Δ2GFP-S2I, which was degraded efficiently by Ubr1 **(Figure 3F)**. Trypsin digestion confirmed the relative structural stability of GFP-S2I when compared to Δ2GFP-S2I **(Figure 3G).** These results indicate that degradation of a substrate via the N-end rule pathway is significantly enhanced when the substrate is also misfolded. This may be a result of the N-terminus of a folding-compromised substrate being aberrantly exposed, either spatially or temporally, making binding by Ubr1p binding more kinetically favorable. This is similar to the model of protein quality control proposed for proteins possessing Ac/N-degrons. (Shemorry et al., 2013; Hwang et al., 2010). Alternatively, it may be due to an increased availability of a misfolded substrate's polyubiquitination-competent lysine residues, as one of the requirements of Ubr1-dependent degradation is access to a lysine in an unstructured region of the substrate (Varshavsky, 2011).

P2 residues act as cellular location signals that mediate CytoQC via the N-end rule pathway

Studies have shown that mislocalized membrane proteins in the cytosol are targeted for degradation through exposed hydrophobic patches in both yeast and mammalian a cells (Hessa et

al., 2011; Kawahara et al., 2013; Rodrigo-Brenni and Hegde, 2012; Ast et al., 2014; Suzuki & Kawahara, 2016). If mistranslocated proteins *without* Ubr1-QC-compatible P2 residues are primarily membrane proteins, most would be degradable through these pathways and rendered benign. To assess whether Ubr1 N-end rule based CytoQC, along with quality control pathways targeting hydrophobicity, could together provide an effective degradation system for clearing the cytosol of mis-localized proteins, we analyzed a set of 277 ORFs encoding signal sequence-containing proteins used in a previous study to determine the frequency of amino acids encoded at the P2 position in secretory proteins (Forte et al., 2011). This set was filtered for duplicates and other inconsistencies, resulting in 273 proteins which were then categorized as soluble or membrane proteins (refer to Materials and Methods). Strikingly, 87.5% (49/56) of ER proteins encoded with Ubr1-QC-*incompatible* residues are membrane proteins, strongly differing from the full set of signal-sequence bearing proteins, for which only 64.1% (173/270) are membrane proteins ($p < 0.001$, $x^2 = 11.7$) **(Figure 4B; Tables S12-S16)**. An even larger difference was observed when compared to the relative frequency of membrane proteins in the set of Ubr1-QC-compatible P2-residue-encoded proteins (57.9% (124/214); $p < 5 \times 10^{-5}$, $x^2 = 16.8$). These results indicate that the vast majority of ER proteins which are not compatible with Ubr1 are also, at a higher relative frequency, membrane proteins, which is expected to allow their degradation through pathways that target exposed hydrophobicity, such as San1-mediated CytoQC.

Interestingly, statistically significant differences were seen in the relative frequency of proteins encoding Ubr1-QC-compatible P2 residues when comparing the set of soluble proteins with membrane proteins (Chi-square test: $p < 5 \times 10^{-5}$, $x^2 = 16.8$), as well as with the full set of signal-sequence bearing proteins ($p < 0.005$, $x^2 = 9.7$): 92.8% (90/97) of the soluble protein set possess Ubr1-QC-compatible P2 residues, versus 78.8% (215/273) of the full set of signal-sequence bearing proteins, and only 71.7% (124/173) of the membrane proteins. **(Figure 4A; Table S12-S16)**. Thus, while the majority of signal-sequence bearing proteins are susceptible to Ubr1, soluble secretory proteins are particularly suited for degradation through this pathway. In contrast, only ~26.6% of cytosolic proteins possess Ubr1-QC-compatible P2 residues based on an analysis of a representative set of 251 randomly selected cytosolic proteins **(Figure 4A**; **Table S11**; Forte et al., 2011**)**. This minority of cytosolic proteins with Ubr1-QC-compatible P2 residues are expected to fold efficiently in their native environment, thereby escaping targeting by Ubr1 CytoQC, as demonstrated in pulse-chase experiments involving folded vs misfolded GFP **(Figure 3F).**

DISCUSSION

The findings presented here broaden our understanding of how eukaryotic cells ensure the the cytosol is kept in a healthy state mostly clear of foreign actors that could disrupt normal cellular processes such as signaling, protein synthesis, and trafficking. Failures in SRP-mediated ER targeting causes the mistargeting of secretory proteins to the mitochondria, resulting in

mitochondrial dysfunction (Costa et al., 2018), while depletion of nascent polypeptide-associated complex (NAC) results in the incorrect import of mitochondrial proteins into the ER lumen (Gamerdinger et al., 2015). Such aberrant cross-organelle mis-targeting would also be mitigated by the Ubr1 P2-dependent CytoQC pathway, capturing such proteins in the cytosol before mis-targeting is able to occur. We propose that there is an intrinsic pressure for secretory and mitochondrial proteins to be encoded with Ubr1-QC-compatible P2 residues so they can be efficiently degraded in the event of mislocalization as a result of translocation failure. Soluble ER proteins in particular are preferentially targeted by Ubr1. That they would not be recognized by hydrophobicity-based degradation pathways due to their lack of transmembrane domains, and possession of only moderately hydrophobic signal sequences, necessitates Ubr1-QC-compatibility to be efficiently cleared from the cytosol. Together, our findings also suggest that P2 residues are utilized as *de facto* cellular location signals which help to ensure the the fidelity of protein localization. This system of P2-encoded location signaling facilitates N-end rule pathway-mediated degradation of the majority of soluble mis-translocated secretory and mitochondrial proteins in the cytosol. Confoundingly, the localization of Ubr1 is predominantly in the nucleus, where some of its misfolded substrates are imported and ubiquitinated by it (Prasad et al. 2018). A smaller fraction of Ubr1 also operates on substrates in the cytosol itself (Stolz et al., 2013; Heck et al., 2010; Eisele and Wolf, 2008). Whether the majority of mislocalized P2-residue-dependent Ubr1 substrates are engaged by the cytosolic form of Ubr1, or are first trafficked to the nucleus and subsequently recognized by the the nuclear form of Ubr1, remains to be determined. Since there is an even lower usage frequency of Ubr1-QC-compatible P2 residues in the case of nuclear-localized proteins as compared to cytosolically-localized proteins, Ubr1 degradation activity in both compartments would still be congruent with our model for P2-based location signaling and degradation **(Figure 2F)**.

Herpes simplex virus 1-encoded microRNA has been shown promote the accumulation of β-amyloid through the inhibition of Ubr1 activity, while aggregation-prone fragments of neurodegeneration-associated TDP43, Tau, and α-synuclein, are substrates of the N-end rule pathway (Zheng et al., 2018; Brower et al., 2013). An intriguing avenue of research would be to investigate if the P2-residue dependent degradation of translocation-deficient proteins by Ubr1 described here is mirrored in mammalian cells, and if so, whether disruption of the pathway is involved in the accumulation of such detrimental disease factors.

An additional surprising corollary to our discovery is that the N-end rule pathway, through the recognition of P2-encoded cellular location signals, plays a primary role in enforcing the compartmentalization of secretory and mitochondrial proteins that would otherwise adversely affect critical cellular activities in the cytosol. It is widely accepted that mitochondria and plastids in eukaryotes evolved from endosymbiotic partnerships, whereby host prokaryotic cells engulfed ancient aerobic bacteria and gained the benefit of increased energy production, while the new cellular tenants were consequently shielded from the environment (Archibald, 2015).

Over time, gene transfer from endosymbionts to the host cell genome occurred that conferred physiological fitness benefits, and evolved further into complex systems that utilized transit peptides and specialized import machinery to direct host-encoded proteins in the cytosol back to the endosymbiont (Henze, 2001). However, how did early eukaryotes deal with endosymbiont-derived proteins that could not make it to their intended destination? Based on the results presented here, we propose that the N-end rule pathway played a critical role in the evolution of advanced eukaryotic cells by providing a mechanism to destroy proteins intended for endosymbiotic organelles that failed to successfully transit. In this way, the N-end rule pathway may have been the key enabler for eukaryotic cells to fully exploit the benefits of possessing endosymbiotic organelles by compensating for the negative physiological effects associated with mistargeted proteins.

Acknowledgements. I thank my ex-graduate advisor Davis Ng for his expert guidance and exchange of ideas throughout this work. I am also grateful to Nassira Bedford for assistance with several unpublished experiments, and Chengchao Xu for his critical review of the manuscript and invaluable advice. This work was supported by funds from the National University of Singapore Research Scholarship awarded to Anthony Tran.

Author contributions. AT designed the studies, performed the experiments, and wrote the manuscript.

Competing interests. The author declares that there are no competing interests.

FIGURES

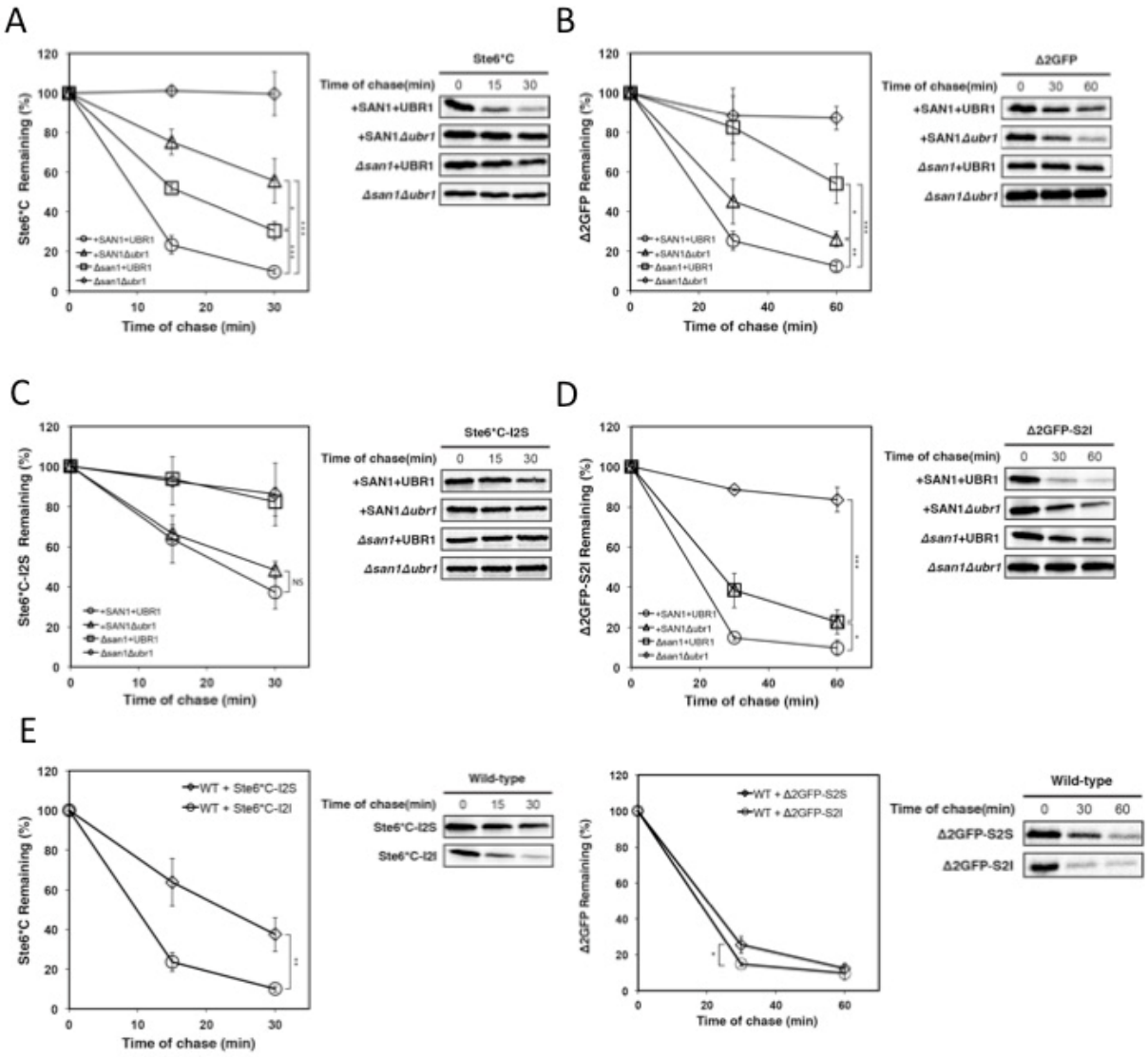

Figure 1. Efficient degradation of Ste6*C and Δ2GFP by the Ubr1 quality control pathway is dependent on P2 residue identity. (Figures adapted from Tran, 2013). (A) and (B) Turnover of Ste6*C and Δ2GFP in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells by metabolic pulse-chase analysis. (C) and (D) Turnover of Ste6*C-I2S and Δ2GFP-S2I in +SAN1+UBR1(wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells was examined by metabolic pulse-chase analysis. (E) Comparison of turnover of P2-serine substrates (Ste6*C-I2S, Δ2GFP-S2S) with P2-isoleucine substrates (Ste6*C-I2I, Δ2GFP-S2I) in wild-type cells by metabolic pulse-chase analysis. (A-E) Metabolic pulse-chase: cells were grown to log phase and shifted to 30°C for 30 minutes followed by pulse-labeling with [35S]methionine/cysteine (5 min for Ste*6C and 10 min for Δ2GFP) and chased at the times indicated. Proteins were immunoprecipitated using anti-HA antibody and resolved by SDS- PAGE, then visualized by phosphoimager analysis. Error bars, mean +/- SD of at least three independent experiments. Student's t-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; Not Significant (NS), $p > 0.5$.

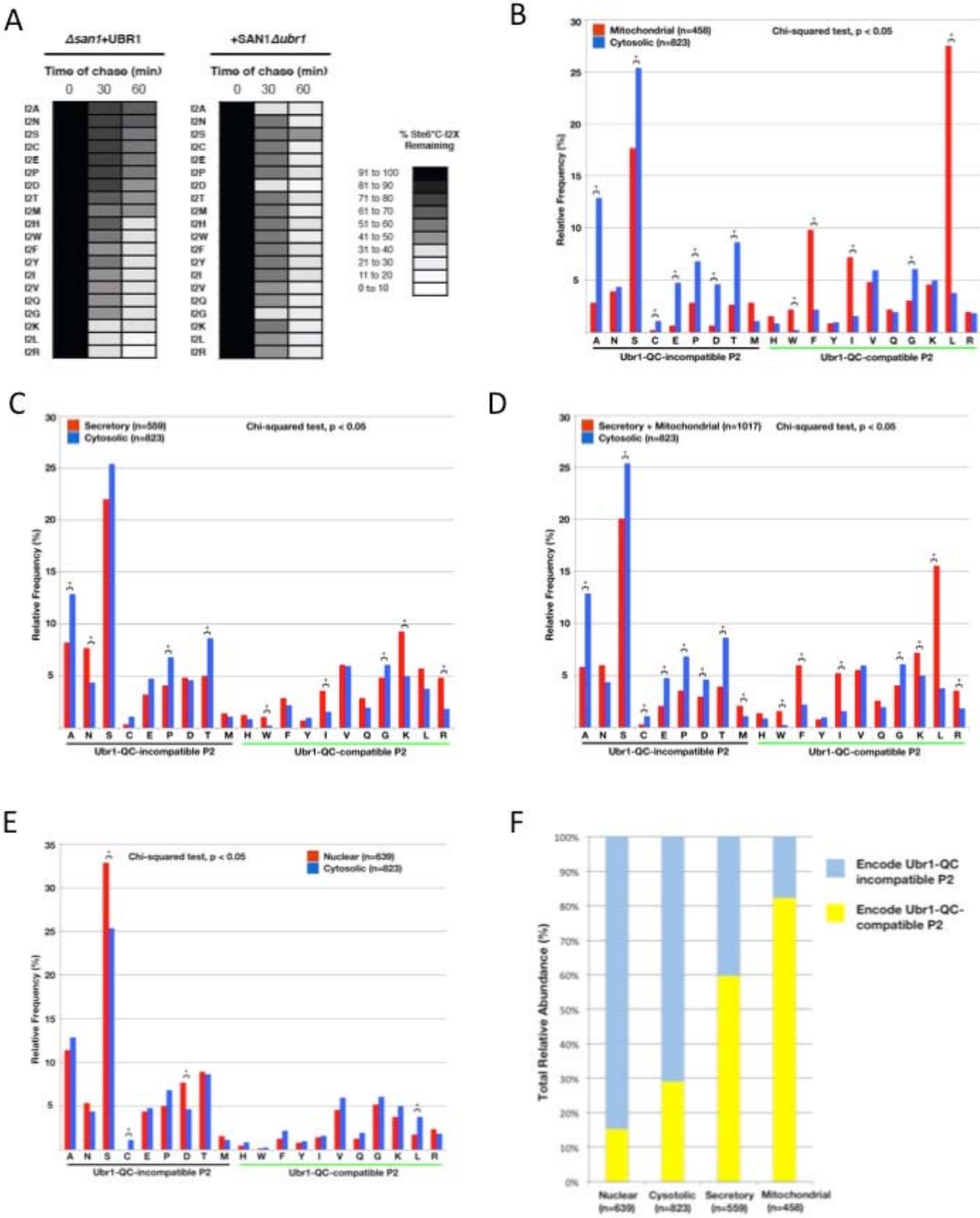
A
Δsan1+UBR1
+SAN1Δubr1
Time of chase (min)
0 30 60
I2A
I2N
I2S
I2C
I2E
I2P
I2D
I2T
I2M
I2H
I2W
I2F
I2Y
I2I
I2V
I2Q
I2G
I2K
I2L
I2R
% Ste6*C-I2X Remaining
91 to 100
81 to 90
71 to 80
61 to 70
51 to 60
41 to 50
31 to 40
21 to 30
11 to 20
0 to 10
B
Mitochondrial (n=458)
Cytosolic (n=823)
Chi-squared test, p < 0.05
Relative Frequency (%)
Ubr1-QC-incompatible P2
Ubr1-QC-compatible P2
C
Secretory (n=559)
Cytosolic (n=823)
Chi-squared test, p < 0.05
D
Secretory + Mitochondrial (n=1017)
Cytosolic (n=823)
Chi-squared test, p < 0.05
E
Nuclear (n=639)
Cytosolic (n=823)
Chi-squared test, p < 0.05
F
Total Relative Abundance (%)
Encode Ubr1-QC incompatible P2
Encode Ubr1-QC-compatible P2
Nuclear (n=639)
Cysotolic (n=823)
Secretory (n=559)
Mitochondrial (n=458)

Figure 2. The P2-residue identity of a CytoQC substrate specifically effects degradation efficiency of the Ubr1 quality control pathway. Proteins localized to secretory pathway components and mitochondria exhibit bias for encoding UBR1-QC-compatible P2 residues when compared to cytosolically-localized proteins. (A) Turnover rates of nineteen Ste6*C P2 amino acid residue mutants (Ste6*C-I2X, where X = A[alanine], N[asparagine], S[serine], C[cysteine], E[glutamic acid], P[proline], D[aspartic acid], T[threonine], M[methionine], H[histidine], W[tryptophan], F[phenylalanine], Y[tyrosine], V[valine], Q[glutamine], G[glycine], K[lysine], L[leucine], R[arginine]) were compared to the turnover rate of the original Ste6*C substrate (Ste6*C-I2I) in Δsan1+UBR1 and +SAN1Δubr1 cells by pulse chase analysis: cells were grown to log phase and shifted to 30°C for 30 minutes followed by pulse-labeling with [35S]methionine/cysteine for 5 minutes and chased at the times indicated. Proteins were immunoprecipitated using anti-HA antibody and resolved by SDS- PAGE, then visualized by phosphoimager analysis. Heat-map reflects results presented in Figures S2A, S2B, S2C, S2D, S2E, and S2F, and is based on the mean of at least three independent experiments. (Figure adapted from Tran, 2013) (B) Relative frequency of P2 amino acid usage of proteins localized to the secretory pathway components (n=559) vs cytosol (n=823). Frequency distribution differed significantly between the two sets ($p < 1 \times 10^{-6}$, $x^2 = 64.33$, $df = 19$; Table S9). Relative usage frequency of amino acids Ala (A), Asn (N), Pro (P), Thr (T), Trp (W), Ile (I), Lys (K), and Arg (R) differed significantly between the two sets (pair-wise Chi-square analysis; $p < 0.01$ for A, N, K, R; $p < 0.05$ for P, T, W, I). (C) Relative frequency of P2 amino acid usage in proteins localized to mitochondria (n=458) vs cytosol (n=823). Frequency distribution differed significantly between the two sets ($p < 1 \times 10^{-55}$, $x^2 = 317.85$, $df = 19$; Table S9). Relative usage frequency of amino acids Ala (A), Ser(S), Cys (C), Glu (E), Pro (P), Asp (D), Thr (T), Trp (W), Phe (F), Ile (I), Gly (G), and Leu (L) differed significantly between the two sets (pair-wise Chi-square analysis; $p < 0.0001$ for A, S, E, P, D, T, F, I, L; $p < 0.05$ for C, W, G) (D) Relative frequency of P2 amino acid usage in proteins localized to either mitochondria or secretory pathway (n=1017) vs cytosol (n=823). Frequency distribution differed significantly between the two sets ($p < 1 \times 10^{-30}$, $x^2 = 196.34$, $df = 19$; Table S10). Relative usage frequency of amino acids Ala (A), Ser (S), Cys (C), Glu (E), Pro (P), Thr (T), Trp (W), Phe (F), Ile (I), Gly (G), Leu (L), Arg (R) differed significantly between the two sets (pair-wise Chi-square analysis; $p < 0.0001$ for A, T, F, I, L, R; $p < 0.01$ for S, E, P, W; $p < 0.05$ for C, G). (E) Relative frequency of P2 amino acid usage in proteins localized to nucleus (n=639) vs cytosol. Frequency distribution differed less significantly between the two sets ($p > 0.005$, $x^2 = 36.33$, $df = 19$; Table S10). Relative usage frequency of amino acids Ser (S), Cys (C), Asp (D), Leu (L) differed significantly between the two sets (pair-wise Chi-square analysis; $p < 0.001$ for S; $p < 0.01$ for C, D, L). (F) Estimated total relative abundance of proteins that are encoded with Ubr1-QC-compatible P2 residues by localization category. Abundance levels based on Huh et al., 2003.

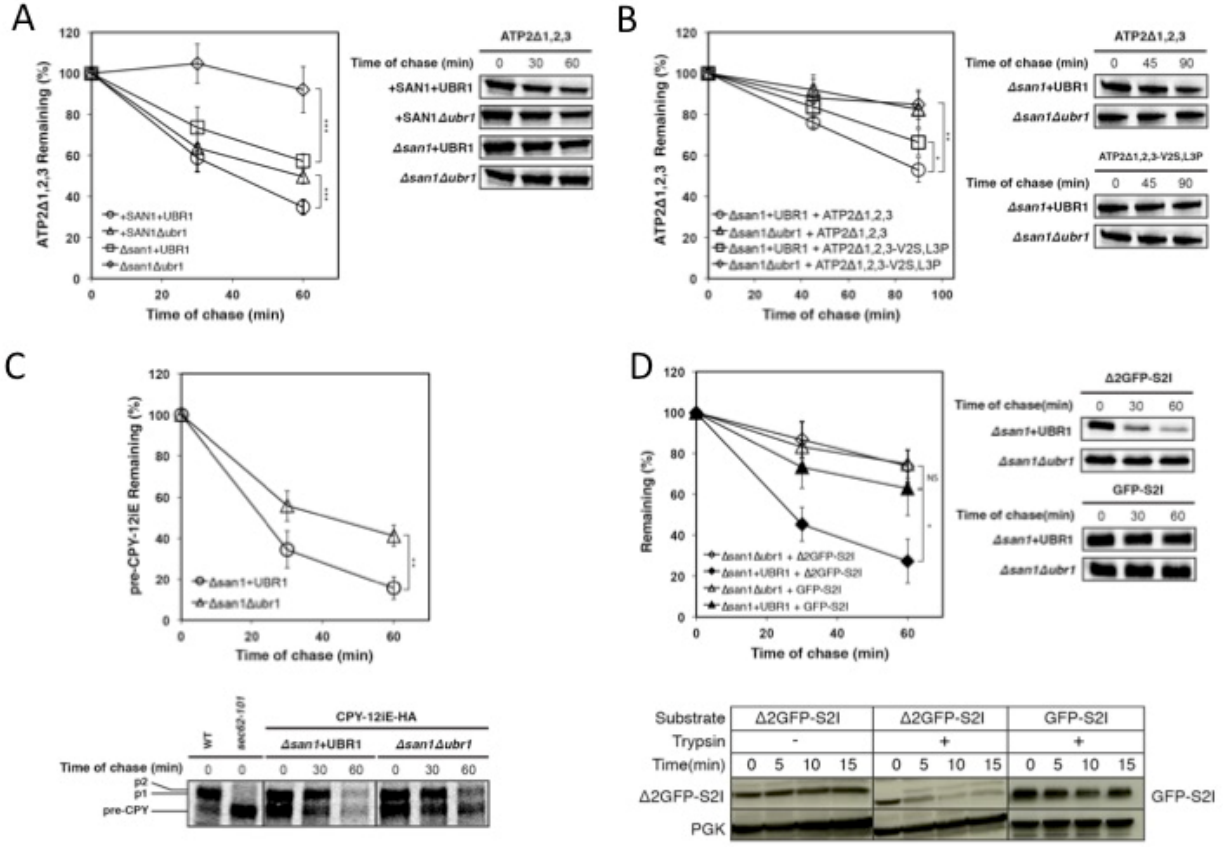

Figure 3. Mistranslocated forms of mitochondrial and secretory proteins are substrates of cytosolic quality control mediated by Ubr1 and the N-end rule pathway. Efficient recognition of CytoQC substrates by Ubr1 is reliant on a bi-partite degradation signal comprising of a Ubr1-QC-compatible P2 residue and structural aberrancy. (Figures adapted from Tran, 2013) (A) ATP2Δ1,2,3 turnover was determined by metabolic pulse-chase analysis in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 strains. (B and C) The turnover of ATP2Δ1,2,3 , ATP2Δ1,2,3-V2S,L3P, or CPY-12iE-HA was determined by metabolic pulse-chase analysis in Δsan1+UBR1, and Δsan1Δubr1 strains. Samples from Δsan1Δubr1 cells expressing CPY-12iE-HA were exposed to a phosphor screen for an extended period (5 days vs 3 days for Δsan1+UBR1 cells) for imaging and quantitative analysis due to a lower overall detection level of the expressed substrate compared to Δsan1+UBR1 cells. (D) Turnover of GFP-S2I and Δ2GFP-S2I was compared by pulse-chase analysis in Δsan1Δubr and Δsan1+UBR1 cells. Trypsin sensitivity assay: postnuclear lysates were prepared from Δsan1Δubr cells and treated with 5.0μg/ml trypsin for the durations indicated. Protein was analyzed by immunoblotting with monoclonal anti-HA antibody. Endogenously expressed phosphoglycerate kinase (PGK) was assayed to serve as a folded protein control. (A-D) Metabolic pulse-chase: Cells were grown to log phase and shifted to 30°C for 30 minutes followed by pulse-labeling with [35S]methionine/cysteine for 10 minutes (ATP2, CPY-12iE-HA) and chased at the times indicated. Proteins were immunoprecipitated using anti-HA antibody and resolved by SDS- PAGE, then visualized by phosphoimager analysis. Error bars, mean +/- SD of at least three independent experiments. Student's t-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; Not Significant (NS), $p > 0.5$.

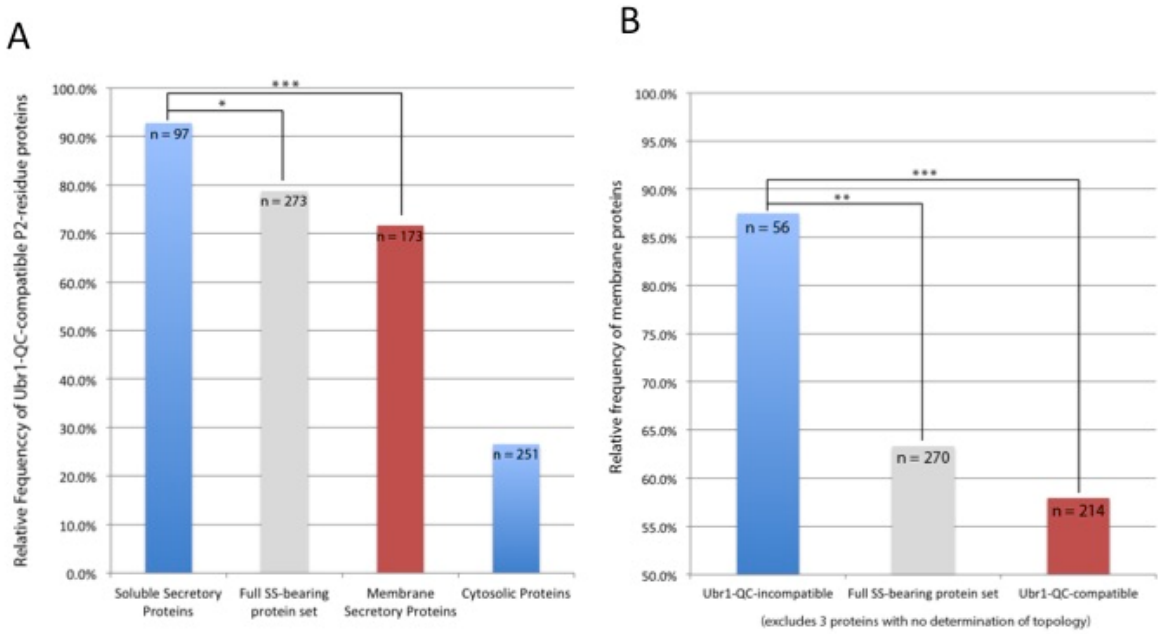

Figure 4. The vast majority secretory proteins encode Ubr1-QC-compatible P2 residues, while the vast majority of secretory proteins encoding Ubr1-QC-*incompatible* P2-residues are membrane proteins. (A) The percentages of proteins encoded with Ubr1-QC-compatible P2 residues was calculated for three categories of proteins derived from the full set of 273 signal-sequence bearing proteins: those determined to be soluble, membranous, and the full set (see Materials and Methods; Tables S12, S14A, S15A, S16). Chi-square test: Soluble vs full set: *, $p < 0.005$, $x^2 = 9.7$, $df = 2$. Soluble vs membrane set: ***, $p < 5 \times 10^{-5}$, $x^2 = 16.8$, $df = 2$. Cytosolic protein set (n = 251) from Forte et al., 2011. (B) The percentages of proteins that are membrane-bound was calculated for three categories derived from the full set of 270 signal-sequence bearing proteins for which topology was determined: those encoded with Ubr1-QC-compatible P2 residues, those encoded with Ubr1-QC-*incompatible* P2 residues, and the full set (see Materials and Methods; Tables S12,S14A, S15A, S16). Chi-square test: Ubr1-incompatible vs full set: **, $p < 0.001$, $x^2 = 11.7$, $df = 2$; Ubr1-incompatible vs Ubr1-compatible: ***, $p < 5 \times 10^{-5}$, $x^2 = 16.8$, $df = 2$.

MATERIALS AND METHODS

Note: All non-bioinformatic biological experimental methods described here are excerpted from the original thesis work performed by Anthony Tran at National University of Singapore for his doctoral dissertation (Tran, 2013) in the lab of Associate Professor Davis Ng at the National University of Singapore.

Yeast strains used in this study are listed in Table S28. Anti-HA monoclonal antibody (HA.11) was sourced from Covance (Princeton, New Jersey). Monoclonal anti-3-phosphoglycerate kinase (PGK) was sourced from Invitrogen (Carslbad, California). Anti-CPY antibody was a gift from Reid Gilmore (University of Massachusetts, Worcester, MA). Anti-Gas1p antibody was raised against amino acids 40 to 289 of Gas1p (Davis Ng, NUS).

Plasmids and primers

Standard cloning procedures were utilized for the construction of plasmids (Sambrook, et al., 1989). Unless otherwise stated, exogenously expressed substrates possess an engineered single hemagglutinin (HA) epitope tag attached to the C-terminus. Ste6*C, Δ2GFP , ATP2Δ1,2,3, and their derivatives, were expressed under control of a high expression, constitutive TDH3 (glyceraldehyde- 3-phosphate dehydrogenase) promoter in yeast centromeric plasmids. HA-tagged Prc1p (CPY/carboxypeptidase Y) and its derivatives were placed under control of its native constitutive endogenous promoter in yeast centromeric plasmids. Site-directed mutagenesis of the original constructs expressing Ste6*C, Δ2GFP, and ATP2Δ1,2,3 was performed to generate mutant substrates.

pAT1: A fragment carrying the PRC1 promoter was PCR amplified from pSW119 with BamHI and NotI restriction ends. The amplified fragment and pSW119 were digested with NotI and BamHI and ligated to generate pAT1.
pAT32: A fragment encoding the TDH3 promoter, followed by Ste6*C-HA, followed by the ACT1 terminator sequence, was PCR amplified from pRP22 with primers AT270 and AT273 and digested with NotI and XhoI. The fragment was then ligated into an empty pRS316 vector to generate pAT32.
pAT33-pAT51: pAT33 through pAT51 (expressing Ste6*C-I2K, Ste6*C-I2Y, Ste6*C-I2F, Ste6*C-I2A, Ste6*C-I2L, Ste6*C-I2E, Ste6*C-I2V, Ste6*C-I2G, Ste6*C-I2R, Ste6*C-I2M, Ste6*C-I2P, Ste6*C-I2W, Ste6*C-I2N, Ste6*C-I2D, Ste6*C-I2H, Ste6*C-I2Q, Ste6*C-I2C, Ste6*C-I2T, Ste6*C-I2S) were constructed by mutation of the base-pairs encoding the 2nd residue of Ste6*C through site-directed mutagenesis using primers AT21-AT39 and pRP22 as a template.
pAT52: A fragment encoding residues 1201-1290 of the STE6 ORF followed by a hemagglutinin epitope (HA-tag) sequence was PCR amplified from yeast genomic DNA using

primers AT40 and AT41. The fragment was digested with BamHI and XbaI and ligated to pRP22 digested with BamHI and XbaI generating pAT52.
pAT55-pAT57: pAT55 through pAT57 (expressing Δ2GFP-S2I, Δ2GFP-S2F, Δ2GFP-S2K) were constructed by mutation of the base-pairs encoding the 2nd residue of Δ2GFP through site-directed mutagenesis using primers AT18-AT20 and pRP44 as a template.
pAT61: A 741bp fragment of the GFP ORF followed by the hemagglutinin (HA) tag sequence was PCR amplified using mutational primers AT189 and AT190 and pAT7 as a template. The resultant PCR product carrying the GFP ORF with the 2nd residue mutated to an isoleucine was digested with BamHI and XbaI and ligated to pAT7 digested with BamHI and XbaI generating pAT61.
pAT64: A 1563-bp fragment carrying the ATP2 ORF followed by the hemagglutinin epitope (HA-tag) sequence was PCR amplified from yeast genomic DNA using primers AT244 and AT226. The fragment was digested with BglII and XbaI and ligated to pAT7 digested with BamHI and XbaI generating pAT64.
pAT65: pAT66 was digested with ClaI and XhoI to release a 2671-bp fragment encoding the TDH3 promoter, ATP2Δ1,2,3-HA, and ACT1 terminator sequences. This fragment was ligated into an empty pRS316 vector digested with ClaI and XhoI to generate pAT65.
pAT66: A 1506-bp fragment carrying the ATP2 ORF with deletions of residues 5- 12, 16-19, and 28-34, followed by the hemagglutinin epitope (HA-tag) sequence was PCR amplified from yeast genomic DNA using primers AT246 and AT226. The fragment was digested with BglII and XbaI and ligated to pAT7 digested with BamHI and XbaI generating pAT66.
pAT68: pAT68 (expressing ATP2Δ1,2,3-V2S,L3P-HA) was constructed by mutation of the sequences encoding the 2nd and 3rd residues of ATP2Δ1,2,3-HA, from valine to serine at the P2 position, and leucine to proline at the 3rd position, through site-directed mutagenesis using primer AT275 and pAT66 as a template.
pAT72: pAT72 (expressing PRC1-12iE) was constructed by the insertion of a three base-pair sequence encoding a glutamic acid residue at the 12th codon position of the PRC1 ORF through site-directed mutagenesis using primer AT280 and pXW92 as a template.

Metabolic Pulse-Chase Assay

Yeast cells were grown to log phase at 30°C (25°C for temperature sensitive strains). 3 OD_{600} units of cells were resuspended in 0.9ml of SC or SC selective media and incubated at 30°C (37°C for temperature sensitive strains) for 30 minutes. Pulse labeling was then initiated with the addition of 82.5 μCi of [35S]Met/Cys (EasyTagTM EXPRESS 35S, PerkinElmer) for 5 or 10 minutes depending on the labeling efficiency of the substrate of interest. Label was chased with the addition of excess cold methionine and cysteine to a final concentration of 2mM. At the appropriate timepoints, pulse labeling/chase was terminated by the addition of 100% TCA to a final concentration of 10%. Immunoprecipitation of samples and resolution by SDS-PAGE were carried out as described (Vashist *et al*., 2001). Phosphor screens exposed to gels (1 to 5 day

exposure, depending on substrate expression level) were scanned with a Typhoon™ phosphoimager and the visualized bands of interest quantified using ImageQuant TL software (GE Healthcare Life Sciences, Uppsala, Sweden). Background signal from screen exposure was subtracted. All results presented are the mean ± SD of at least three independent experiments.

Trypsin Sensitivity Assay

Yeast cells expressing the substrate of interest were grown to log phase (0.4-0.6 OD) and resuspended in cytosol buffer (20 mM HEPES-KOH, pH 7.4, 14 % glycerol, 100 mM KOAc, and 2 mM MgOAc) at a concentration of 20 OD/mL. 1mL of this resuspension was transferred to a 2ml screw-cap tube and homogenized by vortexing for 30 seconds in the presence of 1ml of 0.5mm diameter zirconium beads followed by a 1 minute incubation at 4°C. This was performed for 5 cycles. The homogenate was transferred to 1.5ml eppendorf tubes. 0.6ml of fresh cytosol buffer was used to wash the beads and pooled with the original homogenate. Post-nuclear lysate was isolated by pelleting at 500xg for 5 minutes and transferring the supernatant to a fresh tube. The post-nuclear lysate was incubated at 30°C for 5 minutes, followed by the addition of trypsin to a concentration of 5µg/ml. Samples were vortexed and incubated for 30°C, with 100ul aliquots taken at the indicated timepoints and mixed with 11.1ul of 100% TCA in fresh 1.5ml eppendorf tubes. Aliquots were kept on ice for 5 minutes and pelleted at 14000rpm for 20 minutes at 4°C. Supernatant was discarded, sample pelleted again briefly, and supernatant again discarded. The pellet was resuspended in 10ul of TCA resuspension buffer (100 mM Tris-HCl pH 11.0, 3 % SDS, 1mM PMSF) by cycles of boiling at 100°C and vortexing. Samples were pelleted at 4°C to remove SDS and other insoluble particles, and the resultant supernatant transferred to a fresh tube. Analysis by SDS-PAGE/Western Blotting was performed using the appropriate antibodies.

Edman Degradation N-terminal Seqeuncing

Yeast cells expressing the protein of interest were grown to 1 OD/mL in selective media. 800OD of yeast cells were harvested at 3000xg for 15 minutes, washed once with 1x PBS, and pelleted again at 3000xg. Cells were washed in IP/NP40/PIC/DTT (50mM Tris-HCl, pH8 150mM NaCl, 0.1% NP-40, 0.1mM DTT), pelleted at 3000xg, and resuspended in IP/NP40/DTT containing protease inhibitors (complete, mini, EDTA-free protease inhibitor cocktail tablet, Roche) at a concentration of 50 OD/ml. 1 mL aliquots of the resuspension were transferred to a 2ml screw-cap tubes and homogenized by beadbeating for 30 seconds using a Mini-BeadBeater cell disrupter (Biospec Products) followed by a 5 minute incubation on ice; beadbeating and incubation on ice was repeated for 6 cycles in the presence of 1ml of 0.5mm diameter zirconium beads. The homogenate was transferred to 1.5ml eppendorf tube. 0.6ml of fresh cytosol buffer was used to wash the beads and pooled with the original homogenate. Lysate was cleared by centrifugation at 14000 rpm and transferred to a fresh tube. Lysate was incubated with 65uL of Roche Anti-HA affinity matrix per 4mL of lysate for 2 hours or overnight at 4°C. Affinity matrix

was spun down at 2700rpm for 1 minute and washed with ice cold IP/NP40/PIC/DTT three times, followed by one wash with cold IP buffer to remove residual NP40. Bound proteins were eluted from matrix through the addition of protein loading buffer (PLB) and subsequent boiling at 100°C for 10 minutes. Immunoprecipitated proteins were resolved on 12% SDS-PAGE gels and transferred to a PVDF membrane. Membrane was washed with dIH2O for 1 minute (3 x 1 minute, shaking at 70rpm) to eliminate traces of SDS, Tris, glycine, and other reagents that have the potential to interfere with Edman chemistry. The membrane was then stained with Coomassie Brilliant Blue (0.1% CBB, 5% acetic acid, 50% methanol) for 5 minutes by shaking at 70rpm. Membrane was quickly destained with 50% methanol (3 x 1 minute, shaking at 70rpm). Band containing the protein of interest was excised from the membrane and cut into smaller pieces to facilitate sample analysis. Membrane fragments were loaded into an ABI Procise 494 Sequencer for sequencing using standard manufacturer recommended protocols.

Bioinformatic analysis

The raw data file containing protein translations for systematically-named ORFs was obtained from Saccharomyces Genome Databse (SGD) . To facilitate parsing with PHP, quotation marks in protein descriptions were removed, and a termination character ("@") was added to the end of the file. (http://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/; orf_trans.fasta.gz) A PHP script was written to extract from this data file the systematic name of each ORF and the first 2 residues in its respective protein sequence, and subsequently output into an SQL database; a list of 5887 ORFS and their N-terminal sequence were produced (Table S1). The list of proteins exclusively localized to each of the main protein localization categories (nuclear, mitochondrial, cytoplasmic, or secretory) was determined based on complete or partial localization to that category, as determined by the localization terms assigned to each protein through a GFP-fusion localization method (Huh et al., 2003). Nuclear proteins included those assigned the following localization terms: *nucleus, nucleolous,* and *nuclear periphery*. Secretory proteins included those assigned the following terms: *ER, Golgi, vacuole, endosome,* and *peroxisome*. Proteins assigned the term *mitochondrion* were categorized as mitochondrial. Proteins assigned with the term *cytoplasm* were categorized as cytosolic. To determine the relative total abundance of proteins with nuclear, cytosolic, secretory, or mitochondrial localization that encode Ubr1-QC-compatible or Ubr1-QC-incompatible P2 residues, we summed the abundance levels of proteins in each subset as quantified in the GFP-localization study (Huh et al. 2003), and divided it by the aggregate abundance of all proteins within the full localization category (Table S17-S27).

Categorization of signal-sequence bearing proteins

A published set of 277 signal-sequencing bearing proteins was analyzed (Table S12; Forte et al., 2011). Two of this original set were duplicate entries (AIM6/YDL237W duplicate of LRC1,

FLO11/YIR019C duplicate of MUC1) and were not included in the analysis. Two were not present in the YeastSGD database and were also excluded (YCR012C, and YJL052C) (Table S13). After these exclusions, a manual review was performed on each protein to identify literature that supported soluble or membranous topology. 211 of the proteins were successfully categorized through the review of literature (Tables S14-S16, S14A, S15A). For 59 proteins for which supporting literature could not be found, protein sequences of each were analyzed with SignalP 4.1 using the default optimized parameters. Proteins that were not predicted to possess a signal sequence cleavage site were categorized as membrane proteins. To assess the topology of proteins that were predicted to possess a signal sequence cleavage site, Kyte-Doolittle hydropathy profiles were obtained, and membrane prediction with Phobius performed. Kyte-Doolittle analysis was performed with a window size of 19. Proteins with a hydrophobicity score of greater than 1.8 after the predicted cleavage site were categorized as membrane proteins. Proteins that did not have a hydrophobicity score of greater than 1.8 after the predicted cleavage site were categorized as soluble proteins. 3 proteins which did not have consensus between Kyte-Doolittle and Phobius analysis were excluded from relative frequency calculations involving topology.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request, or are included in the supplemental tables provided.

REFERENCES

1. Archibald, J. Endosymbiosis and Eukaryotic Cell Evolution. Current Biology 25, R911-R921 (2015).
2. Ast, T., Aviram, N., Chuartzman, S. & Schuldiner, M. A cytosolic degradation pathway, prERAD, monitors pre-inserted secretory pathway proteins. *Journal of Cell Science* 127, 3017-3023 (2014).
3. Bedwell, D., Klionsky, D. & Emr, S. The yeast F1-ATPase beta subunit precursor contains functionally redundant mitochondrial protein import information. *Molecular and Cellular Biology* 7, 4038-4047 (1987).
4. Brower, C., Piatkov, K. & Varshavsky, A. Neurodegeneration-Associated Protein Fragments as Short-Lived Substrates of the N-End Rule Pathway. Molecular Cell 50, 161-171 (2013).
5. Buchberger, A., Bukau, B. & Sommer, T. Protein Quality Control in the Cytosol and the Endoplasmic Reticulum: Brothers in Arms. *Molecular Cell* 40, 238-252 (2010).
6. Costa, E., Subramanian, K., Nunnari, J. & Weissman, J. Defining the physiological role of SRP in protein-targeting efficiency and specificity. *Science* 359, 689-692 (2018).

7. Currais, A., Fischer, W., Maher, P. and Schubert, D. (2017). Intraneuronal protein aggregation as a trigger for inflammation and neurodegeneration in the aging brain. *The FASEB Journal*, 31(1), pp.5-10.
8. Dubnikov, T., Ben-Gedalya, T. and Cohen, E. (2016). Protein Quality Control in Health and Disease. *Cold Spring Harbor Perspectives in Biology*, 9(3), p.a023523.
9. Eisele, F. & Wolf, D. Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Letters* 582, 4143-4146 (2008).
10. Fredrickson, E. & Gardner, R. Selective destruction of abnormal proteins by ubiquitin-mediated protein quality control degradation. *Seminars in Cell & Developmental Biology* 23, 530-537 (2012).
11. Fredrickson, E., Gallagher, P., Clowes Candadai, S. & Gardner, R. Substrate Recognition in Nuclear Protein Quality Control Degradation Is Governed by Exposed Hydrophobicity That Correlates with Aggregation and Insolubility. *Journal of Biological Chemistry* 288, 6130-6139 (2013).
12. Fredrickson, E., Rosenbaum, J., Locke, M., Milac, T. & Gardner, R. Exposed hydrophobicity is a key determinant of nuclear quality control degradation. *Molecular Biology of the Cell* 22, 2384-2395 (2011).
13. Forte, G., Pool, M. & Stirling, C. N-Terminal Acetylation Inhibits Protein Targeting to the Endoplasmic Reticulum. *PLoS Biology* 9, e1001073 (2011).
14. Gamerdinger, M., Hanebuth, M., Frickey, T. & Deuerling, E. The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science* 348, 201-207 (2015).
15. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *Cellular and Molecular Life Sciences* 61, (2004).
16. Goldberg, A. Protein degradation and protection against misfolded or damaged proteins. *Nature* 426, 895-899 (2003).
17. Hartl, F., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324-332 (2011).
18. Haynes, C. & Ron, D. The mitochondrial UPR - protecting organelle protein homeostasis. *Journal of Cell Science* 123, 3849-3855 (2010).
19. Heck, J., Cheung, S. & Hampton, R. Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proceedings of the National Academy of Sciences* 107, 1106-1111 (2009).
20. Henze, K. How do mitochondrial genes get into the nucleus?. Trends in Genetics 17, 383-387 (2001).
21. Hessa, T. et al. Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* 475, 394-397 (2011).
22. Hwang, C., Shemorry, A. & Varshavsky, A. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* 327, 973-977 (2010).

23. Huh, W. et al. Global analysis of protein localization in budding yeast. *Nature* 425, 686-691 (2003).
24. Kim, H. et al. The N-Terminal Methionine of Cellular Proteins as a Degradation Signal. *Cell* 156, 158-169 (2014).
25. Kim, Y., Hipp, M., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry* 82, 323-355 (2013).
26. Krimmer, T. et al. Biogenesis of Porin of the Outer Mitochondrial Membrane Involves an Import Pathway via Receptors and the General Import Pore of the Tom Complex. *The Journal of Cell Biology* 152, 289-300 (2001).
27. Laminet, A.A., Plückthun A. The precursor of beta-lactamase: purification, properties and folding kinetics. *The EMBO Journal* 8, 1469-77 (1989).
28. Moerschell R.P., Hosokawa Y., Tsunasawa S., Sherman F. The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes c created by oligonucleotide transformation. *Journal of Biological Chemistry* 265, 19638-19643 (1990).
29. Neupert, W. Protein Import Into Mitochondria. *Annual Review of Biochemistry* 66, 863-917 (1997).
30. Nillegoda, N. et al. Ubr1 and Ubr2 Function in a Quality Control Pathway for Degradation of Unfolded Cytosolic Proteins. *Molecular Biology of the Cell* 21, 2102-2116 (2010).
31. Pearce, M. and Kopito, R. (2017). Prion-Like Characteristics of Polyglutamine-Containing Proteins. *Cold Spring Harbor Perspectives in Medicine*, 8(2), p.a024257.
32. Polevoda, B. & Sherman, F. N-terminal Acetyltransferases and Sequence Requirements for N-terminal Acetylation of Eukaryotic Proteins. *Journal of Molecular Biology* 325, 595-622 (2003).
33. Prasad, R., Kawaguchi, S. & Ng, D. A Nucleus-based Quality Control Mechanism for Cytosolic Proteins. *Molecular Biology of the Cell* 21, 2117-2127 (2010).
34. Prasad, R., Kawaguchi, S. & Ng, D. Biosynthetic mode can determine the mechanism of protein quality control. *Biochemical and Biophysical Research Communications* 425, 689-695 (2012).
35. Prasad, R., Xu, C. & Ng, D. Hsp40/70/110 chaperones adapt nuclear protein quality control to serve cytosolic clients. *The Journal of Cell Biology* jcb.201706091 (2018). doi:10.1083/jcb.201706091
36. Rodrigo-Brenni, M., Gutierrez, E. & Hegde, R. Cytosolic Quality Control of Mislocalized Proteins Requires RNF126 Recruitment to Bag6. *Molecular Cell* 55, 227-237 (2014).
37. Shemorry, A., Hwang, C. & Varshavsky, A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Molecular Cell* 50, 540-551 (2013).

38. Sherman, F., Stewart, J. & Tsunasawa, S. Methionine or not methionine at the beginning of a protein. BioEssays 3, 27-31 (1985).
39. Stevens, F. & Argon, Y. Protein folding in the ER. *Seminars in Cell & Developmental Biology* 10, 443-454 (1999).
40. Stolz, A., Besser, S., Hottmann, H. & Wolf, D. Previously unknown role for the ubiquitin ligase Ubr1 in endoplasmic reticulum-associated protein degradation. *Proceedings of the National Academy of Sciences* 110, 15271-15276 (2013).
41. Szoradi, T. et al. SHRED Is a Regulatory Cascade that Reprograms Ubr1 Substrate Specificity for Enhanced Protein Quality Control during Stress. Molecular Cell 70, 1025-1037.e5 (2018).
42. Strobel, G., Zollner, A., Angermayr, M. & Bandlow, W. Competition of Spontaneous Protein Folding and Mitochondrial Import Causes Dual Subcellular Location of Major Adenylate Kinase. *Molecular Biology of the Cell* 13, 1439-1448 (2002).
43. Suzuki, R. & Kawahara, H. UBQLN4 recognizes mislocalized transmembrane domain proteins and targets these to proteasomal degradation. *EMBO reports* 17, 842-857.
44. Tran, A. N-terminal Determinants of Cytosolic Protein Quality Control. PhD Dissertation. National University of Singapore. http://scholarbank.nus.edu.sg/handle/10635/51986 (2013).
45. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Science* 20, 1298-1345 (2011).
46. Zheng, K. et al. HSV-1-encoded microRNA miR-H1 targets Ubr1 to promote accumulation of neurodegeneration-associated protein. Virus Genes 54, 343-350 (2018).

SUPPLEMENTAL DATA

See PDF and Excel files for Supplemental Tables.

Supplemental Figures S1-S6 (All figures adapted from Tran, 2013)

Edman's Degradation sequencing of Ste6*C-I25 isolated from Δsan1Δubr1 cells

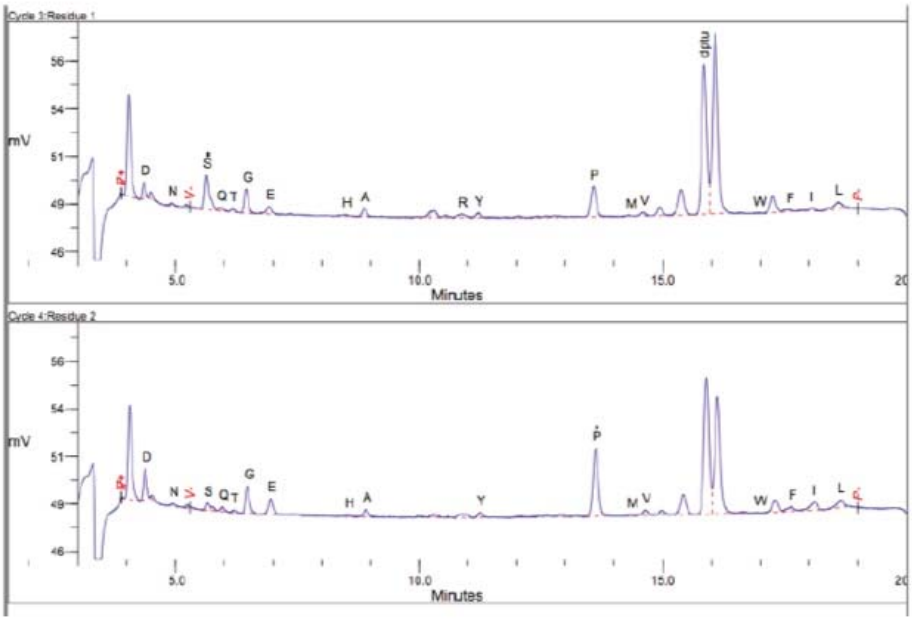

Figure S1A

Edman's Degradation sequencing of Ste6*C isolated from Δsan1Δubr1 cells

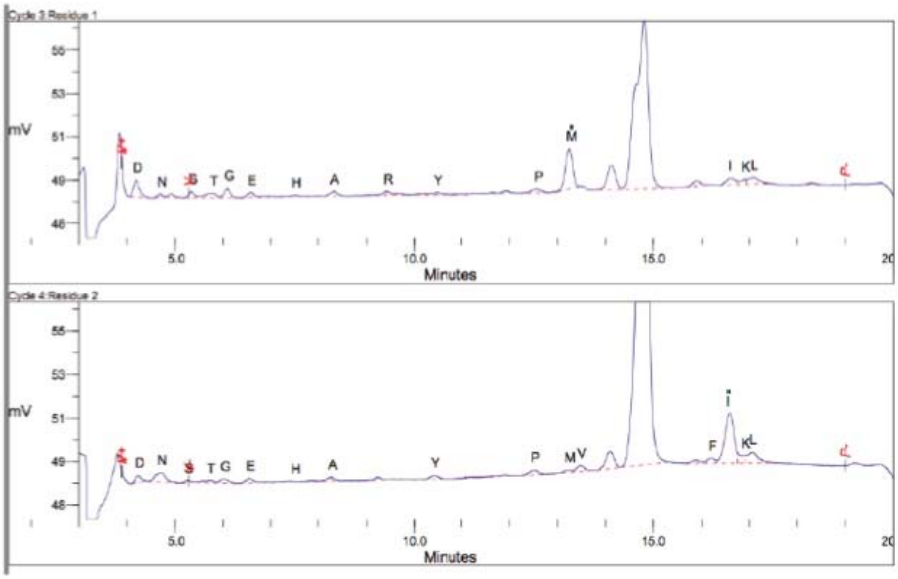

Figure S1B

Edman's Degradation sequencing of Ste6*C-I2V isolated from Δsan1Δubr1 cells

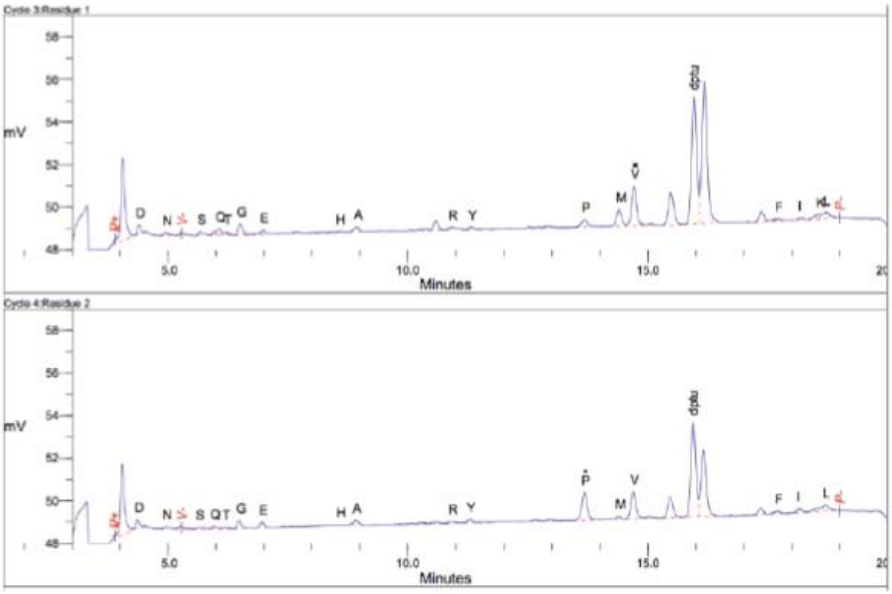

Figure S1C

Figure S1. N-terminal sequencing of Ste6*C constructs. (A) Lysate was prepared from Δsan1Δubr1 cells expressing Ste6*C-HA. Protein was immunoprecipitated from lysate with anti-HA affinity matrix (Roche), resolved by SDS-PAGE, and transferred to PVDF membrane. Protein band carrying Ste6*C-HA was excised from the membrane and sequenced via Edman degradation. (B) Lysate was prepared from Δsan1Δubr1 cells expressing Ste6*C-HA-I2S. Protein was immunoprecipitated from lysate and sequenced as described in Figure S1A. (C) Lysate was prepared from Δsan1Δubr1 cells expressing Ste6*C-HA-I2V. Protein was immunoprecipitated from lysate and sequenced as described in Figure S1A.

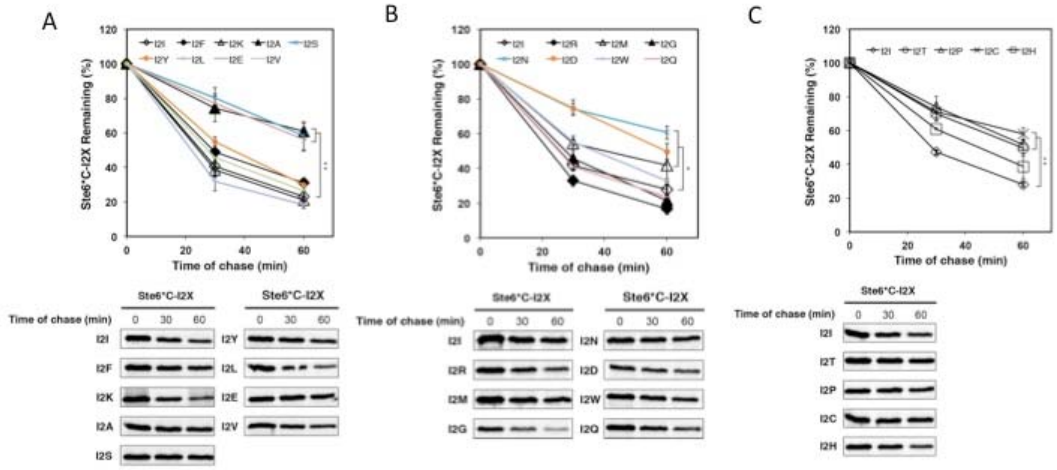

Figure S2

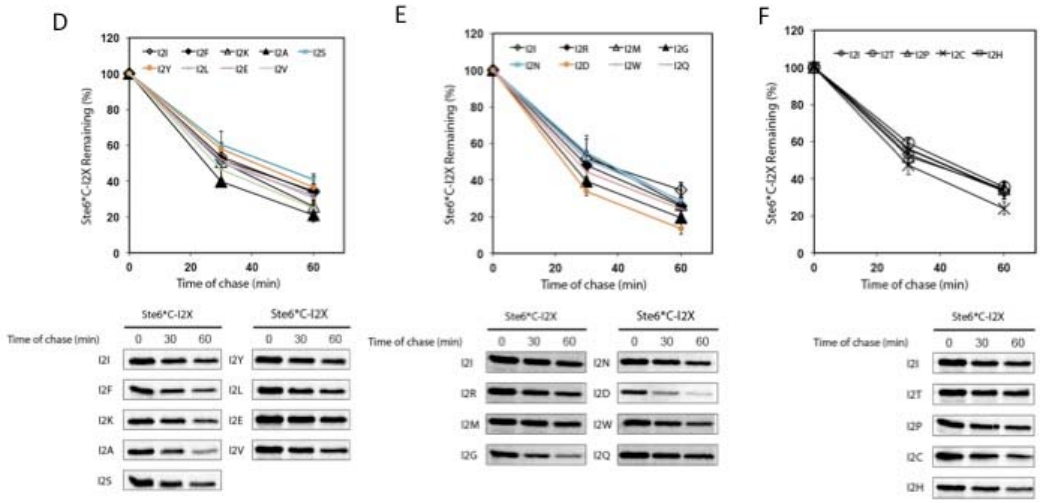

Figure S2

Figure S2. Efficient degradation of Ste6*C by the Ubr1p quality control pathway is highly dependent on P2 residue identity. (A), (B), and (C) Turnover rates of nineteen Ste6*C P2 amino acid residue mutants (Ste6*C-I2X, where X = A[alanine], N[asparagine], S[serine], C[cysteine], E[glutamic acid], P[proline], D[aspartic acid], T[threonine], M[methionine], H[histidine], W[tryptophan], F[phenylalanine], Y[tyrosine], V[valine], Q[glutamine], G[glycine], K[lysine], L[leucine], R[arginine]) was compared to the turnover rate of the original Ste6*C substrate (Ste6*C-I2I) in Δsan1+UBR1 background cells by pulse chase analysis as described in Figure 1A. (D), (E) and (F) Same as described in S2A, but in +SAN1Δubr1 background cells. Error bars, mean +/- SD of at least three independent experiments.

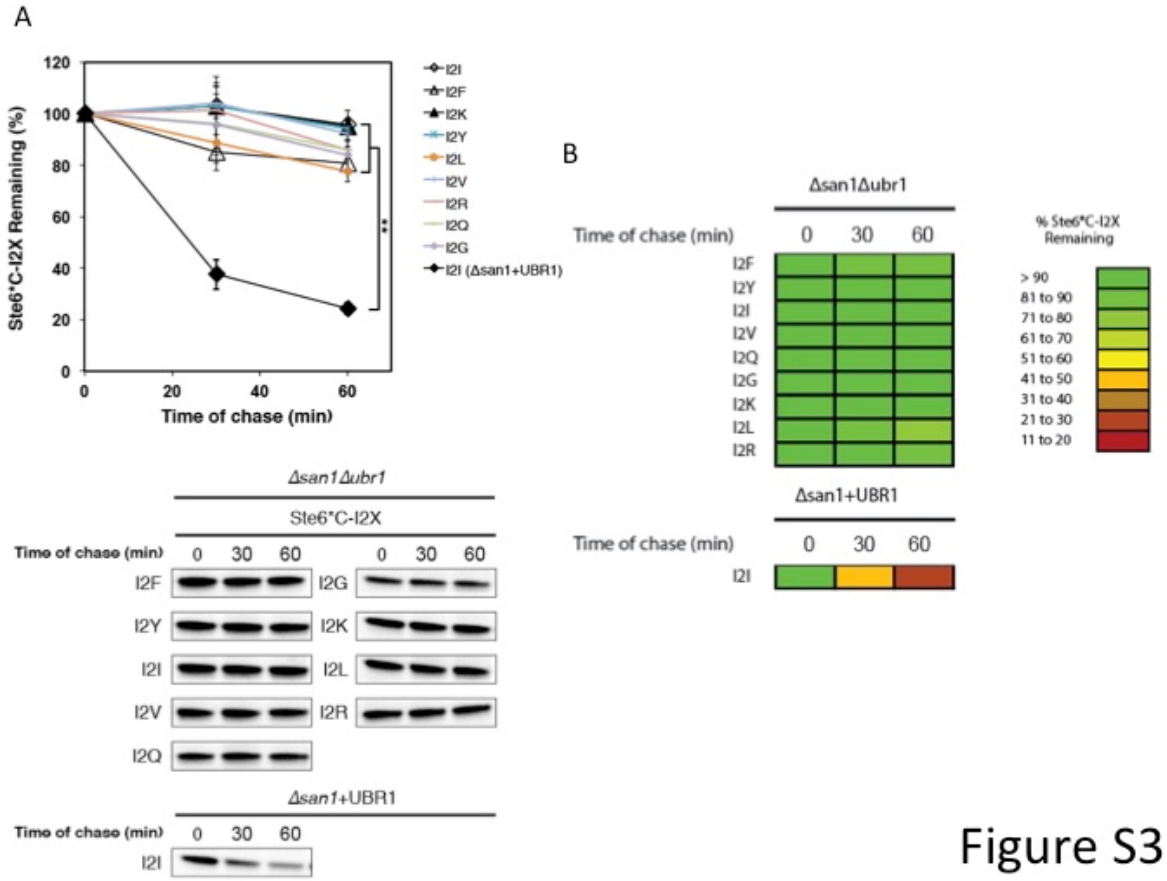

Figure S3. Ste6*C-I2X mutants bearing destabilizing P2 residues are stabilized in Δsan1Δubr1 cells. (A) Turnover rates of N-terminally destabilized Ste6*C-I2X mutants in Δsan1Δubr1 cells were assessed by pulse-chase analysis as described in Figure 1A. Degradation of Ste6*C-I2I in Δsan1+UBR1 cells was also examined as a control. Error bars, mean +/- SD of three independent experiments. (B) Heat map reflecting turnover rates of Ste6*C-I2X substrates in Δsan1Δubr1 as presented in S3A.

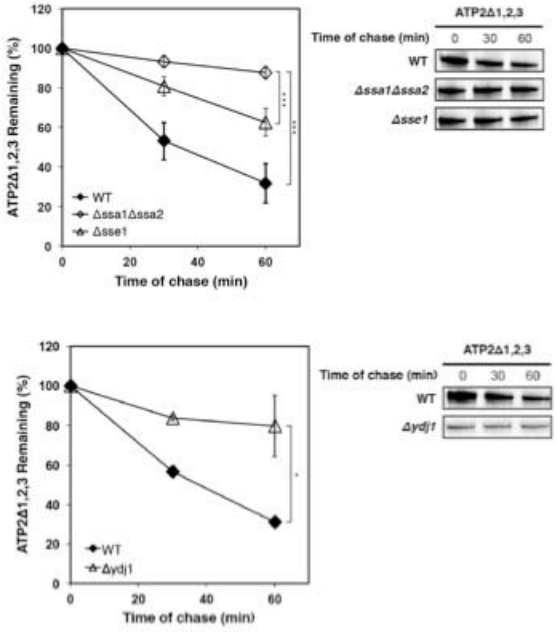

Figure S4

Figure S4. Degradation of ATP2Δ1,2,3 is dependent on chaperones involved in CytoQC degradation. Turnover of ATP2Δ1,2,3 in WT, Δydj1, Δsse1Δsse2, and Δssa1 cells was analyzed by pulse-chase as described in in 5A. Student's t-test: *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; Not Significant (NS), $p > 0.5$. Error bars, mean +/- SD of at least three independent experiments.

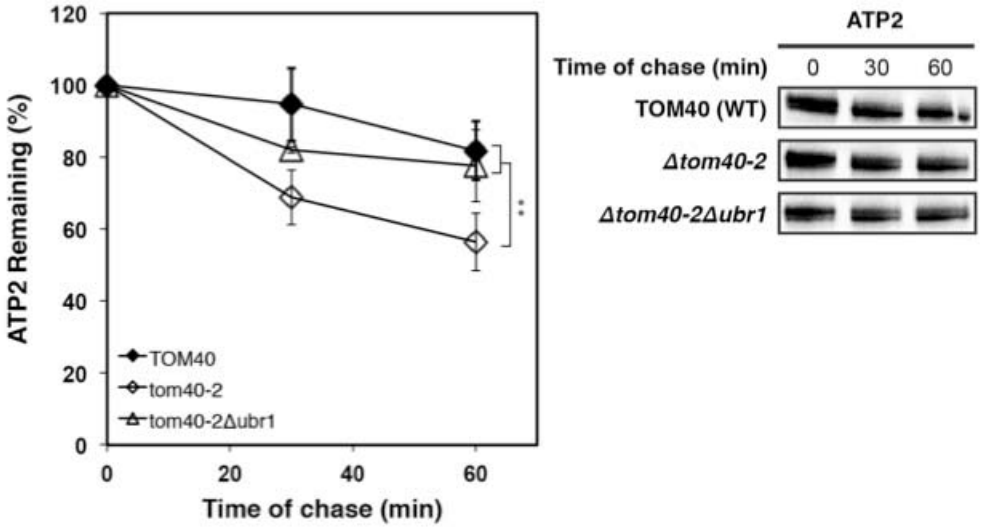

Figure S5

Figure S5. Wild-type Atp2p is degraded by Ubr1p in mitochondrial import mutant strain *tom40-2*. HA-tagged wild-type ATP2 protein expressed in TOM40 (wild-type), tom40-2, and tom40-2Δubr1 cells was analyzed by pulse-chase analysis as described in Figure 1A. Error bars, mean +/- SD of at least three independent experiments.

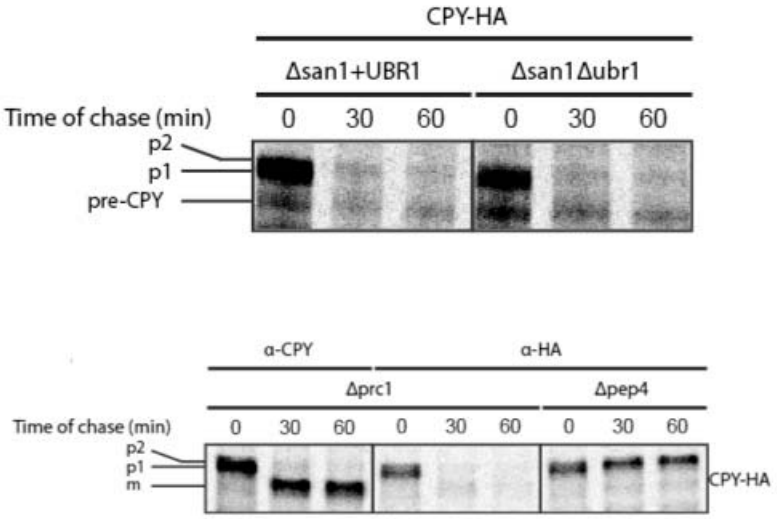

Figure S6

Figure S6. C-terminal HA-tag is removed from HA-tagged CPY in the vacuole. Pulse-chase analysis of CPY-HA was performed in Δprc1, Δpep4, Δsan1+UBR1, and Δsan1Δubr1 cells as described in Figure 1A. CPY was probed with anti-CPY and anti-HA antibodies in Δprc1 cells, and anti-HA in Δpep4 cells. Error bars, mean +/- SD of at least three independent experiments.