The N-end Rule Pathway and Ubr1 enforce protein compartmentalization via P2-encoded cellular location signals

Authors: Anthony Tran$^{12*}$

Affiliations: $^1$National University of Singapore, Department of Biological Sciences, Singapore 117604

Current Address: $^2$5200 Iron Horse Parkway, Dublin, CA 94568

*Correspondence to: anthonytran17@gmail.com

Keywords: N-end rule pathway, Ubr1, ubiquitin-mediated degradation, proteolysis, protein quality control, protein misfolding, CytoQC, N-degron
**Summary Statement:**
The N-end Rule Pathway and Ubr1 enforce the compartmentalization of secretory and mitochondrial proteins through the recognition of cellular location signals embedded into the P2-residues of most proteins.

**Abstract:**
The Arg/N-end Rule Pathway and Ubr1, an E3 ligase conserved from yeast to humans, is involved in the degradation of misfolded proteins in the cytosol. However, the root physiological purpose of this activity is not completely understood. Through a systematic examination of single residue P2-position mutants of misfolded proteins, and global and targeted bioinformatic analyses of the yeast proteome, we determined that Ubr1 preferentially targets mistranslocated secretory and mitochondrial proteins in the cytosol. Degradation by Ubr1 is dependent on the recognition of cellular location signals that are naturally embedded into the 2nd amino acid residue of most proteins. This P2-encoded location signaling mechanism may shed light on how Ubr1 and the N-end rule pathway are involved in neurodegenerative diseases such as Alzheimer’s and Parkinson’s. A corollary to this discovery is that the N-end rule pathway enforces the compartmentalization of secretory and mitochondrial proteins by degrading those that fail to reach their intended sub-cellular locations. The N-end rule pathway is therefore likely to have been critical to the evolution of endosymbiotic relationships that paved the way for advanced eukaryotic cellular life.
**List of Abbreviations:**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPY</td>
<td>Carboxypeptidase Y</td>
</tr>
<tr>
<td>CytoQC</td>
<td>Cytosolic protein quality control</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>MAP</td>
<td>Methionine aminopeptidase</td>
</tr>
<tr>
<td>P2</td>
<td>Position 2</td>
</tr>
<tr>
<td>P2CLS</td>
<td>P2-encoded cellular location signal</td>
</tr>
<tr>
<td>PGK</td>
<td>Phosphoglycerate kinase</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethylsulphonylfluoride</td>
</tr>
<tr>
<td>PQC</td>
<td>Protein quality control</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
</tr>
<tr>
<td>UPS</td>
<td>Ubiquitin-proteasome system</td>
</tr>
</tbody>
</table>
Main Text:
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 et al., 2018; Dubnikov et al., 2017). Cytosolic quality control (CytoQC) pathways, a subset of PQC, specifically mediate the clearance of cytosolically-localized aberrant proteins. In yeast, several E3s have been implicated in CytoQC pathways with varying client substrate scopes and enzymatic activities that are in several cases also dependent on specific physiological insults, such as heat or chemical stresses (Szoradi et al., 2018; Fredrickson et al., 2012; Prasad et al., 2010; Heck et al., 2010; Fang et al., 2011; Buchberger et al., 2010). The San1 and Ubr1 E3 ligases have partially overlapping specificities in the degradation of several cytosolic model misfolded substrates (Prasad et al., 2010; Heck et al., 2010). San1p targets substrates with exposed hydrophobicity consisting of contiguous sequences of at least five or more hydrophobic residues, and was first identified as an E3 ligase that mediates quality control of nuclear proteins (Fredrickson et al., 2013; Fredrickson et al., 2011). Ubr1 is best known for its role as an N-recognin in the N-end rule pathway, which relates the half-life of a protein to its N-terminal residue (Varshavsky, 2011). Ubr1p possesses two domains, the UBR box and N-domain, which recognize and bind to proteins with compatible N-terminal residues, leading to their ubiquitin-dependent degradation. In eukaryotes, four branches of the N-end rule pathway exist: Arg/N-end rule, Ac/N-end rule, Pro/N-end rule, and fMet/N-end rule. Ubr1p mediates the Arg/N-end branch, while the E3 ligases Doa10p and Not4p, Gid4p, and Psh1p are N-recognins of the latter three branches, respectively. Ubr1p recognizes and degrades proteins that possess one of several destabilizing N-terminal residues, described in more detail below. Doa10p and Not4p recognize N-terminally acetylated proteins. Gid4p targets proteins with N-terminal proline residues when present with specific adjacent sequence motifs. Psh1p targets proteins with formylated N-terminal methionine residues in the cytosol. Together, these distinct branches have been demonstrated to partake in regulating a wide range of biological processes, including oxygen sensing, DNA replication, autophagy, cell migration, and nutritional stress responses (Kim et al., 2018; Chen et al., 2017; Shemorry et al., 2013; Varshavsky, 2011).
While there had been earlier conflicting reports on the role of the N-end rule pathway in Ubr1-mediated CytoQC (Nillegoda et al., 2010; Heck et al., 2010), it was later demonstrated that ΔssC\textsuperscript{22-519}Leu\textsubscript{2}myc, a model misfolded cytosolic substrate, was recognized by Ubr1 through its N-terminal Met-Ile, a Met-Φ N-degron, defined as an N-terminal Met followed by a Leu, Phe, Tyr, Trp, or Ile residue at position 2 (P2) (Kim et al., 2014; Eisele and Wolf, 2008). (Note: The 2\textsuperscript{nd} residue of a nascent protein is traditionally referred to as the P1’ or penultimate residue (Wingfield, 2017). It has been referred to as P2 in other literature, and will be referred to as P2 in this report). The latter finding definitively linked Ubr1-mediated CytoQC with N-degrons of the Arg/N-end rule for a subset of substrates.

The P2 position residue of a nascent protein is the main determinant of whether N-terminal methionine excision occurs, and thus dictates the N-terminal sequence of a protein prior to downstream processing steps such as N-terminal acetylation (Nt-acetylation) or signal sequence cleavage (Wingfield, 2017; Giglione et al., 2004; Sherman et al., 1985). Therefore, absent other N-terminal modifications such as acetylation, which blocks N-terminal recognition by Ubr1, the P2-residue of a protein is the primary factor for whether the native, non-post-translationally modified N-terminal sequence of a nascent protein is compatible with recognition by the Arg/N-end rule after translation in the cytosol. The enzyme responsible for N-terminal methionine excision is methionine aminopeptidase (MAP), which cleaves initiating methionines off nascent proteins co-translationally when the side-chain of the P2-residue has a radius of gyration of 1.29 Å or less. Thus, the initiating Met is generally cleaved when the P2-residue is Gly, Ala, Ser, Thr, Cys, Pro, or Val. Residues exposed after Met cleavage are then typically acetylated by an N-terminal acetyltransferase. Consequentially, N-terminally acetylated proteins are vulnerable to degradation by the Ac/N-end rule. Though acetylation effectively blocks N-terminal recognition by Ubr1, it is not of consequence in the case of small P2-residues since they are not destabilizing, aside from P2 Cys, which only acts as a destabilizing residue under conditions where it is oxidized. However, acetylation of un-cleaved N-terminal Met often occurs, which blocks recognition of some substrates possessing Met-Φ N-degrons by Ubr1 (Kim et al., 2014).

Despite recent confirmations that recognition of some misfolded substrates in the cytosol by Ubr1 is N-degron-dependent, the full scope and underlying physiological role of the Arg/N-end
rule in the context of CytoQC is not clearly understood. To help decipher it, we examine the effect that P2-residue identity has on the degradation of misfolded proteins by Ubr1. We characterize the array of P2-residues which produce an expanded set of Ubr1 compatible N-degrons specific to CytoQC by conducting a systematic biochemical analyses of P2-residue variants of a model misfolded CytoQC substrate, Ste6*C (Prasad et al., 2012). Ste6*C was selected as an optimal test substrate because it possesses a P3 proline, which prevents Nt-acetylation for Met-cleaved proteins, and partially inhibits Nt-acetylation for Met-retaining proteins (Lange et al., 2009; Polevoda and Sherman, 2003). This allows us to remove Nt-acetylation as a possible factor for any inhibition of Ubr1-mediated degradation that is observed for Ste6*C P2 mutants that are Met-cleaved. It also greatly increases the likelihood that any Ubr1-compatible N-degrons resulting from a P2 mutation will remain detectable. The expanded set of N-degrons we uncover differs from the original set of destabilizing Arg/N-end rule residues, which only included single position primary destabilizing residues exposed at the N-terminus (basic residues - His, Lys, Arg; bulky hydrophobic residues - Leu, Phe, Trp, Tyr, Ile), or secondary (Asp, Glu, Cys) and tertiary (Asn, Gln) destabilizing residues, to which a primary destabilizing Arg residue is ultimately appended as a result of a one or two step N-terminal modification process. Instead, the expanded N-degron set we define here partially overlaps with the recently discovered Met-Φ N-degrons, which entail a leading methionine residue followed by a compatible bulky hydrophobic P2-residue (Kim et al., 2014; Tran, 2013). While the originally characterized single position Arg/N-end rule destabilizing residues generally do not exist at the N-termini of proteins in their native cytosolic or pre-translocated forms because their size prevents Met-cleavage, Met-Φ N-degrons are naturally abundant, as all proteins are first translated with a leading methionine. Our approach allowed us to identify, in vivo, additional N-degrons relevant to misfolded proteins that were not previously detected using peptide arrays on membrane support (SPOT), which utilizes short peptides for in vitro binding assays. Next, we perform detailed global bioinformatic analyses of P2 amino acid usage frequencies in the S. cerevisiae proteome, considering the expanded P2-dependent N-degron set, and discover that Ubr1’s specificity is designed to preferentially degrade secretory and mitochondrial proteins that fail to translocate and become mislocalized to the cytosol. Our precise in vivo analyses of turnover rates of endogenous translocation-inhibited secretory and mitochondrial proteins, in the presence and absence of Ubr1, confirm this activity. Finally, our analysis of a representative set
of signal-sequence bearing proteins reveals 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 facilitate the rapid degradation of displaced proteins in the cytosol. Together, our findings demonstrate that the N-terminal P2-encoding of cellular location and the N-end rule pathway serve essential cognate roles in proofreading of the cytosolic proteome for abnormalities while simultaneously enforcing the fidelity of protein compartmentalization in eukaryotic cells.

Results

P2-residue identity enables or inhibits degradation of misfolded substrates by Ubr1

Ste6*C is a truncated form of Ste6p, a plasma membrane ATP-binding cassette transporter, containing only a truncated cytosolic tail of the protein (residues 1002-1248), void of the native C-terminus (residues 1249-1290 deleted). 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 a combination of the Ubr1 and San1 CytoQC pathways, albeit with different levels of dependency for each pathway (Prasad et al., 2012; Prasad et al., 2010). To begin investigating the effect of P2-residue identity on CytoQC degradation, we first experimentally reconfirmed that the efficient degradation of Ste6*C and Δ2GFP 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 (Fig 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 an N-terminal Ser residue upon methionine excision. N-terminal Met-Ile, an N-degron of the Arg/N-end rule, should help direct Ste6*C towards the Ubr1 pathway, while Δ2GFP, with a stabilizing N-terminal Ser, should be shielded from Ubr1 N-degron recognition. Since the N-terminal sequence of a misfolded protein impacts its recognition by the Ubr1 pathway, we surmised that by altering the N-terminus of Ste6*C and Δ2GFP, it would be possible to modulate the amount of degradation which proceeds via Ubr1. We generated P2 mutants of Ste6*C and Δ2GFP that matched the P2-residue of the other to assess the impact on Ubr1-mediated degradation. We modified the N-terminus of
Ste6*C to mimic that of ∆2GFP 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 (Fig S1A). In contrast to Ste6*C-I2I, the presence of Ubr1p did not increase degradation rates of Ste6*C-I2S in either Δsan1 or +SAN1 backgrounds (Δsan1Δubr1 vs Δsan1+UBR1; +SAN1Δubr1 vs +SAN1+UBR1) (Fig 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 (Fig 1A,C; +SAN1Δubr1).

The inverse experiment was performed with ∆2GFP (∆2GFP-S2S) by substituting the P2 Ser with Ile, generating ∆2GFP-S2I. Degradation of ∆2GFP-S2I was far more efficient in Δsan1+UBR1 cells than Δsan1Δubr1 cells, in contrast to ∆2GFP, for which the presence of Ubr1p (Δsan1+UBR1) conferred only a minimal increase of degradation efficiency (Fig 1B,D). As the case with Ste6*C-I2S, degradation efficiency of ∆2GFP via the San1-mediated pathway remained relatively unaffected by the S2I substitution (Fig 1B,D; +SAN1Δubr1). 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 (Fig 1E).

In the Arg/N-end rule pathway, both N-terminal Met-Ile and and N-terminal Ile are destabilizing. Therefore, while Ubr1 could potentially recognize Ste6*C-I2I through its native N-terminal Met-Ile, an unconventional Met-cleavage would expose the P2 isoleucine residue and enable recognition via an N-terminal Ile instead. To determine which of the two possibilities was occurring, we performed N-terminal sequencing on Ste6*C-I2I isolated from Δsan1Δubr1 cells. Sequencing confirmed that the leading methionine is retained by Ste6*C-I2I, as predicted (Fig S1B). Therefore, recognition of Ste6*C by Ubr1p is through an N-terminal Met-Ile as opposed to an N-terminal Ile.

We observed that nine amino acids at the P2 position resulted in significant inhibition of Ubr1-mediated degradation of Ste6*C in Δsan1+UBR1 cells: Ala, Asn, Ser, Cys, Glu, Pro, Asp, Thr, and Met, which we refer to as Ubr1-QC-incompatible P2 amino acid residues. 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 acid residues (Figs 2A,B,C,G). The same effect was not seen in +SAN1Δubr1 cells, for which no statistically significant decrease of turnover rate relative to the original Ste6*C substrate was observed (Fig 2D,E,F,G). The Ubr1-dependent degradation of the 9 most rapidly degraded Ste6*C P2 mutants was confirmed in Δsan1Δubr1 cells (Fig 3A,B). Ste6*C P2 mutants inhibited for Ubr1 degradation was not a result of Ubr1 blockage by Nt-acetylation since Ste6*C is encoded with a P3 proline, which prevents Nt-acetylation. 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 (Fig S1C).

Interestingly, P2 His, Lys, Arg, and Gln, all of which cause retention of initiating methionines based on the rules of methionine excision, are not predicted to produce N-termini with known N-degrons of the N-end rule. It is important to note that unconventional N-terminal methionine excision has been demonstrated to occur for some reporter substrates and endogenous proteins with P2 Asn, Gln, and His residues, for which Ubr1-mediated turnover was discovered to be reliant on the atypical removal of the initiating methionine residues, and resultant exposure, N- amidation and Nt-arginylation of the P2-residues in the case of P2 Asn and Gln, or simply the exposure of the P2-residue in the case of P2 His (Nguyen et al., 2019; Kats et al., 2018; Sadis et al., 1995). It is therefore also a possibility that the Ste6*C mutants possessing P2 His, Lys, Arg, or Gln, undergo non-canonical methionine cleavage, resulting in N-terminally exposed His, Lys or Arg, primary destabilizing residues, or Gln, a tertiary destabilizing residue (Varshavsky et al., 2011). Unconventional initiating methionine removal mechanisms in the case of P2 Lys and Arg residues have yet to be demonstrated. Thus, our experiments suggest that P2 His, Lys, Arg, and Gln residues on a misfolded substrate may either induce non-conventional N-terminal methionine excision that produce classical N-degrons, or alternatively, generate novel N-degrons. Further analysis of individual Ste6*C variants would help to determine which specific N-terminal processing mechanisms, if any, lead to their recognition and degradation by Ubr1.

P2 Gly and Val residues are predicted to permit leading methionine removal and become exposed at the N-terminus. However, N-terminal Gly and Val have not been characterized as N-
degrons. Interestingly, Met-Gly and Met-Val N-termini have also not been previously demonstrated to function as N-degrons. Therefore, even in the unexpected events where the leading methionines are retained for these variants, our experiments with Ste6*C suggest that Gly or Val encoded at the P2 position results in novel N-degrons that are compatible with Ubr1.

That the Ubr1-compatibility of P2 His, Lys, Arg, Gln, with the initiating methionine retained, or P2 Gly and Val, with the initiating methionine removed, was not detected in assays involving short peptides in previous studies indicates that additional allosteric interaction with Ubr1 may be taking place in the case of misfolded substrates that increases the range of Ubr1’s specificity for N-degron sequences. The reason for this extended specificity is unclear. Again, an alternative explanation for the compatibility of P2 His, Lys, Arg, or Gln with Ubr1 in our experiments is that non-conventional N-terminal methionine excision, and consequent exposure of the P2-residues, might occur under physiological conditions, but does not occur in *in vitro* assays (Kim et al., 2014). It is worth noting that in a high-throughput analysis of N-degrons utilizing mCherry-sfGFP as a reporter, while reporters with Met-Φ degrons were on average less stable than other variants, the observed destabilization was independent of Ubr1 (Kats et al., 2018). mCherry-sfGFP reporters with P2 His, Lys, Arg, Gln, Gly, and Val were moderately or highly stable. The lack of correlation between the results obtained with mCherry-sfGFP reporters and those achieved using our misfolded substrates suggests that conformational aberrancy may be a key determinant in N-degron compatibility with Ubr1, which we explore later.

Interestingly, the mutation of P2 Leu to Lys in the case of ΔssC22-519Leu2myc inhibited its Ubr1-mediated degradation (Kim et al., 2014), in contrast to Ste6*C, for which both P2 Leu and Lys were compatible with Ubr1 degradation (*Fig 2A,G*). However, as this mutation was assessed in wild-type cells in the cited study, and not in N-terminal acetyltransferase deficient cells, it is possible that acetylation of the substrate’s N-terminus was induced by the mutation, thus inhibiting recognition by Ubr1. So although a concomitant change in acetylation state was not ruled out for ΔssC22-519Leu2myc as a result of the P2-mutation, the difference in outcomes suggest that there are indeed context and substrate specific influences on P2-residue/Ubr1 compatibility, and the Arg/N-end rule’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 extend to other targets and physiological states not explored here.

The majority of secretory and mitochondrial 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 (Huh et al., 2003) (Tables S2-S4). P2 amino acid usage frequency was analyzed for proteins with exclusive nuclear, cytosolic, secretory pathway, or mitochondrial localization (Tables S1, S5-S8; refer to Materials and Methods). Frequency distribution across all residues was significantly different for secretory (n=559, p < 1 x 10^(-6), \(x^2 = 64.33, df = 19\)) and mitochondrial proteins (n=458, p < 1 x 10^(-55), \(x^2 = 317.85, df = 19\)) when compared against cytosolic proteins (n=823) (Fig 4A,B,C; Tables S9, 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, \(x^2 = 36.33, df = 19\), with only four significant differences in relative frequencies of individual amino acids, and no apparent bias for or against Ubr1-QC P2 compatible residues (Fig 4D; Table S10).

We determined the total relative abundance of proteins within each localization category based on abundance level data also gathered from the study conducted by Huh and colleagues (Huh et al., 2003) (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) (Fig 4E; Tables S19-S29).

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 et al., 2010; Argon et al., 1999). Spontaneous attempts at folding by mitochondrial proteins in the cytosol have been proven to occur, thereby preventing import (Strobel et al., 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 shown to be significantly lower than for that of their mature, signal-sequence cleaved counterparts (Laminet and Plückthun, 1989; Neupert, 1997). Thus, a translocation failure of mitochondrial and secretory proteins should generate conditions optimal for recognition and degradation by Ubr1p: availability of an uncleaved N-terminal sequence encoding an N-degron, accompanied by impaired 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 (Fig 5A). We substituted the P2 Val residue with a Ubr1-QC-incompatible Ser residue, and also mutated the P3 Leu to a Pro residue, producing ATP2Δ1,2,3-V2S,L3P. A P3 proline prevents Nt-acetylation of the P2-residue after the Met residue is cleaved, ensuring that any observed effect is not due to N-terminal blockage. The degradation of ATP2Δ1,2,3-V2S,L3P was retarded when compared to that of ATP2Δ1,2,3 in Δsan1+UBR1 cells, while both forms of the substrate were stabilized to similar levels in Δsan1Δubr1 cells, indicating that the decreased
turnover rate of ATP2Δ1,2,3-V2S,L3P was the result of a reduction in Ubr1-mediated degradation due to the N-terminal mutations (Fig 5B). We observed that a P3 proline in another misfolded substrate did not prevent Ubr1 recognition, indicating it is unlikely to have contributed to inhibition of Ubr1-mediated degradation for ATP2Δ1,2,3-V2S,L3P (Fig S3B). Degradation of ATP2Δ1,2,3 was dependent on Sse1, Ydj1, and Ssa1 and Ssa2 chaperones (Fig 5C), a hallmark of UPS-dependent CytoQC substrates, and a strong indication that ATP2Δ1,2,3 is misfolded (Heck et al., 2010; Prasad et al., 2010). Degradation of wild-type Atp2p in a temperature sensitive mitochondrial-import mutant, tom40-2 (Krimmer et al., 2001), was also dependent on Ubr1 CytoQC (Fig 5D), suggesting that wild-type ATP2 is misfolded when trapped in the cytosol. These results confirm that translocation-deficient Atp2p is an endogenous substrate of Ubr1-mediated CytoQC, and that its efficient degradation is dependent on its native Ubr1-QC-compatible P2 Val residue.

To assess the significance of structural aberrancy to Ubr1-recognition of a substrate via a Ubr1-QC-compatible P2-residue, we examined wild-type GFP, which is expected to escape CytoQC-based degradation as it is a stable protein in yeast. The P2 Ser of GFP was substituted with Ile to generate GFP-S2I. GFP-S2I was significantly less susceptible to Ubr1 compared to ∆2GFP-S2I, which was degraded efficiently by Ubr1 (Fig 6A). Trypsin digestion confirmed the relative structural stability of GFP-S2I vs ∆2GFP-S2I (Fig 6B). These results indicate that Ubr1-mediated degradation of a substrate via recognition of a compatible P2-residue is dependent on the substrate also being structurally aberrant. 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 akin to the model of protein quality control proposed for misfolded proteins possessing Ac/N-end N-degrons (Kim et al., 2013; 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).

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 (Fig S2). CPY encodes a Ubr1-QC-compatible P2 Lys. Degradation of pre-CPY-12iE was significantly reduced by the absence of Ubr1p (Fig 5E, Δsan1+UBR1 vs Δsan1Δubr1), confirming that a subset of mis-translocated secretory proteins are also targets of Ubr1.

**P2-residues encode cellular location signals that mediate CytoQC of mistranslocated proteins**

Studied have shown that mislocalized membrane proteins in the cytosol are targeted for degradation through exposed hydrophobic patches in both yeast and mammalian cells (Suzuki and Kawahara, 2016; Rodrigo-Brenni et al., 2014; Ast et al., 2014; Kawahara et al., 2013; Hessa et al., 2011). If mistranslocated proteins without Ubr1-QC-compatible P2-residues are primarily membrane proteins, the majority of them would be degradable through these pathways. To assess whether Ubr1 N-end rule based CytoQC, together with quality control pathways targeting hydrophobic domains, could together provide an effective degradation system for clearing the cytosol of translocation-deficient, mislocalized proteins, we analyzed a set of 277 ORFs encoding signal sequence-containing proteins that was used previously to determine the frequency of amino acids encoded at the P2 position of 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 based on existing literature. If existing literature supporting soluble or membranous topology for a particular ORF did not exist, data from SignalP, Kyte-Doolittle hydrophilicity profiling, and Phobius, were together utilized for the determination of topology (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, $\chi^2 = 11.7$) (Fig 7A; Tables S12-S18). 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 x 10$^{-5}$, $\chi^2 = 16.8$). These results indicate that the vast majority of ER proteins which are not compatible with Ubr1 are...
also, at a much higher relative frequency, membrane proteins, which is expected to facilitate 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. (Fig 7B; Table S12-S18).

Thus, while the majority of signal-sequence bearing proteins are susceptible to Ubr1, soluble secretory proteins in particular are primed 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 (Forte et al., 2011) (Fig 7B; Table S11). This minority of cytosolic proteins with Ubr1-QC-compatible P2-residues should fold efficiently, as they are in their native folding environment, thereby escaping targeting by Ubr1 CytoQC, as demonstrated to be the case in stability experiments comparing folded vs misfolded GFP (Fig 6).

It should be noted that while some native cytosolic proteins have been shown to be targeted via Met-Φ N-degrons, the turnover rates of these proteins were only significantly affected by Ubr1 deletion when NatC Nt-acetyltransferase, which acylates Met-Φ proteins, was concomitantly deleted, thereby placing these substrates in an inherently aberrant, unacetylated state (Kim et al., 2014). Therefore, Ubr1 appears to also possess a role in quality checking the minority of cytosolic proteins that possess P2-based N-degrons. Indeed, a role for Ubr1 in the QC of misfolded cytosolic proteins has been demonstrated previously. An analysis of these substrates, and their alignment with our characterization of Ubr1/P2-residue compatibility, is provided in Discussion.

The biased distribution of Ubr1-QC-compatible vs Ubr1-QC-incompatible P2-residues we detected across the proteome (Fig 4), reflected again in the results of targeted analyses of representative sets of secretory and cytosolic proteins (Fig 7B), indicates that P2-residue identity is inherently linked with cellular localization. That the P2-residues predominantly featured in
secretory proteins and mitochondrial proteins are Ubr1-QC-compatible, and those predominantly featured in cytosolic and nuclear proteins are Ubr1-QC-incompatible, strongly suggest that P2-residues also function as localization-based degradation signals that help the cell screen for and degrade proteins that are aberrantly localized due to mistranslocation. We therefore expect that the minority of secretory proteins that do not possess Ubr1-QC-compatible P2-residues (~21.2%; Fig 7B) to be primarily membrane proteins, allowing them to be effectively degraded by pathways recognizing membrane domains. Indeed, we found that almost all of the secretory proteins (~87.5%) in the representative secretory protein dataset that possess Ubr1-QC-incompatible P2-residues, are also membrane proteins (Fig 7A). Importantly, we also expect a strong pressure for soluble secretory proteins to possess Ubr1-QC-compatible P2-residues, since they are not as amenable to membrane domain-dependent degradation pathways. This turns out to be true, based on our finding that ~93% of soluble secretory proteins encode Ubr1-QC-compatible P2-residues (Fig 7B). These latter two results demonstrate further that across the proteome, P2-residues function as cellular locations signals which facilitate targeted, non-promiscuous quality control in the cytosol by the Arg/N-end rule pathway and Ubr1.

Discussion

The findings presented here broaden our understanding of how eukaryotic cells ensure 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. The data suggests 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 when mislocalized to the cytosol due to translocation failure. Soluble ER proteins are particularly targetable by Ubr1 since the vast majority possess Ubr1-QC-compatible P2-residues. That they would not be readily 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. Collectively, our findings suggest that a majority of proteins in the eukaryotic proteome possess P2-residues that are utilized as de facto cellular location signals which help to ensure the fidelity of protein localization. We term these P2CLS-residues (P2-encoded Cellular Location Signal residues). This system of P2-encoded
location signaling facilitates efficient N-end rule/Ubr1-mediated degradation of a majority of mis-translocated secretory and mitochondrial proteins in the cytosol, as they encode Ubr1-QC-
compatible P2CLS-residues. At the same time, Ubr1-QC-incompatible P2CLS-residues are encoded by the majority of native cytosolic proteins, thereby protecting them from promiscuous degradation by Ubr1. Failures in SRP-mediated ER targeting cause the mistargeting of secretory proteins to the mitochondria, resulting in mitochondrial dysfunction (Costa et al., 2018), while the 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 may also be mitigated by the P2CLS-dependent Ubr1 CytoQC pathway, capturing such proteins in the cytosol before mis-targeting occurs.

While we demonstrated here that mistranslocated secretory and mitochondrial proteins trapped in the cytosol have the highest propensity for targeting by the Arg/N-end rule, roughly 30% of the cytosolic proteome also encode Ubr1-QC-compatible P2CLS-residues (Figs 4E,7B). This subset of cytosolic proteins should also be subject to Arg/N-end rule, N-degron-dependent degradation when they are misfolded or structurally aberrant in other ways. Indeed, several misfolded cytosolic Ubr1 substrates have been identified. Not every misfolded Ubr1 substrate is expected to possess a Ubr1-QC-compatible P2CLS-residue, as we still observed some capacity for Ubr1-mediated degradation even in the case of substrates with Ubr1-QC-incompatible P2CLS-residues, such as ∆2GFP. However, degradation by Ubr1 is significantly enhanced when a Ubr1-QC-compatible P2CLS-residue is present (Fig 1). We would expect to see a similar pattern of effect with natively localized cytosolic substrates. A brief analysis of previously reported cytosolic misfolded substrates suggests a similar pattern of dependency on P2CLS/Ubr1 compatibility for efficient degradation by Ubr1:

1) N-terminally HA-tagged Tpk2 (HA-Tpk2) was strongly stabilized by the deletion of Ubr1 when folding was inhibited by treatment with geldanamycin, with ~37% of substrate remaining in wild-type cells at the 40 minute time-point, versus more than 65% of substrate remaining in ∆ubr1, a ~28 percentage point difference due to deletion of Ubr1 (Nillegoda et al., 2010). The HA tag possesses a P2 Tyr (Y) residue, which we detected as a Ubr1-QC-compatible P2. On the other hand, another substrate they analyze that does not possess Ubr1-QC-compatible P2,
Ste11ΔN^K44R, only exhibited a ~5 percentage point difference in remaining substrate by the 60-minute time-point due to deletion of Ubr1 relative to wild-type (~20% vs ~15%, respectively).

2) Summers et al., analyzed the turnover of slGFP (short-lived GFP) and observed a substantial difference in slGFP remaining after 90 minutes when comparing wild-type and Δubr1 strains (<5% remaining versus ~50% remaining, respectively) (Summers et al., 2013). slGFP encodes a P2 Gly (G), a Ubr1-QC-compatible residue, so its degradation is, as expected, strongly dependent on Ubr1. Importantly, P2 Gly is one of the novel N-degrons which we discovered through the mutational analysis of Ste6*C, helping to demonstrate its applicability to other folding impaired substrates.

3) Wolf and colleagues examined the degradation of orphaned Fas2, a subunit of fatty acid synthase. Fas2 encodes a Ubr1-QC-compatible P2 Lys. A large difference in substrate levels remaining was detected at 2, 4 and 6 hours when comparing Δfas1 to Δfas1Δubr1 strains (<10% remaining at 6 hours, versus >50% remaining, respectively), demonstrating a strong dependency on Ubr1 for degradation (Scazzari et al., 2015).

4) Khosrow-Khavar et al. discovered cytosolic temperature sensitive substrates that exhibited some dependency on Ubr1 for degradation, namely, ugpl-3 and glnl-3. Both do not possess Ubr1-QC-compatible P2-residues. While dependency on Ubr1 for degradation was detected for both substrates, the percentage of substrate remaining was still greater than 60% for both substrates even by the 3-hr time-point (cycloheximide chase), suggestive of inefficient targeting by Ubr1, as expected based on our model of Ubr1 P2-residue compatibility (Mayor et al., 2012).

5) Comyn et al identified a temperature sensitive allele of yeast Guk1 guanylate kinase, Guk1-7-GFP, that is degraded by Ubr1, however, only a 10% difference in substrate remaining was observed after 2 hours in wild-type cells vs Δubr1 (Comyn et al., 2016). Guk1 encodes a Ubr1-QC-incompatible P2 Ser.

6) Heck et al., discovered several N-terminally HA-tagged misfolded substrates that were strongly stabilized by deletion of Ubr1 (Heck et al., 2010). HA-tags encode a Ubr1-QC-
compatible P2 Tyr residue, and thus, a strong dependency on Ubr1 for degradation is expected. One of the substrates analyzed in that study is triple HA-tagged stGnd1 (3HA-stGnd1). Data from our own experiments showed that moving the HA tag from the N-terminus of stGnd1 to the C-terminus (stGnd1-HA) significantly inhibits Ubr1-mediated degradation, with virtually none detected at the 30 minute time-point, and a minor amount at 60 minutes (Fig S3A). The P2-residue of stGnd1-HA is a Ubr1-QC-incompatible P2 Ser residue. Upon mutation of the P2-residue to Ubr1-QC-compatible P2-residues (in addition to mutation of P3 to Pro) the substrate was rapidly degraded by Ubr1 (Fig S3B,C). stGnd1-HA-S2IA3K was assessed to demonstrate that the P3 proline is not the causal factor in the activation of Ubr1-mediated degradation (Fig S3B).

Together, the data in existing literature reporting on misfolded proteins that are degraded by Ubr1, along with results from our own experiments, strongly supports the conclusion that efficient degradation of misfolded proteins by Ubr1 is highly dependent on P2-residue compatibility.

The discovery of Met-Φ N-degrons (Kim et al., 2014; Tran, 2013) was surprising for several reasons. Firstly, N-terminal Met was originally classified as a stabilizing residue. Secondly, the existence of Met-Φ sequences as N-degrons meant that a large number of proteins translated in the cytosol are susceptible to degradation by Ubr1. Our characterization of Met-Φ N-degrons as part of an expanded set of Arg/N-end rule N-degrons that are recognized when possessed by folding-impaired, mistranslocated proteins, helps to identify a specific common physiological occurrence under which these N-degrons take effect: translocation failure. Our findings thus bring to light large swaths of secretory and mitochondrial proteins that are degradable via the Arg/N-end rule. Importantly, they do not require an endoproteolytic cleavage or N-terminal modification to expose destabilizing residues, as the case with the majority of previously characterized substrates in Arg/N-end rule-dependent mechanisms. Instead, they rely on the qualities associated with misfolding and mislocalization as a result of mistranslocation to enable the utilization of their natural N-terminal N-degrons as effective degradation signals.
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 P2CLS-residue-dependent Ubr1 substrates are engaged by the cytosolic form of Ubr1, or are first trafficked to the nucleus and subsequently recognized by the nuclear form of Ubr1, remains to be determined. However, since we observed an even lower usage frequency of Ubr1-QC-compatible P2CLS-residues in the case of nuclear-localized proteins as compared to cytosolically-localized proteins, Ubr1 presence and degradation activity in both nuclear and cytosolic compartments would still be congruent with our model of P2CLS-based location signaling and degradation (Fig 4E).

There is mounting evidence that herpes simplex virus type 1 (HSV) plays a role in the development of Alzheimer’s disease (Itzhaki, 2018). 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 (Zheng et al., 2018; Brower et al., 2013). An intriguing avenue of research would be to investigate whether the P2CLS-residue dependent degradation of translocation-deficient proteins by Ubr1 described here is mirrored in mammalian cells. Since the Arg/N-end rule pathway is highly conserved from yeast through mammals, this is likely to be the case. Determining whether disruption of the pathway is either directly or indirectly involved in the accumulation of such detrimental disease factors would be of high priority for future studies.

An additional surprising corollary to our discovery is that the N-end rule, 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 et al., 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 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. It is thus no surprise that the N-terminal signal sequences of ER and mitochondrial proteins largely coincide with the set of P2CLS-residues that are Ubr1-compatible, with soluble translocated proteins in particular exhibiting nearly total concordance based on our analyses of unbiased datasets. In this way, the N-end rule pathway may have been the key enabler for eukaryotic cells to fully exploit the benefits of harboring endosymbiotic organelles by compensating for the negative physiological effects associated with mistargeted proteins.

Materials and Methods

Study Design

Research objectives and units of investigation
The over-arching goal of the project was to systematically determine the effects that altering N-terminal sequences had on the degradation of misfolded proteins by Ubr1 and the N-end rule pathway. Specifically, this was done by selecting a known cytosolic quality control substrate, Ste6*C, creating a full set of P2-residue variants derived from it, and assessing degradation rates of the substrates in yeast cell lines that differed in the presence or absence of Ubr1 and San1, the E3 ligases of interest. The P2 position was selected for analysis as it is the primary determinant in the N-terminal processing of proteins and the resultant N-terminal sequence of a protein prior to downstream processing of signal sequences (Giglione et al., 2004; Sherman et al., 1985). Based on these initial degradation assays, global bioinformatic analysis was used to determine the cellular compartments that had the highest frequency of proteins with Ubr1-QC compatible P2-residues encoded. The finding that mitochondrial and secretory proteins encode Ubr1-QC-
compatible P2-residues more frequently led us to formulate the follow-up hypothesis that they are the primary targets of Ubr1 quality control in the cytosol in the event of translocation failure. This hypothesis then drove us to perform the targeted in-depth bioinformatic and biochemical analysis of representative secretory and mitochondrial protein sets and substrates in relevant cell-lines and translocation conditions to prove the validity of the hypothesis, finally leading to formulation of the new model for Ubr1-mediated CytoQC.

Replicate and selection of end-points
Pulse-chase experiments, used to determine the rates of degradation of endogenous and engineered substrates in various cell backgrounds, were performed in sets of at least 3 biological replicates unless otherwise noted. Three replicates is an accepted standard minimum for many biological experiments, and is also widely utilized in previously published reports which employed the same protocol. Furthermore, it allowed us to determine means, standard deviations, and verify statistically significant differences (or lack of differences) between strains and/or substrates. Intermediate end-points for pulse-chase assays were selected either based on those used in previous studies analyzing the same substrate type, or through pilot experiments to determine durations that would allow discernable, statistically significant deviations to surface in relevant strains. Replicates were generally performed on separate days and times to account for possible fluctuations in laboratory environment and equipment performance. Strains expressing analyte substrates were freshly inoculated for each replicate experiment to account for normal or unexpected variations in growth environment, growth medium, and biological activity.

Experimental Design
Biological experiments were conducted in controlled laboratory environments with specific incubation temperatures, times, treatment quantities, durations, and measurement methods, as described below in the sections applicable for each experiment.

Sample sizes
For bioinformatic analyses assessing the relative frequencies of P2-residue amino acid usage in the yeast proteome based on cellular localization and ORF sequences, a widely-cited GFP-fusion localization study (Huh et al., 2003) was utilized in order to determine and compare P2 amino
acid residue encoding frequency for proteins in different cellular compartments. For the best assessments of this parameter, the largest sample size available from the study for each cellular compartment analyzed was utilized, and is described in more detail in the Bioinformatic Analysis section.

For targeted bioinformatic analyses of P2 amino acid usage and topological determination of secretory pathway proteins based on a combination of literature review and Kyte Doolittle/Phobius methods, we utilized a published set of 277 signal-sequence bearing proteins/ORFS that was previously used for determining P2-residue frequency for proteins in the secretory pathway (Forte et al., 2011). Using a previously vetted and published secretory protein set eliminated chance of bias for the purpose of our analysis of P2 Ubr1-QC-compatibility. Based on the same logic, the set of 251 randomly selected cytosolic proteins from the same study (Forte et al., 2011) was used for our analysis of P2 amino acid residue frequencies in the cytosol to prevent potential bias in selecting cytosolic proteins for our analysis.

**Statistical Analysis**

A standard of 3 replicates for biological experiments was performed in the determination of means, standard deviations, and statistical significance, unless otherwise noted. Unpaired, two-tailed Student’s t-test was performed to assess statistical significances in differences in degradation rates based on pulse-chase experiments.

Chi-square analysis was performed to assess statistically significant differences in P2-residue amino acid frequency distributions between protein sets across all 20 possible residues using 19 degrees of freedom. Pair-wise chi-square analysis was performed when comparing the relative frequency of a single residue between two protein sets using 1 degree of freedom. Pair-wise chi-square analysis was performed when comparing the relative frequency of Ubr1-QC-compatible proteins between two protein sets, and when comparing the relative frequency of membrane proteins between two protein sets, using 1 degree of freedom.
Detailed N values (either biological experimental replicates, or samples) are included in the respective figure legends.

P-value thresholds were set to 0.05 or less, and are indicated in the respective figure legends as asterisks between the relevant sample data, or multiple asterisks in the case of lower p-value thresholds.

Strains and Antibodies

Yeast strains used in this study are listed in Table S30 (available in Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7). Anti-HA monoclonal antibody (HA.11) was sourced from Covance (Princeton, New Jersey). Strains were tested for contamination before usage in experiments. Monoclonal anti-3-phosphoglycerate kinase (PGK) was sourced from Invitrogen (Carlsbad, California). Anti-CPY antibody was a gift from Reid Gilmore (University of Massachusetts, Worcester, MA).

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) or ImageJ. Background signal from screen
exposure was subtracted. All results presented are the mean ± SD of three independent experiments (N=3, biological replicates) unless otherwise specified.

**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 Sequencing**

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 dH2O 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 Database (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, and divided it by the aggregate abundance of all proteins within the full localization category (Table S19-S29). Dataset tables available in Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7.

**Categorization of signal-sequence bearing proteins**

A published set of 277 signal-sequence bearing proteins was analyzed (Forte et al., 2011) (Table S12). 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-S17; Dataset tables available in Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7). 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 (Peterson et al., 2011). 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 (Käll et al., 2004; Kyte and Doolittle, 1982). 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.

**Plasmids and Plasmid Construction**

Standard 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. stGnd1-HA and its derivatives were expressed under control of the endogenous Gnd1 promoter in yeast centromeric plasmids.

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.

pAT11: A fragment containing the GND1 promoter and encoding residues 1-149 of the GND1 ORF was PCR amplified from yeast genomic DNA using primers AT60 and AT61. The fragment was digested with NcoI and NotI and ligated to pAT1 digested with NcoI and NotI, generating pAT11.

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.
pAT79: pAT79 (expressing stGnd1-HA-S2I,A3P) was constructed by mutation of the base-pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primers AT158 and pAT78 as a template.
pAT80: pAT79 (expressing stGnd1-HA-S2I,A3K) was constructed by mutation of the base-pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primers AT159 and pAT78 as a template.
pAT82: pAT82 (expressing stGnd1-HA-S2L,A3P) was constructed by mutation of the base-pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primers AT160 and pAT78 as a template.
pAT85: pAT85 (expressing Gnd1-HA-S2I,A3P) was constructed by mutation of the base-pairs encoding the 2nd and 3rd residue of Gnd1 through site-directed mutagenesis using primers AT158 and pAT84 as a template.
pAT87: pAT87 (expressing stGnd1-HA-S2Y,A3P) was constructed by mutation of the base-pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primers AT161 and pAT78 as a template.
Acknowledgments:

I would like to thank my ex-graduate advisor, Davis Ng, for his expert guidance and input. I am grateful to Nassira Bedford for her assistance with unpublished experiments, and Chengchao Xu and Eric Fredrickson for their critical review of the manuscript. I am grateful to the Nikolaus Pfanner lab for sharing yeast strains. I thank Alexander Varshavsky for his insightful feedback on my thesis data.

Funding:

This work was supported by funds from the National University of Singapore Research Scholarship awarded to Anthony Tran by the National University of Singapore (Faculty of Science).

Author contributions:

AT conceived of the overall project, designed the studies, performed the experiments and analyses, and wrote the manuscript.

Competing Interests:

The author declares that there are no competing interests.

Data availability:

All data is available in the main text, supplementary materials, or supplementary tables uploaded to the Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7. Other raw data, code, gel scans, etc., are available upon request from the author.
References:


Fig. 1. Efficient degradation of Ste6*C and Δ2GFP by the Ubr1 quality control pathway is dependent on P2-residue identity. (A) and (B) Turnover rates of Ste6*C and Δ2GFP in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells were assessed by metabolic pulse-chase analysis. (C) and (D) Turnover rates of Ste6*C-I2S and Δ2GFP-S2I in +SAN1+UBR1(wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells were assessed
by metabolic pulse-chase. (E) Turnover rates of P2 Ser substrates (Ste6*C-I2S, Δ2GFP-S2S) with P2 Ile substrates (Ste6*C-I2I, Δ2GFP-S2I) in wild-type cells were assessed by metabolic pulse-chase. (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 three independent experiments (N=3, biological replicates). Student’s t-test: *, p < 0.05; **, p < 0.01; ***, p < 0.005; Not Significant (NS), p > 0.5.
Fig. 2. Efficient degradation of Ste6*C by the Ubr1 quality control pathway, but not the San1 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]) were compared to the turnover rate of the original Ste6*C substrate (Ste6*C-I2I) in ∆san1+UBR1 cells. Turnover rates were assessed by metabolic pulse-chase as described in Fig 1. (D), (E) and (F) Same as described in Fig 2A,B,C, but in +SAN1∆ubr1 cells. Error bars, mean +/- SD of three independent experiments (N=3, biological replicates) unless otherwise noted. Due to an N=2 for Ste6*C-I2K expressed in +SAN1∆ubr1 cells because of contamination (Fig 2D), lack of change versus Ste6*C-I2I could not be statistically confirmed; however, in both biological replicates of the experiment, the Ste6*C-I2K substrate was degraded to a further extent than Ste6*C-I2I, indicating that the I2K mutation does not inhibit degradation. (G) Heat-map reflecting results presented in Fig 2A-F.
Fig. 3. Rapidly degraded Ste6*C P2 variants are degraded in a Ubr1-dependent manner. (A) Turnover rates of N-terminally destabilized Ste6*C-I2X mutants in Δsan1Δubr1 cells were assessed by metabolic pulse-chase as described in Fig 1. Degradation of Ste6*C-I2I in Δsan1+UBR1 cells was also examined as a control. Error bars, mean +/- SD of three independent experiments (N=3, biological replicates). (B) Heat map reflecting turnover rates of Ste6*C-I2X substrates in Δsan1Δubr1 as presented in Fig 3A.
Fig 4. Proteins localized to secretory pathway components and mitochondria exhibit a bias for encoding UBR1-QC-compatible P2-residues when compared to cytosolically-localized proteins. (A) Relative frequency of P2 amino acid usage of proteins localized to the secretory pathway components (N=559, proteins) vs cytosol (N=823, proteins). Frequency distribution differed significantly between the two sets (p < 1 x 10^{-6}, \chi^2 = 64.33, df = 19; Table S9). Pair-wise Chi-square analysis: *, p < 0.01 for A, N, K, R; p < 0.05 for P, T, W, I. (B) Relative frequency of P2 amino acid usage in proteins localized to mitochondria (N=458, proteins) vs cytosol (N=823, proteins). Frequency distribution differed significantly between the two sets (p < 1 x 10^{-5}, \chi^2 = 317.85, df = 19; Table S9). 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. (C) Relative frequency of P2 amino acid usage in proteins localized to either mitochondria or secretory pathway (N=1017, proteins) vs cytosol (N=823, proteins). Frequency distribution differed significantly between the two sets (p < 1 x 10^{-3}, \chi^2 = 196.34, df = 19; Table S10). 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. (D) Relative frequency of P2 amino acid usage in proteins localized to nucleus (N=639, proteins) vs cytosol (N=823, proteins). Frequency distribution differed less significantly between the two sets (p > 0.005, \chi^2 = 36.33, df = 19; Table
S10). Pair-wise Chi-square analysis: *, p < 0.001 for S; p < 0.01 for C, D, L. (E) Estimated total relative abundance of proteins that are encoded with Ubr1-QC-compatible P2-residues by localization category (Tables S19-S29). Data available in Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7.
Fig. 5. Mistranslocated forms of mitochondrial and secretory proteins are substrates of cytosolic quality control mediated by Ubr1 and the N-end rule pathway. (A) Turnover rate of ATP2Δ1,2,3 in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells was assessed by metabolic pulse-chase (N=5, biological replicates) (B) Turnover rates of ATP2Δ1,2,3 and ATP2Δ1,2,3-V2S,L3P in Δsan1+UBR1 and Δsan1Δubr1 cells were assessed by metabolic pulse-chase (N=4, biological replicates). (C) Turnover rate of ATP2Δ1,2,3 in WT, Δydj1, Δsse1Δsse2, and Δssa1 cells was assessed by metabolic pulse-chase. (D) Turnover rate of
HA-tagged wild-type ATP2 expressed in TOM40 (wild-type), *tom40-2*, and *tom40-2Δubr1* cells was assessed by metabolic pulse-chase. (E) Turnover rate of CPY-12iE-HA in *Δsan1+UBR1* and *Δsan1Δubr1* cells was assessed by metabolic pulse-chase. (A-E) Metabolic pulse-chase was performed as described in Fig 1 with a 10-minute pulse-labeling step. Experiment involving temperature sensitive *tom40-2* strains was performed at 37°C for all 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. Error bars, mean +/- SD of three independent experiments (N=3, biological replicates) unless otherwise noted. Student’s t-test: *, p < 0.05; **, p < 0.01; ***, p < 0.005; Not Significant (NS), p > 0.5.
Fig. 6. The efficient P2-residue-based Ubr1-mediated degradation of a substrate is dependent on concomitant presence of structural aberrancy. (A) Turnover rates of GFP-S2I and Δ2GFP-S2I in Δsan1Δubr and Δsan1+UBR1 cells were assessed by metabolic pulse-chase as described in Fig 1 (N=3, biological replicates). Student’s t-test: *, p < 0.025; Not Significant (NS), p > 0.2. (B) Trypsin sensitivity assay (N=1): 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.
Fig. 7. Secretory proteins encoding Ubr1-QC-\textit{incompatible} P2-residues are primarily membrane proteins, while soluble secretory proteins primarily encode Ubr1-QC-\textit{compatible} P2-residues. (A) 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-\textit{incompatible} P2-residues, and the full set (see Materials and Methods; Tables S12-S18). Chi-square test: Ubr1-incompatible (N=56, proteins) vs full set (N=270, proteins): **, p < 0.001, $\chi^2 = 11.7$, $df = 1$; Ubr1-incompatible (N=56, proteins) vs Ubr1-compatible (N=214, proteins): ***, p < 5 x 10^{-5}, $\chi^2 = 16.8$, $df = 1$. (B) 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, S14-S17). Chi-square test: Soluble (N=97, proteins) vs full set (N=273, proteins): *, p < 0.005, $\chi^2 = 9.7$, $df = 1$. Soluble (N=97, proteins) vs membrane set (N=173, proteins): ***, p < 5 x 10^{-5}, $\chi^2 = 16.8$, $df = 1$. Cytosolic protein set (N = 251, proteins) from Forte et al., 2011. Data available in Dryad Digital Repository: https://doi.org/10.5061/dryad.r837vv7.
Fig. S1. N-terminal sequencing of Ste6*C constructs. (A) Lysate was prepared from ∆san1Δubr1 cells expressing Ste6*C-HA-I2S. 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 ∆san1Δubr1 cells expressing Ste6*C-HA. Protein was immunoprecipitated from lysate and sequenced as described in Fig S1A. (C) Lysate was prepared from ∆san1Δubr1 cells expressing Ste6*C-HA-I2V. Protein was immunoprecipitated from lysate and sequenced as described in Fig S1A.
Fig S2. C-terminal HA-tag is removed from HA-tagged CPY in the vacuole. Turnover rate of CPY-HA was assessed in ∆prc1, ∆pep4, ∆san1+UBR1, and ∆san1Δubr1 cells by metabolic pulse-chase as described in Fig 1. CPY was probed with anti-CPY and anti-HA antibodies in ∆prc1 cells, and anti-HA in ∆pep4 cells.
Fig S3. Ubr1-mediated degradation of stGnd1 is P2-residue/N-degron dependent. (A) Turnover rate of C-terminally HA-tagged stGnd1 (stGnd1-HA) expressed in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells was assessed by metabolic pulse-chase. Error bars, mean +/- SD of two independent experiments (N=2, biological replicates). (B) Turnover rates of C-terminally HA-tagged stGnd1-HA-S2IA3K and stGnd1-HA-S2IA3P expressed in wild-type and Δubr1 cells were assessed by metabolic pulse-chase. Error bars, mean +/- SD of two or three independent experiments (N=2 for stGnd1-HA-S2IA3K; N=3 for stGnd1-HA-S2IA3P). (C) Turnover rates of C-terminally HA-tagged stGnd1-HA-S2YA3P and stGnd1-HA-S2LA3P expressed in Δsan1+UBR1 and Δsan1Δubr1 cells were assessed by metabolic pulse-chase (N=1, biological replicates). (A-C) Metabolic pulse-chase was performed as described in Fig 1.