

RESEARCH ARTICLE

The N-end rule pathway and Ubr1 enforce protein compartmentalization via P2-encoded cellular location signals

Anthony Tran*

ABSTRACT

The Arg/N-end rule pathway and Ubr1, a ubiquitin 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 *Saccharomyces cerevisiae* proteome, it was 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 second 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 diseases. 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 subcellular 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.

This article has an associated First Person interview with the first author of the paper.

KEY WORDS: N-end rule pathway, Ubr1, Ubiquitin-mediated degradation, Proteolysis, Protein quality control, Protein misfolding, CytoQC, N-degron

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 (Pearce and Kopito, 2018; Currais et al., 2017; Dubnikov et al., 2017). Cytosolic quality control (CytoQC) pathways, a subset of PQC, specifically mediates the clearance of cytosolically localized aberrant proteins. In *Saccharomyces cerevisiae*, several ubiquitin E3 ligase enzymes 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 and Gardner, 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, 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 branches. Ubr1p mediates the Arg/N-end branch, while the E3 ligases Doa10p and Not4p, 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 Met 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 $\Delta\text{ssC}^{22-519}\text{Leu2}_{\text{myc}}$, 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 second 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 Met 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, in the absence of other N-terminal modifications such as acetylation, which blocks N-terminal recognition by Ubr1, the P2 residue of a protein is the primary

National University of Singapore, Department of Biological Sciences, Singapore 117604.

*Author for correspondence (anthonytran17@gmail.com)

 A.T., 0000-0003-3487-9255

Received 11 March 2019; Accepted 22 March 2019

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 pathway after translation in the cytosol. The enzyme responsible for N-terminal Met excision is methionine aminopeptidase (MAP), which cleaves initiating Met residues 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. Although 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 a P2 Cys, which only acts as a destabilizing residue under conditions where it is oxidized (occurs in plants and animals, but not in yeast). 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, here, the effect that P2 residue identity has on the degradation of misfolded proteins by Ubr1 is examined. The array of P2 residues that produce an expanded set of Ubr1 compatible N-degrons specific to CytoQC is characterized 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 Pro residue, which prevents Nt-acetylation for Met-cleaved proteins, and partially inhibits Nt-acetylation for Met-retaining proteins (Goetze et al., 2009; Polevoda and Sherman, 2003). This ensures that Nt-acetylation is not a factor in any inhibition of Ubr1-mediated degradation detected for Ste6**C* P2 mutants that are cleaved of Met. 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 uncovered differs from the original set of destabilizing Arg/N-end rule residues (Varshavsky, 2011), which only included single-position primary destabilizing residues exposed at the N-terminus (basic residues, His, Lys and Arg; bulky hydrophobic residues, Leu, Phe, Trp, Tyr and Ile), or secondary (Asp, Glu, and oxidized Cys) and tertiary (Asn, Gln, and Cys) 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 defined here partially overlaps with the recently discovered Met-Φ N-degrons, which entail a leading Met 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 Met. The approach utilizing Ste6**C* allowed the identification, *in vivo*, of 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, detailed global bioinformatic analyses of P2 amino acid usage frequencies in the *S. cerevisiae* proteome were performed, considering the expanded P2-dependent N-degron set, and it was discovered that the specificity of Ubr1 is designed to preferentially degrade secretory and mitochondrial proteins that fail to translocate and become

mislocalized to the cytosol. A precise *in vivo* analysis of turnover rates of endogenous translocation-inhibited secretory and mitochondrial proteins, in the presence and absence of Ubr1, confirms this activity. Finally, an 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. Taken together, these findings demonstrate that the P2-residue-dependent encoding of cellular location, and the N-end rule pathway, serve essential cognate roles in proofreading 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, 2010). To begin investigating the effect of P2 residue identity on CytoQC degradation, it was 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 Met 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 Met 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, it was surmised that by altering the N-terminus of Ste6**C* and Δ2GFP, it would be possible to modulate the amount of degradation that proceeds via Ubr1. P2 mutants of Ste6**C* and Δ2GFP that matched the P2 residue of the other substrate were generated to assess the impact on Ubr1-mediated degradation. The N-terminus of Ste6**C* was modified to mimic that of Δ2GFP by replacing 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 Δ*san1* or +SAN1 backgrounds (Δ*san1*Δ*ubr1* versus Δ*san1*+UBR1; +SAN1Δ*ubr1* versus +SAN1+UBR1) (Fig. 1C). Thus, by replacing 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 that of Ste6**C*-I2I (Fig. 1A,C; +SAN1Δ*ubr1*).

The inverse experiment was performed with Δ2GFP (Δ2GFP-S2S) by replacing 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 (Fig. 1B,D). As for the case

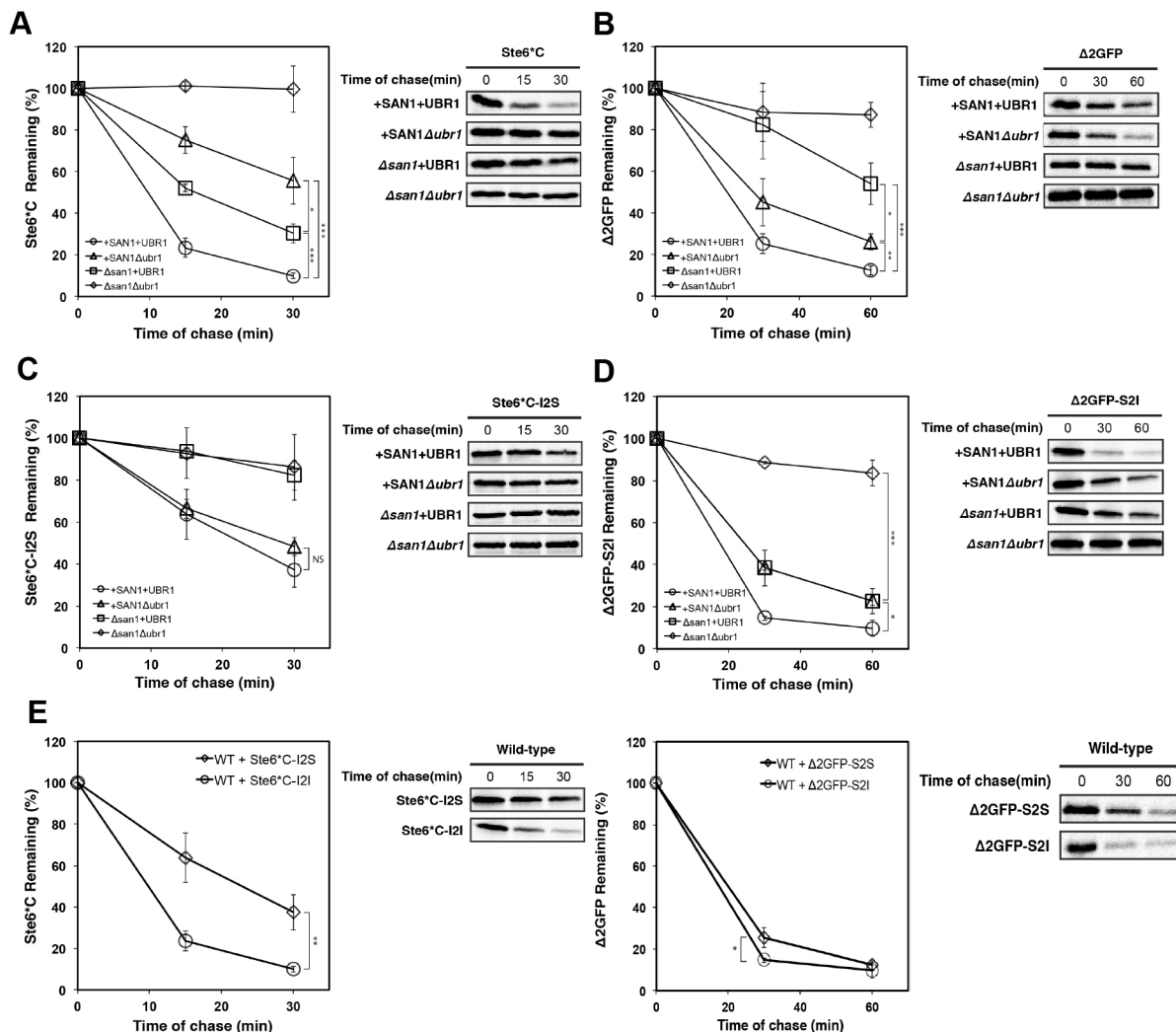


Fig. 1. Efficient degradation of Ste6*C and Δ2GFP by the Ubr1 QC pathway is dependent on P2 residue identity. (A,B) Turnover rates of Ste6*C and Δ2GFP in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells were assessed by metabolic pulse-chase analysis. (C,D) Turnover rates of Ste6*C-I2S and Δ2GFP-S2I in +SAN1+UBR1 (wild-type), Δsan1+UBR1, +SAN1Δubr1, and Δsan1Δubr1 cells were assessed by metabolic pulse-chase analysis. (E) Turnover rates of P2 Ser substrates (Ste6*C-I2S, Δ2GFP-S2S) and P2 Ile substrates (Ste6*C-I2I, Δ2GFP-S2I) in wild-type cells were assessed by metabolic pulse-chase analysis. All results (A–E) are for metabolic pulse-chase: cells were grown to log phase and shifted to 30°C for 30 min followed by pulse-labeling with [35S]Met/Cys (5 min for Ste6*C 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 phosphorimager analysis. Results are mean ± s.d. for three independent experiments ($n=3$, biological replicates). * $P<0.05$; ** $P<0.01$; *** $P<0.005$; NS, not significant, $P>0.5$ (Student's t -test).

with Ste6*C-I2S, degradation efficiency of Δ2GFP via the San1-mediated pathway remained relatively unaffected by the S2I mutation (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 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 Ile residue and enable recognition via an N-terminal Ile instead. To determine which of the two possibilities was occurring, N-terminal sequencing on Ste6*C-I2I isolated from Δsan1Δubr1 cells was performed. Sequencing confirmed that the leading Met 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.

It was found 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, referred to hereafter 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, referred to hereafter as Ubr1-QC-compatible P2 amino acid residues (Fig. 2A–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–G). The Ubr1-dependent degradation of the nine most-rapidly degraded Ste6*C P2 mutants was confirmed in Δsan1Δubr1 cells (Fig. 3A,B). The inhibition of Ubr1-mediated degradation observed for Ste6*C mutants possessing Ubr1-QC-incompatible P2 residues was not a result of Ubr1 blockage by Nt-acetylation since Ste6*C is encoded with a P3 proline, which prevents Nt-acetylation. Leading Met retention is not a prerequisite for compatibility with the Ubr1-QC degradation pathway, as

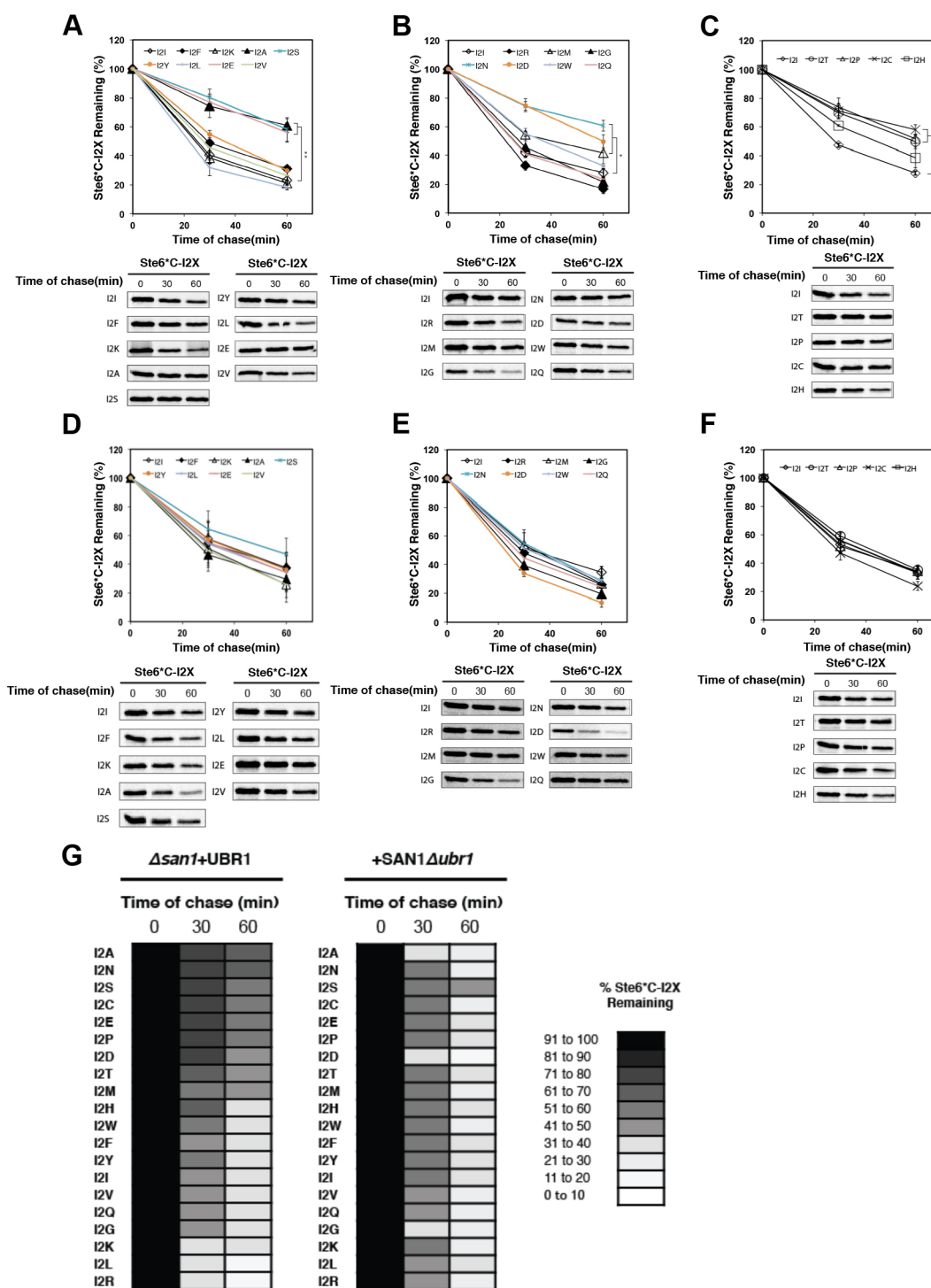


Fig. 2. Efficient degradation of Ste6*C by the Ubr1 QC pathway, but not the San1 pathway, is highly dependent on P2 residue identity. (A–C) Turnover rates of 19 Ste6*C P2 amino acid residue mutants [Ste6*C-I2X, where X=A (alanine), N (asparagine), S (serine), C (cysteine), E (glutamic acid), P (proline), 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) and R (arginine)] were compared to the turnover rate of the original Ste6*C substrate (Ste6*C-I2I) in $\Delta san1+UBR1$ cells. Turnover rates were assessed by metabolic pulse-chase analysis as described in Fig. 1. (D–F) As described in A–C, but in $+SAN1\Delta ubr1$ cells. Results are mean \pm s.d. for three independent experiments ($n=3$, biological replicates) unless otherwise noted. Owing to $n=2$ for Ste6*C-I2K expressed in $+SAN1\Delta ubr1$ cells because of contamination (D), the 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 A–F. * $P<0.05$; ** $P<0.01$ (Student's t -test).

Ste6*C-I2V was demonstrated to be degraded efficiently by Ubr1 and is predicted to undergo Met cleavage, as also confirmed via sequencing (Fig. S1C).

Interestingly, P2 His, Lys, Arg and Gln, all of which cause retention of initiating Met residues based on the rules of Met excision, are not predicted to produce N-termini with known N-degrons of the

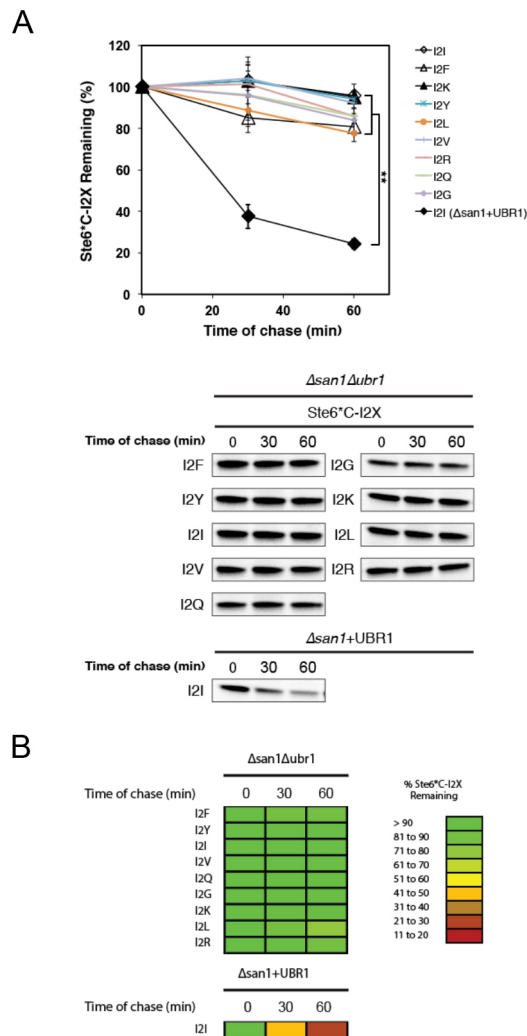


Fig. 3. Rapidly degraded Ste6*C P2 variants are degraded in a Ubr1-dependent manner. (A) Turnover rates of N-terminally destabilized Ste6*C-I2X mutants in $\Delta san1\Delta ubr1$ cells were assessed by metabolic pulse-chase analysis as described in Fig. 1. Degradation of Ste6*C-I2I in $\Delta san1+UBR1$ cells was also examined as a control. Results are mean \pm s.d. for three independent experiments ($n=3$ biological replicates). (B) Heat map reflecting turnover rates of Ste6*C-I2X substrates in $\Delta san1\Delta ubr1$ as presented in Fig. 3A. ** $P<0.01$ (Student's t -test).

N-end rule. It is important to note that unconventional N-terminal Met 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 Met residues, and resultant exposure, Nt-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 a 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 Met cleavage, resulting in N-terminally exposed His, Lys or Arg, as primary destabilizing residues, or Gln, a tertiary destabilizing residue (Varshavsky, 2011). Unconventional initiating Met removal mechanisms in the case of P2 Lys and Arg residues have yet to be demonstrated. Thus, the above experiments suggest that P2 His, Lys, Arg and Gln residues on a misfolded substrate may either induce a non-conventional N-terminal Met excision that produces 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 Met 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 Met residues are retained for these variants, the 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 and Gln, with the initiating Met retained, or P2 Gly and Val, with the initiating Met 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 N-degron sequences recognized by Ubr1. 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 the current experiments is that non-conventional N-terminal Met 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 a P2 His, Lys, Arg, Gln, Gly or Val residue were moderately or highly stable. The lack of correlation between the results obtained with mCherry-sfGFP reporters and those achieved using the misfolded substrates here suggests that conformational aberrancy may be a key determinant in N-degron compatibility with Ubr1, which is explored later.

Interestingly, the mutation of P2 Leu to Lys in the case of $\Delta ssC^{22-519}Leu2_{myc}$ inhibited its Ubr1-mediated degradation (Kim et al., 2014), which is in contrast to what was seen for 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 N-terminus of the substrate was induced by the mutation, thus inhibiting recognition by Ubr1. So although a concomitant change in acetylation state was not ruled out for $\Delta ssC^{22-519}Leu2_{myc}$ 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 activity of the Arg/N-end rule 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 whether there is a bias within specific cellular compartments for proteins to possess P2 residues that are Ubr1-QC-compatible, data from a genome-wide GFP-fusion-based localization study of protein localization spanning 4156 proteins was utilized (Huh et al., 2003) (Tables S2–S4; Tran, 2019). P2 amino acid usage frequency was analyzed for proteins with exclusive nuclear, cytosolic, secretory pathway, or mitochondrial localization (Tables S1, S5–S8, Tran, 2019; also see Materials and Methods). Frequency distribution

across all residues was significantly different for secretory ($n=559$, $P<1\times 10^{-6}$, $\chi^2=64.33$, d.f.=19) and mitochondrial proteins ($n=458$, $P<1\times 10^{-55}$, $\chi^2=317.85$, d.f.=19) when compared against cytosolic proteins ($n=823$) (Fig. 4A–C; Tables S9 and S10; Tran, 2019). Pair-wise χ -squared analysis of individual amino acid usage frequencies demonstrated that biases were seen for mitochondrial and secretory proteins to be encoded with P2 residues that are Ubr1-QC-compatible, with six 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 mitochondrial and secretory pathway protein sets when compared with cytosolic proteins. An equally important finding is that there was a significant usage bias against seven amino acids that do not enable efficient Ubr1 degradation (Ubr1-QC-incompatible P2 residues: Ala, Ser, Cys, Glu, Pro, Asp and Thr) in either one or both mitochondrial and secretory pathway protein sets when compared to cytosolic proteins. In contrast, the frequency distribution between nuclear and cytosolic proteins had a much lower significant difference (nuclear, $n=639$; cytosolic $n=823$; $0.005<P<0.05$, $\chi^2=36.33$, d.f.=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; Tran, 2019).

The total relative abundance of proteins within each localization category was determined 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; Tran, 2019).

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 proteins. 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 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 (Neupert and Herrmann, 2007; Laminet and Plückthun, 1989). Thus, a

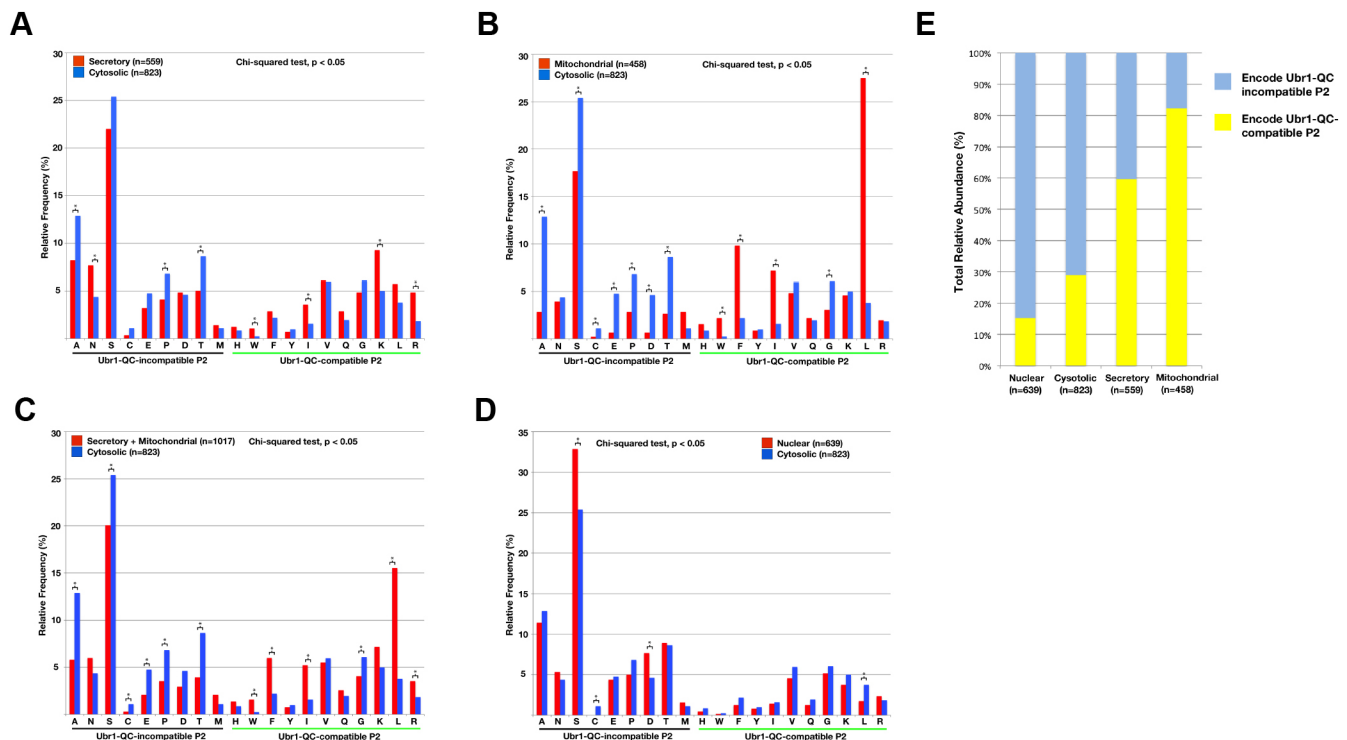


Fig. 4. Proteins localized to secretory pathway components and mitochondria exhibit a bias for encoding UBR1-QC-compatible P2 residues when compared to cytosolically localized proteins. (A) Relative frequency of P2 amino acid usage of proteins localized to the secretory pathway components ($n=559$, proteins) versus cytosol ($n=823$, proteins). Frequency distribution differed significantly between the two sets ($P<1\times 10^{-6}$, $\chi^2=64.33$, d.f.=19; Table S9; Tran, 2019). Pair-wise χ -squared 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) versus cytosol ($n=823$, proteins). Frequency distribution differed significantly between the two sets ($P<1\times 10^{-55}$, $\chi^2=317.85$, d.f.=19; Table S9; Tran, 2019). 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) versus cytosol ($n=823$, proteins). Frequency distribution differed significantly between the two sets ($P<1\times 10^{-30}$, $\chi^2=196.34$, d.f.=19; Table S10; Tran, 2019). Pair-wise χ -squared analysis: $*P<0.0001$ for A, T, F, I, L; $P<0.01$ for S, E, P, W; $P<0.05$ for C, G, R. (D) Relative frequency of P2 amino acid usage in proteins localized to nucleus ($n=639$, proteins) versus cytosol ($n=823$, proteins). Frequency distribution differed less significantly between the two sets ($0.005<P<0.05$, $\chi^2=36.33$, d.f.=19; Table S10; Tran, 2019). Pair-wise χ -squared 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; Tran, 2019).

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, ATP2 Δ 1,2,3, a translocation-defective mutant of Atp2p that is import deficient and degraded in the cytosol (Bedwell et al., 1987), was examined. ATP2 Δ 1,2,3 carries a

Ubr1-QC-compatible P2 Val residue. Degradation was dependent on a combination of San1 and Ubr1 (Fig. 5A). The P2 Val residue was replaced with a Ubr1-QC-incompatible Ser residue, and the P3 Leu was also mutated to a Pro residue, producing ATP2 Δ 1,2,3-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 ATP2 Δ 1,2,3-V2S,L3P was retarded when compared to that of ATP2 Δ 1,2,3 in Δ san1+UBR1 cells, while both forms of the substrate were stabilized to similar levels in Δ san1 Δ ubr1 cells, indicating that the decreased turnover

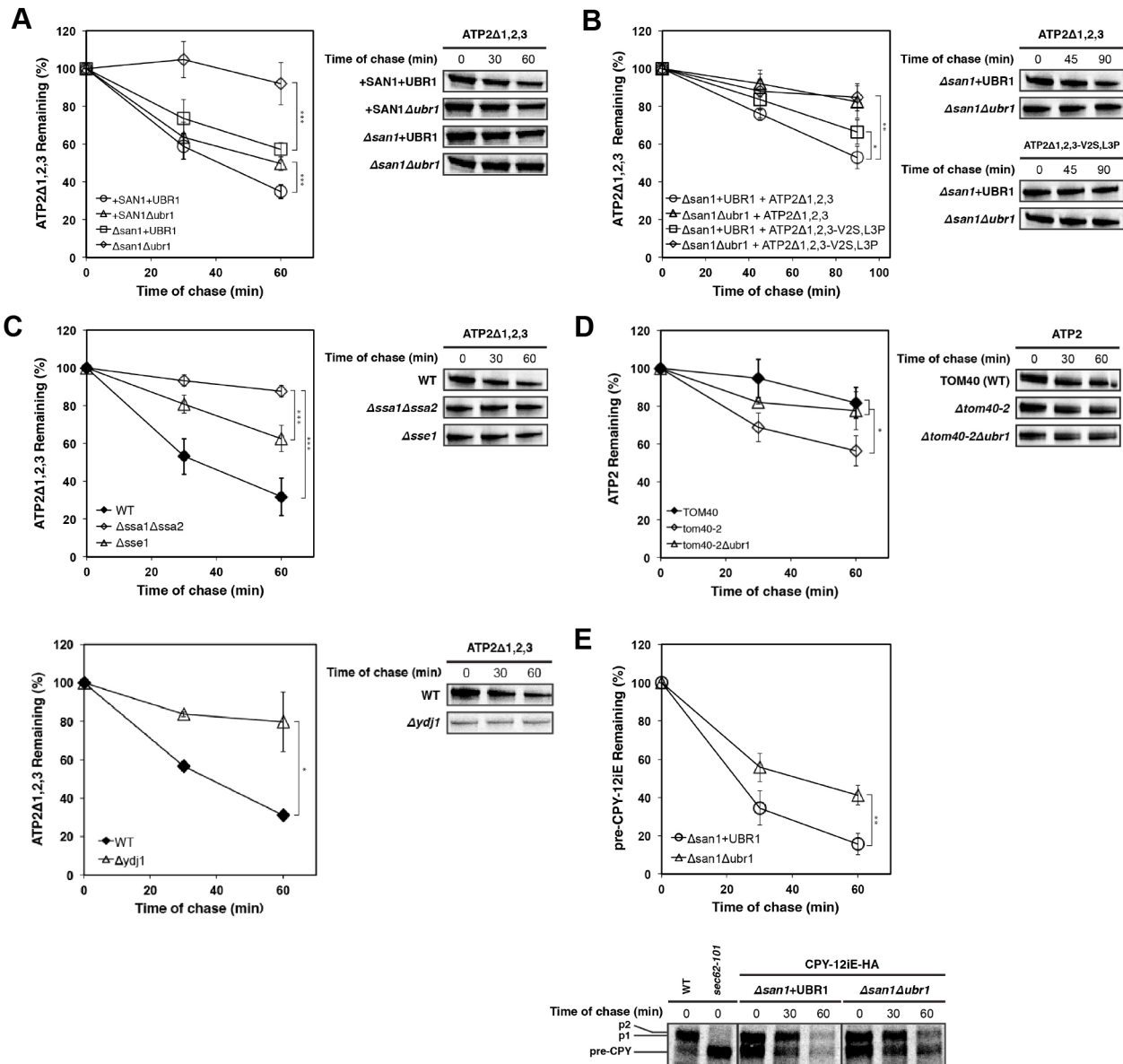


Fig. 5. Mistranslocated forms of mitochondrial and secretory proteins are substrates of cytosolic QC mediated by Ubr1 and the N-end rule pathway. (A) Turnover rate of ATP2 Δ 1,2,3 in +SAN1+UBR1 (wild-type), Δ san1+UBR1, +SAN1 Δ ubr1, and Δ san1 Δ ubr1 cells was assessed by metabolic pulse-chase analysis ($n=5$, biological replicates). (B) Turnover rates of ATP2 Δ 1,2,3 and ATP2 Δ 1,2,3-V2S,L3P in Δ san1+UBR1 and Δ san1 Δ ubr1 cells were assessed by metabolic pulse-chase analysis ($n=4$, biological replicates). (C) Turnover rate of ATP2 Δ 1,2,3 in wild-type, Δ ydj1, Δ ssa1 Δ ssa2, and Δ sse1 cells was assessed by metabolic pulse-chase analysis. (D) Turnover rate of HA-tagged wild-type ATP2 expressed in TOM40 (wild-type), tom40-2, and tom40-2 Δ ubr1 cells was assessed by metabolic pulse-chase analysis. (E) Turnover rate of CPY-12IE-HA in Δ san1+UBR1 and Δ san1 Δ ubr1 cells was assessed by metabolic pulse-chase analysis. All results (A–E) are for metabolic pulse-chase analysis, performed as described in Fig. 1 with a 10-min pulse-labeling step. Experiments involving temperature-sensitive tom40-2 strains were performed at 37°C for all strains. Samples from Δ san1 Δ ubr1 cells expressing CPY-12IE-HA were exposed to a phosphor screen for an extended period (5 days versus 3 days for Δ san1+UBR1 cells) for imaging and quantitative analysis due to a lower overall detection level of the expressed substrate compared to what was found for Δ san1+UBR1 cells. Results are mean \pm s.d. for three independent experiments ($n=3$, biological replicates) unless otherwise noted. * $P<0.05$; ** $P<0.01$; *** $P<0.005$ (Student's t -test).

rate of ATP2 Δ 1,2,3-V2S,L3P was the result of a reduction in Ubr1-mediated degradation due to the N-terminal mutations (Fig. 5B). It was found 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 ATP2 Δ 1,2,3-V2S,L3P (Fig. S3B). Degradation of ATP2 Δ 1,2,3 was dependent on the Sse1, Ydj1, Ssa1, and Ssa2 chaperones (Fig. 5C), a hallmark of ubiquitin-proteasome system-dependent CytoQC substrates, and a strong indication that ATP2 Δ 1,2,3 is misfolded (Heck et al., 2010; Prasad et al., 2010). Degradation of wild-type Atp2p in the temperature-sensitive mitochondrial-import mutant *tom40-2* (Krimmer et al., 2001), was also dependent on Ubr1 CytoQC (Fig. 5D), suggesting that wild-type Atp2p 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, wild-type GFP was examined, which is expected to escape CytoQC-based degradation as it is a stable protein in yeast. The P2 Ser of GFP was replaced with Ile to generate GFP-S2I. GFP-S2I was significantly less susceptible to Ubr1-mediated degradation than Δ 2GFP-S2I, which was degraded efficiently by Ubr1 (Fig. 6A). Trypsin digestion confirmed the relative structural stability of GFP-S2I versus Δ 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 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., 2014; Shemorry et al., 2013; Hwang et al., 2010). Alternatively, it may be due to an increased availability of polyubiquitylation-competent Lys residues in a misfolded substrate, as one of the requirements of Ubr1-dependent degradation is access to a Lys residue in an unstructured region of the substrate (Varshavsky, 2011).

To determine whether translocation-defective secretory pathway proteins are also subject to Ubr1 CytoQC, an import-impaired mutant of vacuolar carboxypeptidase Y (CPY), CPY-12iE-HA, in which a Glu residue is inserted at position 12 to disrupt the hydrophobic core of the native CPY signal sequence, was generated. Since any mature, vacuole-processed CPY-12iE-HA would prevent accurate analysis of its pre-form levels, the fact that C-terminal processing of CPY-HA results in HA tag removal upon maturation was utilized (Fig. S2). CPY encodes a Ubr1-QC-compatible P2 Lys. Degradation of pre-CPY-12iE was significantly reduced upon the absence of Ubr1p (Fig. 5E, Δ san1+UBR1 versus Δ san1 Δ ubr1), 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

Studies 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 could be degraded through these pathways. To assess whether Ubr1 N-end rule-based CytoQC, together with quality control pathways targeting hydrophobic domains, could provide an effective degradation system for clearing the cytosol of translocation-deficient, mislocalized proteins, a set of 277 open-reading frames (ORFs) encoding signal sequence-containing proteins, used previously to determine the frequency of amino acids encoded at the P2 position of secretory proteins, was analyzed (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 hydrophobicity 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; Tran, 2019). 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); $P<5\times 10^{-5}$, $\chi^2=16.8$). 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 (χ -square test: $P<5\times 10^{-5}$, $\chi^2=16.8$), as well as with the full set of signal-sequence bearing proteins ($P<0.005$, $\chi^2=9.7$): 92.8% (90/97) of the soluble protein set possess Ubr1-QC-compatible P2

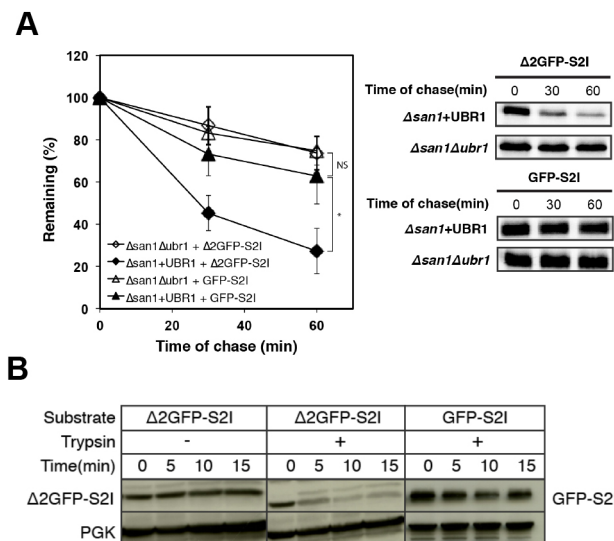


Fig. 6. The efficient P2 residue-based Ubr1-mediated degradation of a substrate is dependent on concomitant presence of structural aberrancy. (A) Turnover rates of GFP-S2I and Δ 2GFP-S2I in Δ san1 Δ ubr1 and Δ san1+UBR1 cells were assessed by metabolic pulse-chase analysis as described in Fig. 1 ($n=3$, biological replicates). * $P<0.025$; NS, not significant, $P>0.2$ (Student's t -test). (B) Trypsin sensitivity assay ($n=1$): postnuclear lysates were prepared from Δ san1 Δ ubr1 cells and treated with 5.0 μ g/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.

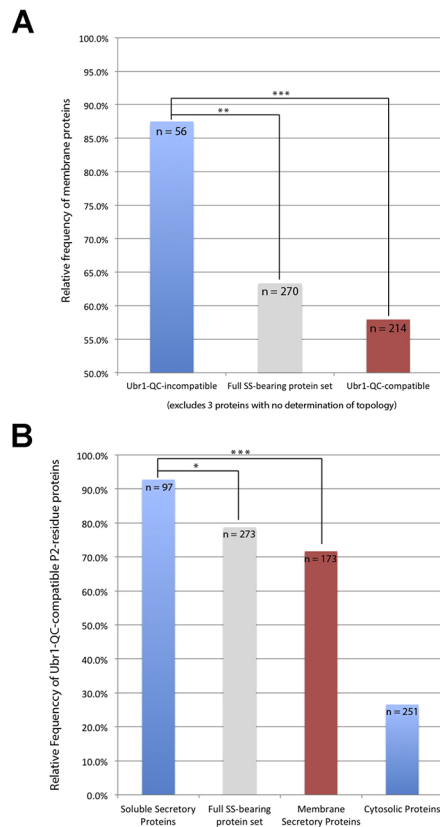


Fig. 7. Secretory proteins encoding Ubr1-QC-incompatible P2 residues are primarily membrane proteins, while soluble secretory proteins primarily encode Ubr1-QC-compatible P2 residues. (A) The percentage of proteins that were membrane-bound was calculated for three categories derived from the full set 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–S18; Tran, 2019). χ -squared test: Ubr1-incompatible ($n=56$, proteins) versus full set ($n=270$, proteins): $^{**}P<0.001$, $\chi^2=11.7$, d.f.=1; Ubr1-incompatible ($n=56$, proteins) versus Ubr1-compatible ($n=214$, proteins): $^{***}P<5\times 10^{-5}$, $\chi^2=16.8$, d.f.=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; Tran, 2019). χ -squared test: soluble ($n=97$, proteins) versus full set ($n=273$, proteins): $^{*}P<0.005$, $\chi^2=9.7$, d.f.=1. Soluble ($n=97$, proteins) versus membrane set ($n=173$, proteins): $^{***}P<5\times 10^{-5}$, $\chi^2=16.8$, d.f.=1. The cytosolic protein set ($n=251$, proteins) was from Forte et al., 2011.

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; Tables S12–S18; Tran, 2019). 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; Tran, 2019). 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 versus 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 acetylates 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 the characterization of Ubr1–P2 residue compatibility, is provided in the Discussion.

The biased distribution of Ubr1-QC-compatible versus Ubr1-QC-incompatible P2 residues that was 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. It would therefore be expected that the minority of secretory proteins that do not possess Ubr1-QC-compatible P2 residues (~21.2%; Fig. 7B) are primarily membrane proteins, allowing them to be effectively degraded by pathways recognizing membrane domains. Indeed, it was 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, a strong pressure for soluble secretory proteins to possess Ubr1-QC-compatible P2 residues would be expected, since they are not as amenable to membrane domain-dependent degradation pathways. This turns out to be true, based on the finding that ~93% of soluble secretory proteins possess Ubr1-QC-compatible P2 residues (Fig. 7B). These latter two results demonstrate further that across the proteome, P2 residues function as cellular locations signals that 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 suggest 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 for them to be efficiently cleared from the cytosol. Collectively, these findings suggest that a majority of proteins in the eukaryotic proteome possess P2 residues that are utilized as *de facto* cellular location signals that help to ensure the fidelity of protein localization. I term these P2-encoded cellular location signal residues (P2CLS residues). This system of P2-encoded location signaling facilitates efficient N-end rule/Ubr1-mediated degradation of the 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 (Gamerding et al., 2015). Such aberrant cross-organelle mistargeting may also be mitigated by the P2CLS-dependent Ubr1 CytoQC pathway, capturing such proteins in the cytosol before mistargeting occurs.

While it was demonstrated here that mistranslocated secretory and mitochondrial proteins trapped in the cytosol have the highest propensity for targeting by the Arg/N-end rule, ~30% of the cytosolic proteome also encode Ubr1-QC-compatible P2CLS-residues (Figs 4E and 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 some capacity for Ubr1-mediated degradation was still observed even in the case of substrates with Ubr1-QC-incompatible P2CLS residues, such as $\Delta 2$ GFP. However, degradation by Ubr1 is significantly enhanced when a Ubr1-QC-compatible P2CLS residue is present (Fig. 1). A similar pattern of effect would be expected 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, as follows.

First, N-terminally HA-tagged Tpk2 (HA–Tpk2) is strongly stabilized upon the deletion of Ubr1 when folding is inhibited by treatment with geldanamycin, with ~37% of substrate remaining in wild-type cells at the 40 min time-point, versus more than 65% of substrate remaining in $\Delta ubr1$, a ~28 percentage point difference due to deletion of Ubr1 (Nillegoda et al., 2010). The HA tag possesses a P2 Tyr residue, which here is detected as a Ubr1-QC-compatible P2. On the other hand, another substrate they analyze that does not possess a Ubr1-QC-compatible P2, Ste11AN^{K44R}, only exhibits a ~5 percentage point difference in remaining substrate by the 60-min time-point upon deletion of Ubr1 relative to wild type (~20% versus ~15%, respectively).

Second, Summers et al., analyzed the turnover of short-lived GFP (slGFP) and observed a substantial difference in slGFP remaining after 90 min when comparing wild-type and $\Delta ubr1$ strains (<5% remaining versus ~50% remaining, respectively) (Summers et al., 2013). slGFP encodes a P2 Gly, 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 that was discovered here through the mutational analysis of Ste6*_C, helping to demonstrate its applicability to other folding impaired substrates.

Third, 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 h when comparing $\Delta fas1$ to $\Delta fas1 \Delta ubr1$ strains (<10% remaining at 6 h versus >50% remaining, respectively), demonstrating a strong dependency on Ubr1 for degradation (Scazzari et al., 2015).

In addition, Khosrow-Khavar et al. discovered cytosolic temperature-sensitive substrates that exhibited some dependency on Ubr1 for degradation, namely, *ugp1-3* and *gln1-3*. Both do not possess Ubr1-QC-compatible P2 residues (Mayor et al., 2012).

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-h time-point (cycloheximide chase), suggestive of inefficient targeting by Ubr1, as expected based on the model of Ubr1–P2 residue compatibility.

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 h in wild-type cells versus $\Delta ubr1$ (Comyn et al., 2016). Guk1 encodes a Ubr1-QC-incompatible P2 Ser.

Finally, Heck et al., discovered several N-terminally HA-tagged misfolded substrates that were strongly stabilized upon 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 the experiments presented here 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 min time-point and only a minor amount at 60 min (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-S21A3K was assessed here to demonstrate that the P3 proline is not the causal factor in the activation of Ubr1-mediated degradation (Fig. S3B).

Taken together, the data in existing literature reporting on misfolded proteins that are degraded by Ubr1, along with results from the experiments presented here, 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. First, an N-terminal Met was originally classified as a stabilizing residue, and second, 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. The 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. The 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 an even lower usage frequency of Ubr1-QC-compatible P2CLS residues was observed 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 the model of P2CLS-based location signaling and degradation presented in this report (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 this 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 and Martin, 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, I 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 these 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 undertaken 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 (Giglion 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 to the formulation of 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 me 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 the formulation of a 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 three 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 that employed the same protocol. Furthermore, it allowed means and standard deviations to be determined to 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 below.

For targeted bioinformatic analyses of P2 amino acid usage frequencies and topological determinations of secretory pathway proteins based on a combination of literature review and Kyte–Doolittle/Phobius methods, 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) was used. Using a previously vetted and published secretory protein set eliminated any chance of bias for the purpose of this analysis of P2 residue 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 the analysis of P2 amino acid residue frequencies in the cytosol to prevent potential bias in selecting cytosolic proteins for this analysis.

Statistical analysis

A standard of three replicates for biological experiments was performed for 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.

χ -squared 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 χ -square analysis was performed when comparing the relative frequency of a single residue between two protein sets using 1 degree of freedom. Pair-wise χ -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 (Tran, 2019). Strains were tested for contamination before usage in experiments. Anti-HA monoclonal antibody (HA.11) was sourced from Covance (MMS-101R-1000; Princeton, NJ) and used at 1:3000 for western blotting and 1:700 for immunoprecipitation. Monoclonal anti-3-phosphoglycerate kinase (PGK) was sourced from Invitrogen (459250; Carlsbad, CA) and used at 1:10,000. Anti-CPY antibody was a gift from Reid Gilmore (University of Massachusetts, Worcester, MA) and was used at 1:3000.

Metabolic pulse-chase assay

Yeast cells were grown to log phase at 30°C (or 25°C for temperature-sensitive strains). Three OD₆₀₀ units of cells were resuspended in 0.9 ml of SC or SC selective medium and incubated at 30°C (37°C for temperature-sensitive strains) for 30 min. Pulse labeling was then initiated with the addition of 82.5 μ Ci of [35S]Met/Cys (EasyTag EXPRESS 35S, PerkinElmer) for 5 or 10 min depending on the labeling efficiency of the substrate of interest. Label was chased with the addition of excess cold Met and Cys to a final concentration of 2 mM. At the appropriate timepoints, pulse labeling/chase was terminated by the addition of 100% trichloroacetic acid (TCA) to a final concentration of 10%. Immunoprecipitation of samples and resolution by SDS-PAGE were carried out as described previously (Vashist et al., 2001). Phosphor screens exposed to gels (1 to 5 day exposure, depending on substrate expression level) were scanned with a TyphoonTM phosphorimager 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 \pm s.d. 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 OD₆₀₀) 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 OD₆₀₀/ml. 1 ml of this resuspension was transferred to a 2 ml screw-cap tube and homogenized by vortexing for 30 s in the presence of 1 ml of 0.5 mm diameter zirconium beads followed by a 1 min incubation at 4°C. This was performed for five cycles. The homogenate was transferred to 1.5 ml Eppendorf tubes. 0.6 ml of fresh cytosol buffer was used to wash the beads and pooled with the original homogenate. Post-nuclear lysate was isolated by pelleting at 500 *g* for 5 min and transferring the supernatant to a fresh tube. The post-nuclear lysate was incubated at 30°C for 5 min, followed by the addition of trypsin to a concentration of 5 μ g/ml. Samples were vortexed and incubated for 30°C, with 100 μ l aliquots taken at the indicated time points and mixed with 11.1 μ l of 100% TCA in fresh 1.5 ml Eppendorf tubes. Aliquots were kept on ice for 5 min and pelleted at 16,000 *g* for 20 min at 4°C. Supernatant was discarded, sample pelleted again briefly, and supernatant again discarded. The pellet was resuspended in 10 μ l of TCA resuspension buffer (100 mM Tris-HCl pH 11.0, 3% SDS, 1 mM 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 OD₆₀₀/ml in selective media. 800 OD₆₀₀ of yeast cells were harvested at 1467 *g* for 15 min,

washed once with 1 \times PBS, and pelleted again at 1467 *g*. Cells were washed in IP/NP40/DTT/PIC (50 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% NP-40, 0.1 mM DTT, Roche cOmplete Mini EDTA-free protease inhibitor cocktail), pelleted at 1467 *g*, and resuspended in IP/NP40/DTT/PIC at a concentration of 50 OD₆₀₀/ml. 1 ml aliquots of the resuspension were transferred to 2 ml screw-cap tubes and homogenized by beadbeating for 30 s using a Mini-BeadBeater cell disrupter (Biospec Products) followed by a 5 min incubation on ice; beadbeating and incubation on ice was repeated for six cycles in the presence of 1 ml of 0.5 mm diameter zirconium beads. The homogenate was transferred to 1.5 ml Eppendorf tubes, and 0.6 ml of fresh cytosol buffer was used to wash the beads and pooled with the original homogenate. Lysate was cleared by centrifugation at 16,000 *g* and transferred to a fresh tube. Lysate was incubated with 65 μ l of Roche anti-HA affinity matrix per 4 ml of lysate for 2 h or overnight at 4°C. Affinity matrix was spun down at 1467 *g* for 1 min and washed with ice-cold IP/NP40/DTT/PIC three times, followed by one wash with cold IP buffer (50 mM Tris-HCl pH 8, 150 mM NaCl) to remove residual NP40. Bound proteins were eluted from matrix through the addition of protein loading buffer (PLB; Bio-Rad) and subsequent boiling at 100°C for 10 min. Immunoprecipitated proteins were resolved on 12% SDS-PAGE gels and transferred to a PVDF membrane. Membrane was washed with deionized water (three \times 1 min, shaking at 70 rpm) 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 min by shaking at 70 rpm. Membrane was quickly destained with 50% methanol (three \times 1 min, shaking at 70 rpm). 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. (https://downloads.yeastgenome.org/sequence/S288C_reference/orf_protein/archive/orf_trans.20110203.fasta.gz). A PHP script was written to extract from this data file the systematic name of each ORF and the first two residues in its respective protein sequence, and subsequently output this into an SQL database; a list of 5887 ORFs and their N-terminal sequences were produced (Table S1; Tran, 2019). 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 include those assigned the following localization terms: 'nucleus', 'nucleolus' 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 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, the abundance levels of proteins in each subset as quantified in the GFP localization study was totaled and then divided by the aggregate abundance of all proteins within the full localization category (<https://doi.org/10.5061/dryad.r837vv7>; Tran, 2019).

Categorization of signal-sequence bearing proteins

A published set of 277 signal-sequencing bearing proteins was analyzed (Forte et al., 2011) (Table S12; Tran, 2019). Two of this original set were duplicate entries (AIM6/YDL237W was a duplicate of LRC1, FLO11/YIR019C was a duplicate of MUC1) and were not included in the analysis. Two were not present in the SGD database and were also excluded (YCR012C and YJL052C) (Table S13; Tran, 2019). 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;

Tran, 2019). 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 (Petersen 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. Three 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, GFP, ATP2Δ1,2,3, ATP2, and their derivatives, were expressed under control of a high expression, constitutive glyceraldehyde-3-phosphate dehydrogenase (TDH3) promoter in yeast centromeric plasmids. HA-tagged CPY/carboxypeptidase Y (Prc1p) and its derivatives were placed under control of its native constitutive endogenous promoter in yeast centromeric plasmids. stGnd1 and its derivatives were expressed under control of the endogenous Gnd1 promoter in yeast centromeric plasmids. Site-directed mutagenesis (Sawano and Miyawaki, 2000) of the original constructs expressing Ste6**C*, Δ2GFP, GFP, CPY, ATP2Δ1,2,3, and stGnd1 was performed to generate mutant substrates. Plasmids were sequenced to verify that the correct mutations were introduced. All expression constructs utilized an ACT1 terminator. Refer to Tables S31 and S32 for full plasmid and primer sequence lists (Tran, 2019). Plasmids were constructed as follows.

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 (expressing stGnd1-HA): a fragment containing the GND1 promoter and encoding residues 1–150 of the GND1 ORF was PCR amplified from yeast genomic DNA using primers AT59 and AT60. The fragment was digested with NcoI and NotI and ligated to pAT1 digested with NcoI and NotI, generating pAT11. The NcoI site placed the stGnd1 coding sequence in-frame with the vector sequence encoding an HA tag followed by the ACT1 terminator sequence.

pAT33–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): pAT33 through pAT51 were constructed by mutation of the base pairs encoding the second residue of Ste6**C* through site-directed mutagenesis using primers AT21–AT39 and pRP22 as a template.

pAT55 (expressing Δ2GFP-S2I): pAT55 was constructed by mutation of the base pairs encoding the 2nd residue of Δ2GFP through site-directed mutagenesis using primer AT18 and pRP44 as a template.

pAT61 (expressing GFP-HA-S2I): a 741-bp 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 (expressing ATP2-HA): 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 (expressing ATP2-HA): pAT64 was digested with ClaI and XhoI to release a 2728-bp fragment encoding the TDH3 promoter, ATP2-HA, and ACT1 terminator sequences. This fragment was ligated into an empty pRS316 vector digested with ClaI and XhoI to generate pAT65.

pAT66 (expressing ATP2Δ1,2,3-HA): 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 (expressing ATP2Δ1,2,3-HA-V2S,L3P): pAT68 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 P3 position, through site-directed mutagenesis using primer AT275 and pAT66 as a template.

pAT72 (expressing CPY-12iE-HA): pAT72 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 (expressing stGnd1-HA-S2L,A3P): pAT79 was constructed by mutation of the base pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primer AT158 and pAT11 as a template.

pAT80 (expressing stGnd1-HA-S2L,A3K): pAT80 was constructed by mutation of the base pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primer AT159 and pAT11 as a template.

pAT82 (expressing stGnd1-HA-S2L,A3P): pAT82 was constructed by mutation of the base pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primer AT160 and pAT11 as a template.

pAT87 (expressing stGnd1-HA-S2Y,A3P): pAT87 was constructed by mutation of the base pairs encoding the 2nd and 3rd residue of stGnd1 through site-directed mutagenesis using primer AT161 and pAT11 as a template.

Acknowledgements

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 comments on the manuscript. I am grateful to Jan Brix (Universität Freiburg, Germany) for sharing yeast strains. I thank Alexander Varshavsky for his insightful feedback on my thesis data.

Competing interests

The author declares no competing or financial interests.

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).

Data availability

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

Supplementary information

Supplementary information available online at <http://jcs.biologists.org/lookup/doi/10.1242/jcs.231662.supplemental>

References

- Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* **25**, R911–R921. doi:10.1016/j.cub.2015.07.055
- Ast, T., Aviram, N., Chuartzman, S. G. and Schuldiner, M. (2014). A cytosolic degradation pathway, prERAD, monitors pre-inserted secretory pathway proteins. *J. Cell Sci.* **127**, 3017–3023. doi:10.1242/jcs.144386
- Bedwell, D. M., Klionsky, D. J. and Emr, S. D. (1987). The yeast F1-ATPase beta subunit precursor contains functionally redundant mitochondrial protein import information. *Mol. Cell. Biol.* **7**, 4038–4047. doi:10.1128/MCB.7.11.4038
- Brower, C. S., Piatkov, K. I. and Varshavsky, A. (2013). Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. *Mol. Cell* **50**, 161–171. doi:10.1016/j.molcel.2013.02.009
- Buchberger, A., Bukau, B. and Sommer, T. (2010). Protein quality control in the cytosol and the endoplasmic reticulum: brothers in arms. *Mol. Cell* **40**, 238–252. doi:10.1016/j.molcel.2010.10.001
- Chen, S.-J., Wu, X., Wadas, B., Oh, J.-H. and Varshavsky, A. (2017). An N-end rule pathway that recognizes proline and destroys gluconeogenic enzymes. *Science* **355**, eaal3655. doi:10.1126/science.aal3655

- Comyn, S. A., Young, B. P., Loewen, C. J. and Mayor, T. (2016). Prefoldin promotes proteasomal degradation of cytosolic proteins with missense mutations by maintaining substrate solubility. *PLoS Genet.* **12**, e1006184. doi:10.1371/journal.pgen.1006184
- Costa, E. A., Subramanian, K., Nunnari, J. and Weissman, J. S. (2018). Defining the physiological role of SRP in protein-targeting efficiency and specificity. *Science* **359**, 689–686. doi:10.1126/science.aar3607
- Currais, A., Fischer, W., Maher, P. and Schubert, D. (2017). Intraneuronal protein aggregation as a trigger for inflammation and neurodegeneration in the aging brain. *FASEB J.* **31**, 5–10. doi:10.1096/fj.201601184
- Dubnikov, T., Ben-Gedalya, T. and Cohen, E. (2017). Protein quality control in health and disease. *Cold Spring Harb. Perspect. Biol.* **9**, a023523. doi:10.1101/cshperspect.a023523
- Eisele, F. and Wolf, D. H. (2008). Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143–4146. doi:10.1016/j.febslet.2008.11.015
- Fang, N. N., Ng, A. H. M., Measday, V. and Mayor, T. (2011). Huf5 HECT ubiquitin ligase plays a major role in the ubiquitylation and turnover of cytosolic misfolded proteins. *Nat. Cell Biol.* **13**, 1344–1352. doi:10.1038/ncb2343
- Forté, G. M. A., Pool, M. R. and Stirling, C. J. (2011). N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. *PLoS Biol.* **9**, e1001073. doi:10.1371/journal.pbio.1001073
- Fredrickson, E. K. and Gardner, R. G. (2012). Selective destruction of abnormal proteins by ubiquitin-mediated protein quality control degradation. *Semin. Cell Dev. Biol.* **23**, 530–537. doi:10.1016/j.semcdb.2011.12.006
- Fredrickson, E. K., Rosenbaum, J. C., Locke, M. N., Milac, T. I. and Gardner, R. G. (2011). Exposed hydrophobicity is a key determinant of nuclear quality control degradation. *Mol. Biol. Cell.* **22**, 2384–2395. doi:10.1091/mbc.e11-03-0256
- Fredrickson, E. K., Gallagher, P. S., Candadai, S. V. C. and Gardner, R. G. (2013). Substrate recognition in nuclear protein quality control degradation is governed by exposed hydrophobicity that correlates with aggregation and insolubility. *J. Biol. Chem.* **288**, 6130–6139. doi:10.1074/jbc.M112.406710
- Gamerding, M., Hanebuth, M. A., Frickey, T. and Deuerling, E. (2015). The principle of antagonism ensures protein targeting specificity at the endoplasmic reticulum. *Science* **348**, 201–207. doi:10.1126/science.aaa5335
- Gigliante, C., Boularot, A. and Meinnel, T. (2004). Protein N-terminal methionine excision. *Cell. Mol. Life Sci.* **61**, 1455–1474. doi:10.1007/s00018-004-3466-8
- Goetze, S., Qeli, E., Mosimann, C., Staes, A., Gerrits, B., Roschitzki, B., Mohanty, S., Niederer, E. M., Laczkó, E., Timmerman, E., et al. (2009). Identification and functional characterization of N-terminally acetylated proteins in *Drosophila melanogaster*. *PLoS Biol.* **7**, e1000236. doi:10.1371/journal.pbio.1000236
- Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332. doi:10.1038/nature10317
- Haynes, C. M. and Ron, D. (2010). The mitochondrial UPR - protecting organelle protein homeostasis. *J. Cell Sci.* **123**, 3849–3855. doi:10.1242/jcs.075119
- Heck, J. W., Cheung, S. K. and Hampton, R. Y. (2010). Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl. Acad. Sci. USA* **107**, 1106–1111. doi:10.1073/pnas.0910591107
- Henze, K. and Martin, W. (2001). How do mitochondrial genes get into the nucleus? *Trends Genet.* **17**, 383–387. doi:10.1016/S0168-9525(01)00231-5
- Hessa, T., Sharma, A., Mariappan, M., Eshleman, H. D., Gutierrez, E. and Hegde, R. S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* **475**, 394–397. doi:10.1038/nature10181
- Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. and O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691. doi:10.1038/nature02026
- Hwang, C.-S., Shemorry, A. and Varshavsky, A. (2010). N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973–977. doi:10.1126/science.1183147
- Itzhaki, R. F. (2018). Corroboration of a major role for herpes simplex virus type 1 in Alzheimer's disease. *Front. Aging Neurosci.* **10**, 1–11. doi:10.3389/fnagi.2018.00324
- Käll, L., Krogh, A. and Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036. doi:10.1016/j.jmb.2004.03.016
- Kats, I., Khmelinskii, A., Kschonsak, M., Huber, F., Knieß, R. A., Bartosik, A. and Knop, M. (2018). Mapping degradation signals and pathways in a eukaryotic N-terminome. *Mol. Cell.* **70**, 488–501.e5. doi:10.1016/j.molcel.2018.03.033
- Kawahara, H., Minami, H. and Yokota, N. (2013). BAG6/BAT3: emerging roles in quality control for nascent polypeptides. *J. Biochem.* **153**, 147–160. doi:10.1093/jb/mvs149
- Kim, H.-K., Kim, R.-R., Oh, J.-H., Cho, H., Varshavsky, A. and Hwang, C.-S. (2014). The N-terminal methionine of cellular proteins as a degradation signal. *Cell* **156**, 158–169. doi:10.1016/j.cell.2013.11.031
- Kim, J. M., Seok, O. H., Ju, S., Heo, J. E., Yeom, J., Kim, D. S., Yoo, J. Y., Varshavsky, A., Lee, C. and Hwang, C. S. (2018). Formyl-methionine as an N-degron of a eukaryotic N-end rule pathway. *Science* **362**, eaat0174. doi:10.1126/science.aat0174
- Kim, Y., Hipp, M., Bracher, A., Hayer-Hartl, M. and Ulrich Hartl, F. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355. doi:10.1146/annurev-biochem-060208-092442
- Krimmer, T., Rapaport, D., Ryan, M. T., Meisinger, C., Kassenbrock, C. K., Blachly-Dyson, E., Forte, M., Douglas, M. G., Neupert, W., Nargang, F. E. et al. (2001). Biogenesis of porin of the outer mitochondrial membrane involves an import pathway via receptors and the general import pore of the TOM complex. *J. Cell Biol.* **152**, 289–300. doi:10.1083/jcb.152.2.289
- Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132. doi:10.1016/0022-2836(82)90515-0
- Laminet, A. A. and Plückthun, A. (1989). The precursor of beta-lactamase: purification, properties and folding kinetics. *EMBO J.* **8**, 1469–1477. doi:10.1002/j.1460-2075.1989.tb03530.x
- Mayor, T., Fang, N. N., Comyn, S. A., Winget, J. M., Khosrow-Khavar, F. and Ng, A. H. M. (2012). The yeast Ubr1 ubiquitin ligase participates in a prominent pathway that targets cytosolic thermosensitive mutants for degradation. *G3* **2**, 619–628. doi:10.1534/g3.111.001933
- Neupert, W. and Herrmann, J. M. (2007). Translocation of proteins into mitochondria. *Annu. Rev. Biochem.* **76**, 723–749. doi:10.1146/annurev-biochem.76.052705.163409
- Nguyen, K. T., Kim, J.-M., Park, S.-E. and Hwang, C.-S. (2019). N-terminal methionine excision of proteins creates tertiary destabilizing N-degrons of the Arg/N-end rule pathway. *J. Biol. Chem.* **294**, 4464–4476. doi:10.1074/jbc.RA118.006913
- Nillegoda, N. B., Theodoraki, M. A., Mandal, A. K., Mayo, K. J., Ren, H. Y., Sultana, R., Wu, K., Johnson, J., Cyr, D. M. and Caplan, A. J. (2010). Ubr1 and Ubr2 function in a quality control pathway for degradation of unfolded cytosolic proteins. *Mol. Biol. Cell.* **21**, 2102–2116. doi:10.1091/mbc.e10-02-0098
- Pearce, M. M. P. and Kopito, R. R. (2018). Prion-Like Characteristics of Polyglutamine-Containing Proteins. *Cold Spring Harb. Perspect. Med.* **8**, a024257. doi:10.1101/cshperspect.a024257
- Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786. doi:10.1038/nmeth.1701
- Polevoda, B. and Sherman, F. (2003). N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J. Mol. Biol.* **325**, 595–622. doi:10.1016/S0022-2836(02)01269-X
- Prasad, R., Kawaguchi, S. and Ng, D. T. W. (2010). A nucleus-based quality control mechanism for cytosolic proteins. *Mol. Biol. Cell.* **21**, 2099–2366. doi:10.1091/mbc.e10-02-0111
- Prasad, R., Kawaguchi, S. and Ng, D. T. W. (2012). Biosynthetic mode can determine the mechanism of protein quality control. *Biochem. Biophys. Res. Commun.* **425**, 689–695. doi:10.1016/j.bbrc.2012.07.080
- Prasad, R., Xu, C. and Ng, D. T. W. (2018). Hsp40/70/110 chaperones adapt nuclear protein quality control to serve cytosolic clients. *J. Cell Biol.* **217**, 2019–2032. doi:10.1083/jcb.201706091
- Rodrigo-Brenni, M. C., Gutierrez, E. and Hegde, R. S. (2014). Cytosolic quality control of mislocalized proteins requires RNF126 recruitment to Bag6. *Mol. Cell* **55**, 227–237. doi:10.1016/j.molcel.2014.05.025
- Sadis, S., Atienza, C. and Finley, D. (1995). Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.* **15**, 4086–4094. doi:10.1128/MCB.15.8.4086
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harb. Lab. Press.
- Sawano, A. and Miyawaki, A. (2000). Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28**, e78. doi:10.1093/nar/28.16.e78
- Scazzari, M., Amm, I. and Wolf, D. H. (2015). Quality control of a cytoplasmic protein complex. *J. Biol. Chem.* **290**, 4677–4687. doi:10.1074/jbc.M114.596064
- Shemorry, A., Hwang, C.-S. and Varshavsky, A. (2013). Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* **50**, 540–551. doi:10.1016/j.molcel.2013.03.018
- Sherman, F., Stewart, J. W. and Tsunasawa, S. (1985). Methionine or not methionine at the beginning of a protein. *BioEssays* **3**, 27–31. doi:10.1002/bies.950030108
- Stevens, F. J. and Argon, Y. (1999). Protein folding in the ER. *Semin. Cell Dev. Biol.* **10**, 443–454. doi:10.1006/scdb.1999.0315
- Stolz, A., Besser, S., Hottmann, H. and Wolf, D. H. (2013). Previously unknown role for the ubiquitin ligase Ubr1 in endoplasmic reticulum-associated protein degradation. *Proc. Natl. Acad. Sci. USA* **110**, 15271–15276. doi:10.1073/pnas.1304928110
- Strobel, G., Zollner, A., Angermayr, M. and Bandlow, W. (2002). Competition of spontaneous protein folding and mitochondrial import causes dual subcellular location of major adenylate kinase. *Mol. Biol. Cell* **13**, 1439–1448. doi:10.1091/mbc.01-08-0396
- Summers, D. W., Wolfe, K. J., Ren, H. Y. and Cyr, D. M. (2013). The type II Hsp40 Sis1 cooperates with Hsp70 and the E3 ligase Ubr1 to promote degradation of terminally misfolded cytosolic protein. *PLoS ONE* **8**, e52099. doi:10.1371/journal.pone.0052099
- Suzuki, R. and Kawahara, H. (2016). UBQLN4 recognizes mislocalized transmembrane domain proteins and targets these to proteasomal degradation. *EMBO Rep.* **17**, 842–857. doi:10.15252/embr.201541402
- Szoradi, T., Schaeff, K., Garcia-Rivera, E. M., Itzhak, D. N., Schmidt, R. M., Bircham, P. W., Leiss, K., Diaz-Miyar, J., Chen, V. K., Muzzey, D. et al. (2018). SHRED is a regulatory cascade that reprograms Ubr1 substrate specificity for

- enhanced protein quality control during stress. *Mol. Cell.* **70**, 1025-1037.e5. doi:10.1016/j.molcel.2018.04.027
- Tran, A.** (2013). N-terminal Determinants of Cytosolic Quality Control. *PhD thesis*, National University of Singapore, Singapore. <http://scholarbank.nus.edu.sg/handle/10635/51986>.
- Tran, A.** (2019). Data from: The N-end Rule Pathway and Ubr1 enforce protein compartmentalization via P2-encoded cellular location signals. Dryad Digital Repository. <https://doi.org/10.5061/dryad.r837vv7>.
- Varshavsky, A.** (2011). The N-end rule pathway and regulation by proteolysis. *Protein Sci.* **20**, 1298-1345. doi:10.1002/pro.666
- Vashist, S., Kim, W., Belden, W. J., Spear, E. D., Barlowe, C. and Ng, D. T. W.** (2001). Distinct retrieval and retention mechanisms are required for the quality control of endoplasmic reticulum protein folding. *J. Cell Biol.* **155**, 355-367. doi:10.1083/jcb.200106123
- Wingfield, P. T.** (2017). N-terminal methionine processing. *Curr. Protoc. Protein Sci.* **2017**, 2016-2019. doi:10.1002/cpps.29
- Zheng, K., Liu, Q., Wang, S., Ren, Z., Kitazato, K., Yang, D. and Wang, Y.** (2018). HSV-1-encoded microRNA miR-H1 targets Ubr1 to promote accumulation of neurodegeneration-associated protein. *Virus Genes* **54**, 343-350. doi:10.1007/s11262-018-1551-6