



# The C-degron pathway eliminates mislocalized proteins and products of deubiquitinating enzymes

Chi-Wei Yeh<sup>1,†</sup>, Wei-Chieh Huang<sup>1,†</sup>, Pang-Hung Hsu<sup>2</sup> , Kun-Hai Yeh<sup>1</sup>, Li-Chin Wang<sup>1,3</sup>, Paul Wei-Che Hsu<sup>1</sup>, Hsiu-Chuan Lin<sup>1,3</sup>, Yi-Ning Chen<sup>1</sup>, Shu-Chuan Chen<sup>1</sup>, Chen-Hsiang Yeang<sup>3,4</sup> & Hsueh-Chi S Yen<sup>1,3,\*</sup> 

## Abstract

**Protein termini are determinants of protein stability. Proteins bearing degradation signals, or degrons, at their amino- or carboxyl-termini are eliminated by the N- or C-degron pathways, respectively. We aimed to elucidate the function of C-degron pathways and to unveil how normal proteomes are exempt from C-degron pathway-mediated destruction. Our data reveal that C-degron pathways remove mislocalized cellular proteins and cleavage products of deubiquitinating enzymes. Furthermore, the C-degron and N-degron pathways cooperate in protein removal. Proteome analysis revealed a shortfall in normal proteins targeted by C-degron pathways, but not of defective proteins, suggesting proteolysis-based immunity as a constraint for protein evolution/selection. Our work highlights the importance of protein termini for protein quality surveillance, and the relationship between the functional proteome and protein degradation pathways.**

**Keywords** C-degron pathway; CRL2 ubiquitin ligase; N-degron pathway; protein spatial quality control; protein termini

**Subject Category** Post-translational Modifications & Proteolysis

**DOI** 10.15252/emj.2020105846 | Received 5 June 2020 | Revised 11 December 2020 | Accepted 15 December 2020 | Published online 20 January 2021

**The EMBO Journal (2021) 40: e105846**

## Introduction

Faithful interpretation of the genome into the proteome is essential for life, yet proteome fidelity is constantly challenged by genetic mutations, mistranscription and mistranslation, faulty segregation, erroneous folding, and post-translational oxidative damage (Harper & Bennett, 2016; Hipp *et al.*, 2019). For instance, the decoding of diverse sorting signals to route proteins to different cellular compartments involves highly complex processes, which entails a risk of mistargeting (Wang & Li, 2014; Vitali *et al.*, 2018). The failure

rates of protein segregation are as high as ~10% for several membrane proteins (Kim *et al.*, 2002; Rane *et al.*, 2010). Accumulation of mislocalized or defective proteins is cytotoxic and has been implicated in various neurodegenerative disorders (Hetz & Mollereau, 2014; Dubnikov *et al.*, 2017). To defend the proteome, robust proteolysis-mediated quality control systems have evolved to specifically remove abnormal proteins from cells (Chen *et al.*, 2011; Pohl & Dikic, 2019).

Nevertheless, the aforementioned proteolytic mechanisms also represent a threat to the normal proteome. Coexistence of these proteolytic systems and the normal proteome imposes an evolutionary pressure on both. An effective balance must be achieved between the proteolytic systems having to effectively and efficiently eliminate only unwanted proteins but leave necessary proteins intact with the need for the functional proteome to evolve evasive mechanisms to prevent proteolytic destruction. A deeper understanding of the former process, i.e., the mechanism by which proteolytic systems recognize and clear aberrant proteins, is quickly emerging (Petroski & Deshaies, 2005; Williamson *et al.*, 2013; Zheng & Shabek, 2017). However, our comprehension of the “compatibility” between protein surveillance systems and the functional proteome, and how functional proteins prevent proteolysis, remains limited.

Protein termini are determinants of protein degradation (Varshavsky, 2019). Proteins containing N-terminal degradation signals (N-degrons) are degraded by N-degron pathways (Sriram *et al.*, 2011; Varshavsky, 2011). N-degrons are created by proteolytic cleavage or enzymatic N-terminal-modification of proteins (Hwang *et al.*, 2010; Piatkov *et al.*, 2014). We previously discovered a CRL2 ubiquitin ligase-mediated C-degron pathway, DesCEND (Destruction *via* C-End Degron), that selectively eliminates proteins with deviant C-termini, i.e., C-degrons (Lin *et al.*, 2015; Lin *et al.*, 2018). In line with its function in protein quality surveillance, this C-degron pathway eliminates truncated selenoproteins generated by translation miscoding and the N-terminal fragment of USP1 from autocleavage. C-degrons are portable short motifs and comprise a few conserved

1 Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

2 Department of Life Science, Institute of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan

3 Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan

4 Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

\*Corresponding author. Tel: +886 2 2788 3593; +886 2 2789 9212; E-mail: hyen@imb.sinica.edu.tw

†These authors contributed equally to this work

residues, with the remainder being rather flexible. Glycine is frequently the last residue of C-degrons. An extreme C-terminal position and structural accessibility are absolutely essential for C-degrons to trigger degradation. CRL2 exploits interchangeable BC-box proteins to target distinct C-degrons, i.e., KLHDC2, KLHDC3, and APPBP2 recognize the diGly, R\_G and \_R\_G\_ C-degrons, respectively (Koren *et al*, 2018; Lin *et al*, 2018; Rusnac *et al*, 2018). A detailed characterization of C-degrons and additional functions of C-degron pathways has yet to be achieved.

In this study, we developed a context-independent random peptide platform to characterize the features of C-degrons. We further demonstrate novel functions of C-degron pathways in protein spatial quality control and clearing the products of deubiquitinating enzymes. We characterize the first example of a single protein simultaneously harboring both an N-degron and a C-degron, and we showcase two modes of collaborative action between N-degron and C-degron pathways in protein elimination. Lastly, we reveal that coronaviruses have adapted to resist proteolysis attack by C-degron pathways and provide evidence to support that avoiding C-degron-mediated degradation shapes eukaryotic protein evolution/selection.

## Results

### C-degron surrounding sequences affect the degree of degradation

The conventional approach to characterizing degrons is *via* identification of consensus residues from alignments of multiple degrons coupled with mutagenesis analysis. Nevertheless, it remains challenging to uncover complex or excluded features when the number of known degron sequences is limited. Furthermore, conclusions drawn from specific mutagenesis analyses can be template- and amino acid-dependent, yet it is laborious to perform mutagenesis on multiple templates and to change each targeted residue to various different amino acids.

To evaluate unbiasedly the role of the extreme C-termini of protein sequences for their stability, we adapted a “context-independent” Global Protein Stability (GPS) random peptide platform. GPS is a fluorescence-based reporter for measuring protein stability with single-cell resolution (Yen *et al*, 2008). We fused a 12-residue random peptide library at the C-terminus of the fluorescence tag GFP, with another fluorescent tag (RFP) serving as a control to normalize protein synthesis (Fig 1A). We introduced our GPS random peptide library into the genome of human HEK293T cells *via* lentiviral transduction at a low multiplicity of infection (MOI) so that each GPS reporter cell only carried one unique GFP fusion peptide. The GFP/RFP ratio, as measured by fluorescence-activated cell sorting (FACS), represented a proxy for the stability of GFP fusion proteins. Cells carrying specific peptides that promote GFP degradation display decreased GFP/RFP ratios, and the percentage of library cells with a reduced GFP/RFP ratio (% degradation, Fig 1A) reflects the percentage of random peptides in the library that can function as C-degrons. Instead of examining loss-of-activity by mutating the targeted feature on a particular template sequence, this method compares activities with or without “adding-in” a targeted feature under randomized sequence backgrounds, thereby enabling

the effect of this targeted feature to be assessed across a heterogeneous pool. To avoid data biases due to uneven or insufficient variant representation, we constructed multiple independent peptide libraries for each examined C-terminal residue or motif, and then analyzed % degradation for each library by multiple  $10^5$ -sized random samplings (Fig 1B–F). Since the sample space of all 12-residue peptides ( $20^X$ , where X is the number of random amino acids and ranges between 5 and 12) far exceeds the size of random sampling ( $10^5$ ), the small standard deviations in Fig 1 further support that the features uncovered by our assay are truly independent of the identity of their neighboring residues.

Approximately 23% of peptides ending in random amino acids (i.e.,  $X_{12}$ ) promoted GFP degradation (Fig 1B), highlighting the potency of protein C-termini in controlling protein stability. To assess the role of the last C-terminal residue for protein half-lives, we constructed twenty GFP-random peptide libraries, each terminating with a distinct amino acid. Consistent with our observations that C-degrons frequently end with Gly and occasionally Ala (Fig EV1A; Lin *et al*, 2018), we found that random peptides ending in Gly or Ala were more likely to stimulate degradation, respectively (Fig 1B). Moreover, Gly-dependent degradation was mediated by the CRL2 ubiquitin ligase, consistent with CRL2 playing a dominant role in C-degron pathways (Fig 1C; Lin *et al*, 2015). The KLHDC2, KLHDC3, and APPBP2 substrates we identified contain a conserved C-terminal diGly, R\_G or \_R\_G\_ motif, respectively (Fig EV1A). Next, we examined if the diGly, R\_G and \_R\_G\_ motifs are sufficient to predict corresponding BC-box protein-mediated degradation. As expected, random peptides terminating in these motifs significantly stimulated degradation (Fig 1B right), and the respective degradation was inhibited only by knocking down their corresponding BC-box proteins (Fig 1D). More than 95% of diGly-ending random peptides stimulated degradation, whereas capping diGly with Leu blocked degradation (Fig 1B), suggesting that exposure of the terminal diGly motif alone is typically adequate for KLHDC2 degron recognition. We were unable to completely abolish KLHDC2 function by shRNA-mediated knockdown (Fig 1D) because KLHDC2 binds diGly degrons with exceptionally high affinities ( $\sim$  nM) (Rusnac *et al*, 2018). The upstream sequences of diGly peptides that promoted degradation (as recovered by single-cell sorting) are rather diverse. These flanking sequences do not define but rather tune the potency of degradation (Table EV1).

The R\_G and \_R\_G\_ motifs alone are insufficient to define KLHDC3- and APPBP2-targeted degrons, respectively (Fig 1B and D). We applied an add-in approach to determine their missing features. We appointed specified positions with specific amino acids, with the rest of the positions remaining indiscriminate, and then compared % degradation in cells with or without assigned BC-box proteins (Fig 1E and F). For KLHDC3 degrons, the residue directly upstream of Gly<sup>-1</sup> (i.e., the -2 position) is crucial. It has to be either charged or polar, with positively charged amino acids being particularly appropriate (Fig 1E). All random peptides terminating with the R\_KG motif promoted degradation (Fig 1B) and the R\_KG motif-stimulated degradation is KLHDC3-mediated (Fig 1D). The spacing between Arg and Gly only mildly affected degradation (Fig 1E, bottom four rows). In contrast, Arg-Gly spacing for APPBP2 degrons was rigorous, with two intervening residues being favored (Fig 1F, right panel, rows 4–11 from top). The number of residues downstream of Gly can range from one to three (Fig 1F,



**Figure 1. Sequences nearby C-degrons affect the potency of degradation.**

- A A schematic depiction of our GPS random peptide assay. The GPS reporter relies on the expression of two fluorescent proteins from a single promoter, enabled by an internal ribosome entry site (IRES). A 12-residue peptide library was fused to the C-terminus of GFP, whereas RFP serves as an internal control. The GFP/RFP ratio represents protein stability. The GPS peptide library was delivered into cells as a single copy so that each reporter cell expressed a unique GFP-peptide fusion construct. The “% degradation”, i.e., the % of library cells displaying a decreased GFP/RFP ratio, reflects the % of peptides within the library that promote GFP degradation.
- B The % degradation of random peptide libraries terminating in the indicated amino acids or C-terminal motifs labeled at bottom. X represents any amino acid residue. To illustrate the reliability of this approach, libraries terminating in each specific amino acid or motif were independently constructed three times and each independent library was analyzed by FACS in triplicate (three random samplings with  $10^5$  cells each time). Data are presented as mean  $\pm$  standard deviation from the nine replicates described above. Since the sample space of all 12-residue peptides far exceeds the size of random sampling, the probability of retrieving repeated peptides in random samples is extremely low. Thus, those random samples are expected to carry distinct peptides and should be considered as biological replicates. *P*-values were calculated by unpaired *t*-test.
- C GPS reporter cells expressing peptide libraries ending with a random (X) or Gly (G) residue were treated with dominant-negative (DN) cullins and analyzed. Data are presented as mean  $\pm$  standard deviation from nine replicates.
- D GPS reporter cells carrying indicated libraries were treated with shRNAs against various BC-box proteins and analyzed. VHL and FEM1C serve as unrelated BC-box protein controls. Data are presented as mean  $\pm$  standard deviation from nine replicates.
- E, F The relative % degradation of indicated libraries in cells with or without assigned BC-box proteins. A value of 1 indicates no difference. BC-box proteins were depleted by shRNA-mediated knockdown and FEM1C served as an irrelevant BC-box protein control. Amino acids with different biophysical properties are marked with different colors, i.e., small aliphatic, hydrophobic, basic, acidic and amide residues are orange, green, blue, magenta, and crimson, respectively. Libraries terminating in each specific motif were independently constructed twice and each independent library was analyzed in triplicate. Data are presented as mean  $\pm$  standard deviation from six replicates.
- G GPS reporter cells expressing indicated diGly-ending full-length human proteins were treated with or without an shRNA against KLHDC2 and analyzed by FACS.

right panel, rows 2–4 from top). The electrical property near the Arg and Gly residues significantly influenced APPBP2-dependent degradation, with a positive charge being advantageous but a negative charge being detrimental (Fig 1F, left panel, rows 6–21 from bottom). Consistently, Asp and Glu were absent from previously identified APPBP2 degrons (Fig EV1A). Taken together, KLHDC3 and APPBP2 degrons comprise a conserved R\_G motif and the remaining residues are “comparatively” changeable. Thus, rather than having a strictly conserved sequence of residues, these adjacent variable residues can tune degradation. It would be difficult to uncover this variability by conventional sequence comparison of infrequently encountered substrate or by taking a mutagenesis approach that is template-dependent, but it is readily apparent from GPS random peptide library assays.

Stochastic peptides terminated with diGly or R\_KG effectively triggered KLHDC2- or KLHDC3-mediated GFP degradation, respectively (Table EV1, Fig 1B and D). We wondered if functional human proteins bearing these C-terminal motifs are KLHDC2 or KLHDC3 substrates. Surprisingly, only half of the diGly-ending proteins we examined (5/10) are KLHDC2 substrates (Fig 1G). All three of the cytosolic R\_KG-ending proteins that we tested were not targeted by KLHDC3 (Fig EV1B). Our data suggest that for functional proteins, harboring C-degrons does not guarantee degradation. Thus, mechanisms must exist to prevent such functional proteins from being eliminated by the C-degron pathway, such as by structural masking of terminal diGly, as in ubiquitin and SUMO2 (Lin *et al*, 2018), or other scenarios that we discuss further below.

### Gly/C-degron shortfall is limited to functional eukaryotic proteomes

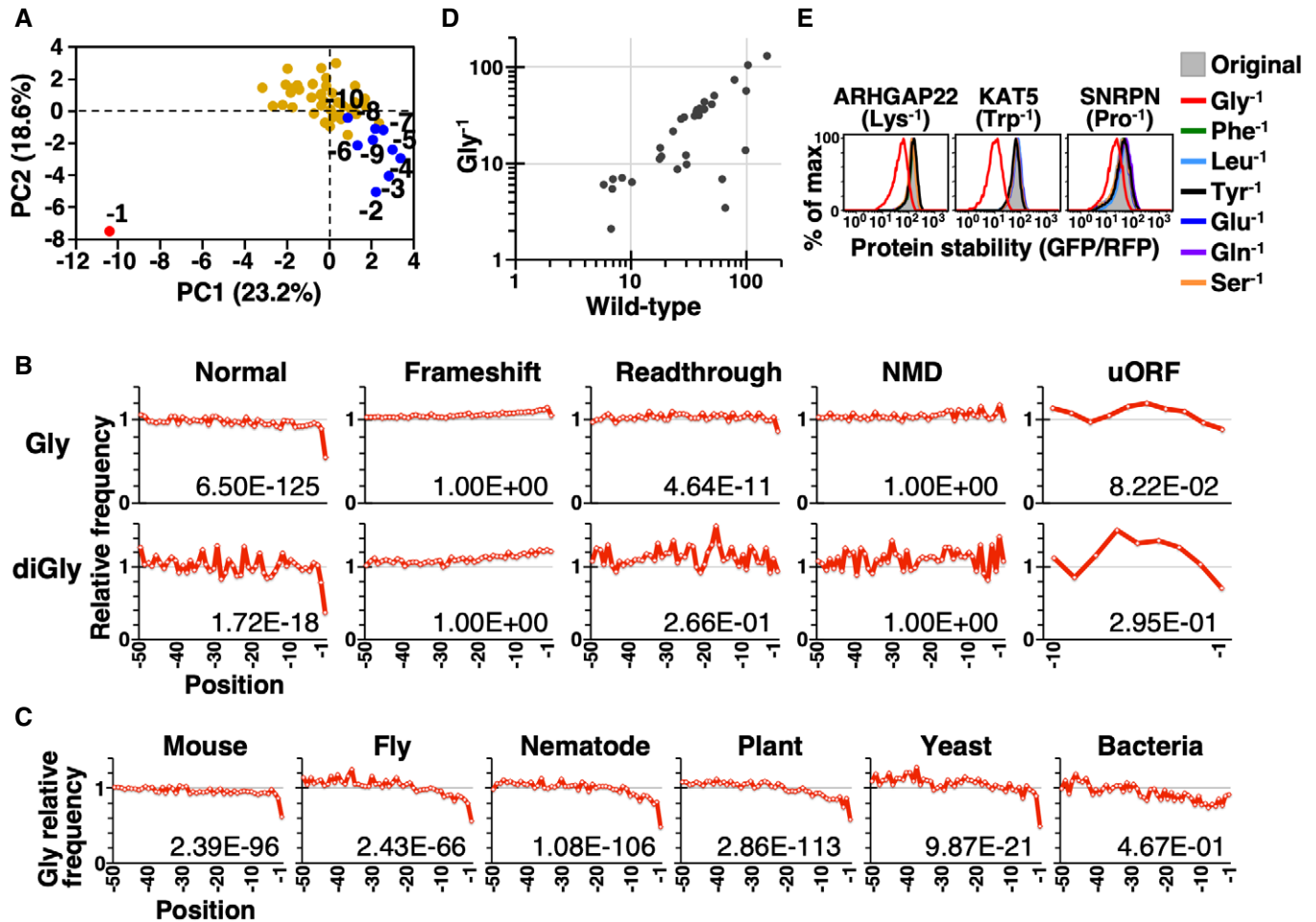
It has been shown that C-degrons are depleted in functional eukaryotic proteomes (Koren *et al*, 2018), but whether this phenomenon is universal or specific to a subset of proteins remains unclear. To investigate the effect of the C-degron pathway on the protein C-terminome, we performed principle component analysis (PCA) on

the amino acid frequency of the final 50 residues of all functional human proteins. We found that amino acid usage of the last five residues deviated from the pattern for the other residues, with the terminal residue (–1) exhibiting the most dramatic difference (Figs 2A and EV1C). Gly is the least preferred amino acid, and Ala is the third least preferred (Table EV2). Intriguingly, this Gly deficit is limited to the last residue of normal proteins conserved across eukaryotic proteomes, but does not occur in abnormal proteins produced from translational frameshift or readthrough errors, NMD (nonsense-mediated mRNA decay) transcripts, or small peptides encoded by upstream open reading frames (uORFs) (Figs 2B top, C, and EV1D left). The pattern for the diGly motif (the KLHDC2 degron) is similar, but is more exaggerated (Figs 2B bottom, and EV1D right, E). We did not perform the same analysis for the R\_G and \_R\_G\_ motifs because the KLHDC3 and APPBP2 degrons are more complex and their ability to stimulate degradation of functional proteins is low (Figs 1E–G and EV1B).

The shortfall of terminal Gly or diGly in functional proteomes may be linked to avoidance of C-degron-mediated protein elimination. To test this hypothesis directly, we selected a panel of cytosolic human proteins and mutated their non-Gly terminal residues into Gly and then examined their protein half-lives. Mutation to Gly frequently reduced the stability of those proteins (18/36 proteins showed  $\geq 15\%$  reduction in stability), indicative of a direct causal relationship between having a Gly terminus and protein instability (Fig 2D, Table EV3). This relationship is Gly-specific (Figs 2E and EV1F). These data collectively suggest that C-degron pathway avoidance may be a driving force for proteome evolution. Thus, in order to prevent the cellular defense mechanism against erroneous proteins targeting functional proteins, strategies have evolved for the proteome to evade C-degron pathways.

### C-degron pathway eliminates mislocalized cellular proteins

To identify the substrates of C-degron pathways, we analyzed the C-termini of human proteins localized in different cellular



**Figure 2. Gly- and diGly-ending proteins are scarce in functional proteomes.**

- A** PCA on the amino acid frequency of the last 50 residues of functional human proteins. Human protein coding sequences were downloaded from NCBI GeneBank. The terminal (-1) and the final 10 residues (-2 to -10) are labeled red and blue, respectively, with the remainder marked yellow.
- B** The relative frequency of Gly (top) and diGly (bottom) in the final 50 residues of functional and aberrant human proteins. We only analyzed the last 10 residues of peptides translated from uORFs due to their inherently smaller size. The number on each graph indicates the left-tailed FDR (false discovery rate)-adjusted *P*-value of the extreme terminal (-1) residue. Analyses using human sequences downloaded from Uniprot database are shown in Fig EV1C and D.
- C** The relative frequency of Gly at the C-termini of indicated proteomes, i.e., mouse (*Mus musculus*), fly (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), plant (*Arabidopsis thaliana*), yeast (*Saccharomyces cerevisiae*), and bacteria (*Escherichia coli*).
- D** Stability (mean GFP/RFP) comparisons between the wild-type protein and its mutant in which the terminal residue has been changed to Gly. Detailed information on tested proteins is presented in Table EV3.
- E** Comparisons of protein stability among wild-type protein and the mutants indicated on the right.

compartments (Fig 3A). Gly<sup>-1</sup> is scarce among cytosolic and nuclear proteins. Intriguingly, we noted a Gly<sup>-1</sup> shortage only for membrane proteins having C-termini facing the cytosol (inward), but not for secreted proteins or membrane proteins having their C-termini facing the extracellular space or ER lumen (outward), suggesting that spatial quarantine protects proteins with C-degrons. We employed model substrates to test if the distribution of a protein in cells dictates its accessibility to C-degron pathways. Firefly luciferase (FLuc) and *Gaussia* luciferase (GLuc) are naturally cytosolic and secreted proteins, respectively (Fig EV2A). We compared their activities with or without fusion with a diGly degron. Although the difference in luciferase activity for FLuc with or without the diGly

degron is > 480-fold (mean ± standard deviation;  $1.18\text{E} + 04 \pm 1.38\text{E} + 03$  versus  $5.69\text{E} + 06 \pm (< 0.01)$ ), it is only two-fold for GLuc ( $1.49\text{E} + 07 \pm 1.73\text{E} + 05$  versus  $3.00\text{E} + 07 \pm (< 0.01)$ ), supporting the notion of localization-restricted degradation. Next, we selected a set of multi-pass transmembrane proteins that are not native KLHDC2 substrates to test the impact of substrate C-terminal topology for C-degron pathway-mediated degradation (Fig EV2B and C). N-terminal GFP tagging does not affect membrane anchoring of these transmembrane proteins (Fig EV2D). Consistently, we found that C-terminal diGly degron-tagging only triggered the destruction of membrane proteins having inward-facing but not outward-facing C-termini (Fig 3B and C), and this degradation

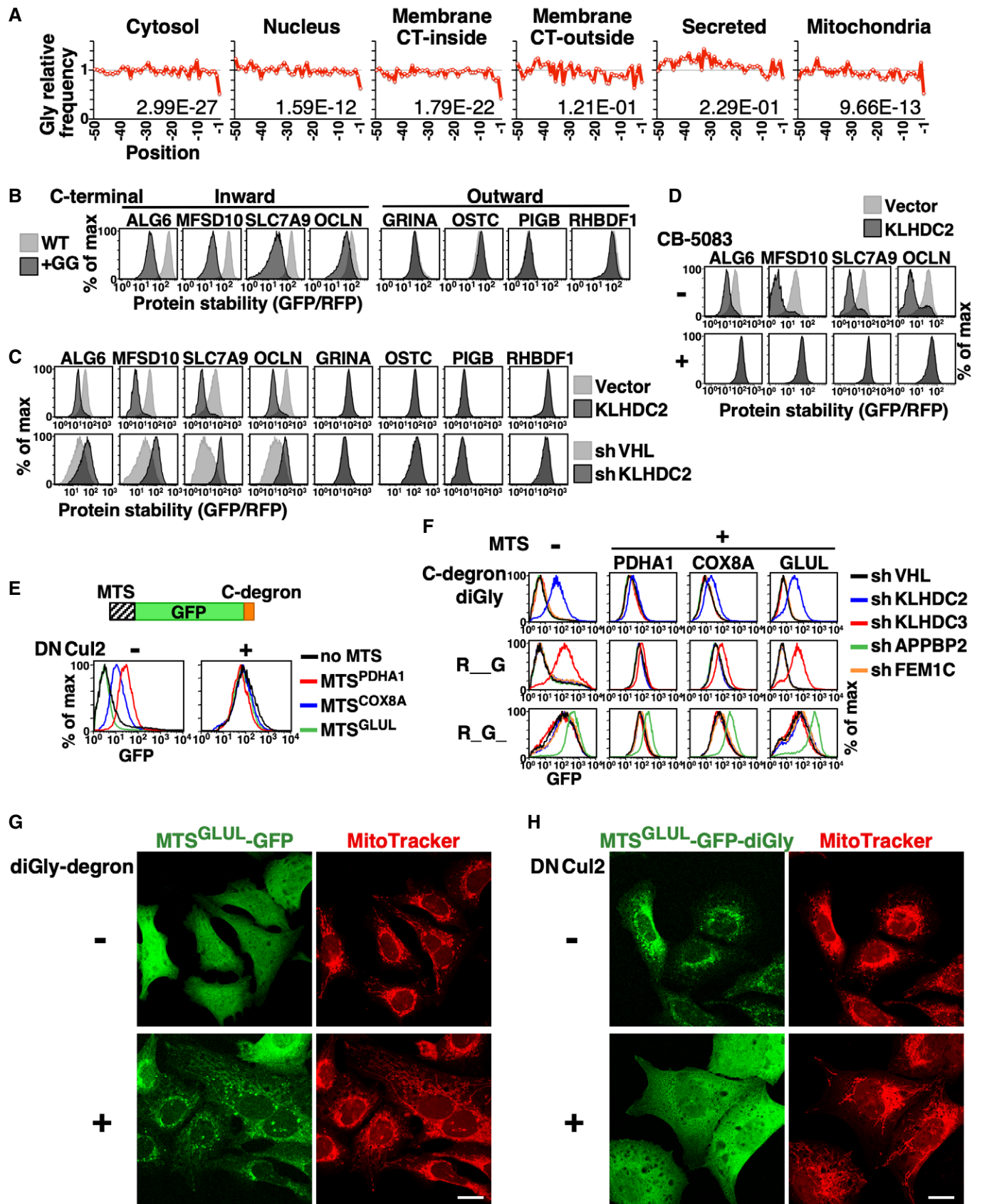


Figure 3.

**Figure 3. The subcellular distribution of a protein dictates its accessibility to the C-degron pathway.**

- A The relative frequency of Gly in the final 50 residues of functional human proteins with different subcellular localizations. The number on each graph denotes the left-tailed *P*-value of the  $-1$  residue. CT-inside and CT-outside represent membrane proteins with their C-termini facing the cytosol or extracellular space, respectively.
- B Comparison of stabilities between wild-type and diGly degron-tagged membrane proteins. The C-terminal topology of tested proteins is indicated at top.
- C GPS reporter cells expressing indicated membrane proteins with a C-terminal diGly degron were treated with or without KLHDC2 overexpression or knockdown and then analyzed.
- D Stability analysis of C-terminal diGly degron-tagged membrane proteins with or without KLHDC2 overexpression and CB-5083 treatment (to inhibit p97/VCP).
- E Abundance analysis of GFP-diGly degron fusion proteins with or without N-terminal tagging with the MTSs of PDHA1, COX8A, or GLUL, and with (+) or without (–) blockage of C-degron pathways by dominant-negative Cul2 (DNCul2).
- F Abundance comparison of MTS-GFP-degron fusion proteins in cells treated with shRNAs against various BC-box proteins. The MTS and C-degrons used are labeled at top and left, respectively. The MTS and C-degron sequences are provided in the Materials and Methods.
- G Confocal images of U2OS cells expressing MTS<sup>GLUL</sup>-GFP proteins with or without a C-terminal diGly degron. MitoTracker staining shows mitochondria. Scale bar = 20  $\mu$ m. DAPI staining and merged images are shown in Appendix Fig S1.
- H Confocal images of U2OS cells expressing MTS<sup>GLUL</sup>-GFP-diGly degron fusion proteins co-stained with MitoTracker, and with or without DNCul2 treatment. Scale bar = 20  $\mu$ m.

required the activity of p97/VCP ATPase (Fig 3D; Meyer *et al*, 2012).

Rather surprisingly, we noted a Gly<sup>-1</sup> scarcity for mitochondrial proteins (Fig 3A). Unlike secreted and membrane proteins that are transported by co-translational targeting, mitochondrial proteins are synthesized in cytosol and then post-translationally delivered to mitochondria. Therefore, we investigated if the C-degron pathway clears mitochondrial proteins that erroneously remain in cytosol. Most proteins that are destined for mitochondria have a mitochondrial targeting sequence (MTS) at their N-terminus (Pfanner & Geissler, 2001). We created model mitochondrial substrates by fusing the MTSs from PDHA1, COX8A, and GLUL proteins to the N-terminus of a GFP-degron fusion reporter (Fig 3E). The efficiency of these MTSs for mitochondrial targeting differ, with that of GLUL (MTS<sup>GLUL</sup>) being the weakest (Fig EV2E, Appendix Fig S1; Matthews *et al*, 2010). We then compared the abundance of these mitochondrial GFP-degron fusion proteins in the presence or absence of functional C-degron pathways (Fig 3E and F). Supporting the notion of localization-selective degradation, the degree of difference was inversely correlated with the efficiency of the MTS tag, i.e., it was positively correlated with the amount of misplaced cytosolic protein. MTS<sup>GLUL</sup>-GFP was localized everywhere in cells, whereas when we added a diGly degron it was only observed in mitochondria (Fig 3G, Appendix Fig S1). Blocking the C-degron pathway rescued the cytosolic distribution of the MTS<sup>GLUL</sup>-GFP-diGly fusion proteins (Fig 3H).

To examine the function of the C-degron pathway in protein spatial quality control, we inferred a panel of physiological secreted and mitochondrial substrates of C-degron pathways based on their C-terminal sequences and subcellular localizations (Fig 4A). Abolition of the predestined localization of those proteins, either by N-terminal capping (Fig 4B–F top) or targeted peptide deletion (Fig 4F bottom, G), specifically stimulated C-degron pathway-dependent protein destruction. Changing the C-termini of these proteins by either deletion or masking completely inhibited C-degron pathway-mediated degradation (Fig 4C). Although APPBP2 knockdown only modestly stabilized BMP5 and BMP7, this effect was not observed for their C-degron mutants ( $\Delta$ G) (Fig 4C). In addition, BMP5 and BMP7, but not their C-degron mutants, were significantly destabilized when the abundance of APPBP2 was elevated (Fig 4D). Moreover, signal peptide-deleted mutant BMP5 and BMP7 ( $\Delta$ SP) accumulated in the cytosol when APPBP2 was inhibited (Fig 4G).

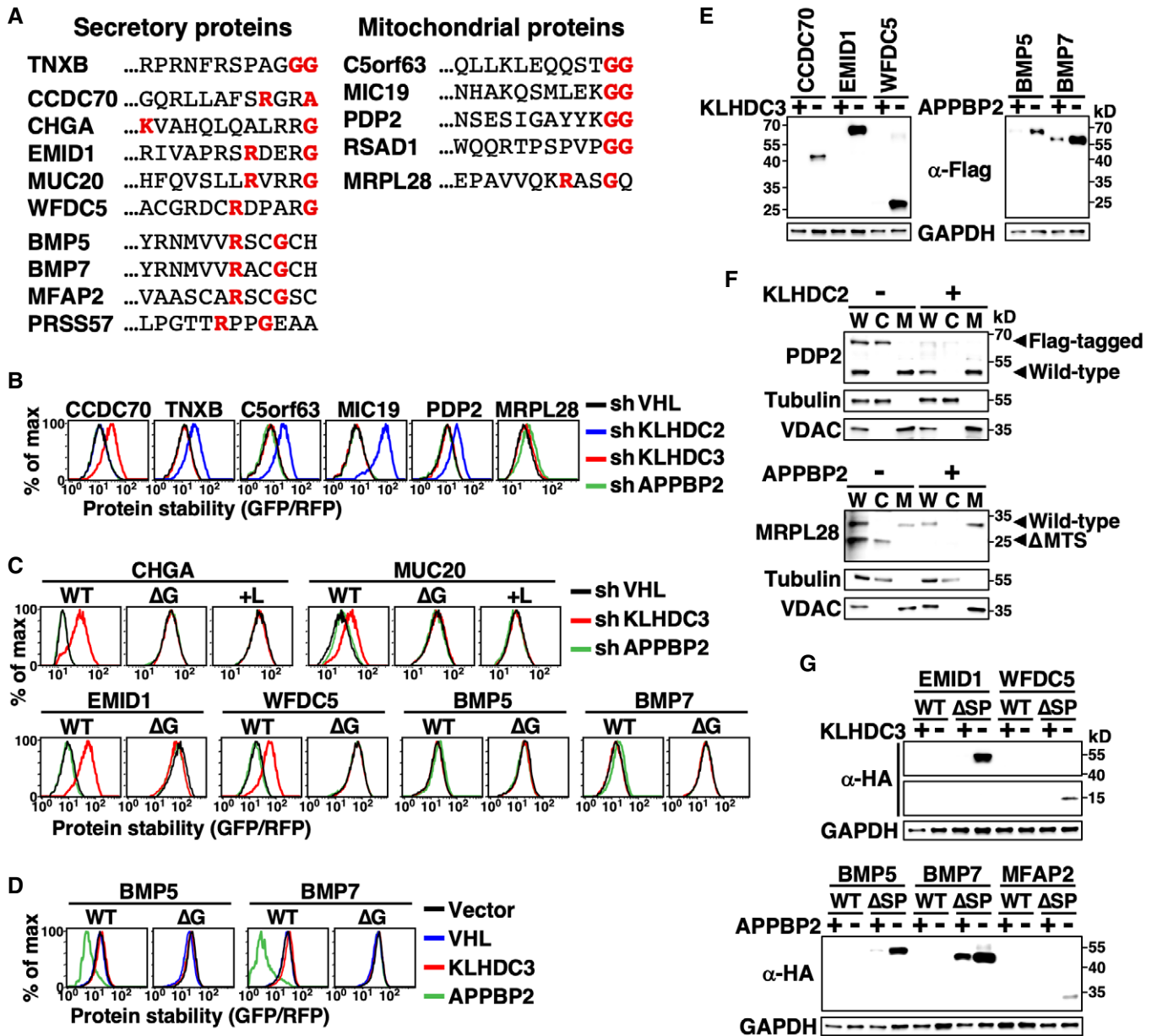
Intracellular protein sorting is known to be error-prone (Levine *et al*, 2004; Vitali *et al*, 2018). Thus, our findings suggest that C-degron pathways potentially eradicate misplaced cellular proteins caused by failed protein targeting.

#### Dual-end inspections safeguard the mitochondrial localization of MIC19

We searched for physiological targets of KLHDC2 by GST pull-down and mass spectrometry-based proteomics (Table 1, Table EV4). Consistent with the idea that C-degron pathways clear mislocalized proteins, we isolated various full-length diGly-ending proteins localized at mitochondria or ER membrane, including MIC19.

MIC19 is a mitochondrial inner membrane protein crucial for crista integrity and mitochondrial function (Darshi *et al*, 2011). Downregulation of MIC19 causes fragmented mitochondria, restricted oxygen consumption and glycolysis, and reduced growth rate. MIC19 orthologs are present throughout metazoans, and their C-terminal diGly tails are highly conserved (Appendix Fig S2). Unlike canonical mitochondrial proteins that rely on MTSs, Gly<sup>2</sup> myristoylation-mediated membrane-targeting and the CHCH domain contribute to the mitochondrial localization of MIC19 (Fig 5A) (Darshi *et al*, 2012; Ueda *et al*, 2019). Consistently, the localization of GFP N-terminally tagged with the 24-residue N-terminal myristoylation motif of MIC19 (MIC19<sup>NT</sup>) resembled that of an ER membrane protein (Figs 5B and EV3A). We examined the role of the diGly/C-degron pathway in localization of endogenous MIC19. MIC19 only resides in mitochondria in normal cells. Nevertheless, blocking the C-degron pathway, either by inhibiting Cul2 or KLHDC2, resulted in trace amounts of MIC19 appearing in the cytosol (Figs 5C and EV3B), suggesting that mitochondrial transport of a proportion of endogenous MIC19 failed. Under normal conditions, that misplaced MIC19 is likely cleared by the C-degron pathway. Facilitating MIC19 mistargeting by treatment with the myristoylation inhibitor IMP-1088 further magnified the abundance of cytosolic MIC19 (Figs 5C and EV3B).

Intriguingly, besides having a conspicuous C-terminal diGly, the N-terminal myristoylation motif of MIC19 resembles a recently described CRL2-mediated Gly/N-degron (Timms *et al*, 2019; Fig 5A), raising the possibility that mitochondrial localization of MIC19 comes under “dual” surveillance by both the N-degron and C-degron pathways. Tagging either the N- or C-terminal motif of



**Figure 4. C-degron pathways clear mislocalized cellular proteins.**

A The extreme C-terminal sequence of indicated secretory and mitochondrial proteins. The diGly, R\_G, and R\_G\_ motifs are labeled red.

B Stability analysis of indicated proteins in cells treated with shRNAs against various BC-box proteins. Since GFP was tagged at the protein N-terminus and signal peptide and MTSs need to be exposed to be functional, all tested proteins were mislocalized.

C, D Protein stability analyses of wild-type (WT) and mutant proteins lacking their extreme terminal Gly ( $\Delta$ G) or having it masked by adding Leu (+L) in cells with or without BC-box protein knockdown (C) or overexpression (D).

E Western blot analysis of Flag-tagged proteins in cells with or without KLHDC3 or APPBP2 expression. The N-terminal Flag abolished protein targeting. GAPDH serves as a loading control.

F Fractionation and Western blot analysis of wild-type and Flag-tagged PDP2 (top), and wild-type and MTS-deleted ( $\Delta$ MTS) MRPL28 (bottom). Flag-PDP2 and  $\Delta$ MTS MRPL28 were mislocalized. W, C, and M denote whole cell lysates, cytosol fraction and mitochondrial fraction, respectively. Tubulin and VDAC serve as fractionation quality controls.

G Protein abundance analysis of HA-tagged wild-type or mutant secretory proteins with their signal peptide deleted ( $\Delta$ SP) in cells with or without KLHDC3 or APPBP2 expression. The HA epitope was inserted immediately after the signal peptide (SP), which had no impact on protein secretion.

Source data are available online for this figure.



**Table 1. KLHDC2-binding proteins identified by GST pull-down and mass spectrometry.**

Gene <sup>a</sup>	DiGly end	Protein name	Subcellular location <sup>b</sup>
<i>AGPAT1</i>	Yes	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	Endoplasmic reticulum
<i>CC2D2A</i>	Yes	Coiled-coil and C2 domain-containing protein 2A	Cytoskeleton; cytosol
<i>CHCHD3</i>	Yes	MICOS complex subunit MIC19	Mitochondrion
<i>DERL2</i>	Yes	Derlin-2	Endoplasmic reticulum membrane; endosome
<i>DNAJB14</i>	Yes	Dnaj homolog subfamily B member 14	Endoplasmic reticulum membrane; nucleus membrane
<i>DNAJC18</i>	Yes	Dnaj homolog subfamily C member 18	Membrane
<i>FAU</i>	No	Ubiquitin-like protein FUBI	–
<i>H4C1</i>	Yes	Histone H4	Nucleus; chromosome
<i>MOCS2A</i>	Yes	Molybdopterin synthase sulfur carrier subunit small subunit	Cytoplasm, cytosol
<i>NMD3</i>	Yes	60S ribosomal export protein NMD3	Nucleus; cytosol
<i>PDGFC</i>	Yes	Platelet-derived growth factor C	Cytoplasm; secreted; nucleus; cytoplasmic granule; cell membrane
<i>PDP2</i>	Yes	Pyruvate dehydrogenase phosphatase catalytic subunit 2	Mitochondrion matrix
<i>PTOV1</i>	Yes	Prostate tumor-overexpressed gene 1 protein	Cell membrane; cytoplasm; nucleus
<i>RSAD1</i>	Yes	Radical S-adenosyl methionine domain-containing protein 1	Mitochondrion
<i>SDE2</i>	No	Replication stress response regulator SDE2	Nucleus
<i>SEC61G</i>	Yes	Protein transport protein Sec61 subunit gamma	Endoplasmic reticulum membrane
<i>SELENOK</i>	No	Selenoprotein K	Endoplasmic reticulum membrane; cell membrane
<i>SELENOS</i>	No	Selenoprotein S	Endoplasmic reticulum membrane; cytoplasm
<i>SLC25A39</i>	Yes	Solute carrier family 25 member 39	Mitochondrion inner membrane
<i>STRADA</i>	Yes	STE20-related kinase adapter protein alpha	Cytosol; nucleus
<i>USP1</i>	No	Ubiquitin carboxyl-terminal hydrolase 1	Nucleus

<sup>a</sup>Listed in alphabetical order.

<sup>b</sup>Based on Uniprot GO (cellular component) or Human Protein Atlas annotations.

MIC19 with GFP promoted CRL2-mediated degradation, confirming that both “ends” of MIC19 are autonomous CRL2-dependent degrons (Fig 5D left). However, these two degrons differ in two ways. Firstly, although the C-degron is insensitive to IMP-1088 treatment, the N-degron is stimulated by IMP-1088, suggesting that myristoylation of Gly<sup>2</sup> abolishes N-degron function. Consistently, we found that a myristoylation-incapable mutant (MIC19<sup>NT</sup>S<sup>6W</sup>, T<sup>7</sup>P-GFP) underwent degradation in the absence of the myristoylation inhibitor (Fig 5B and D). Secondly, the two degrons are recognized by distinct substrate receptors/BC-box proteins, i.e., ZYG11B and KLHDC2 for the N-degron and C-degron, respectively (Figs 5D right, and EV3C). Consistent with the pivotal role of Gly in both degrons, mutating Gly<sup>2</sup> or capping diGly prevented CRL2-mediated degradation (Fig 5D). We noticed that N-degron activity was enhanced when MIC19<sup>NT</sup>-GFP was expressed using the ubiquitin fusion technique (Bachmair *et al*, 1986) (Fig EV3D). Importantly, localization of MIC19<sup>NT</sup>-GFP depends on CRL2 activity. When we inhibited myristoylation, MIC19<sup>NT</sup>-GFP localized at the ER membrane in wild-type cells but became globally distributed upon blocking CRL2 function, suggesting that the CRL2-mediated N-degron pathway eliminated the misplaced cytosolic fraction of MIC19<sup>NT</sup>-GFP (Figs 5E and EV3E, Appendix Fig S3).

Next, we examined the function of the C-degron in full-length MIC19. We placed an HA epitope immediately after the

myristoylation motif in order to analyze exogenously introduced MIC19 variants. The HA insertion did not affect mitochondrial delivery of MIC19 (Fig 5F, top row). We used a MIC19 antibody to detect exogenously expressed HA-MIC19 because the signal from HA antibodies was too weak. We confirmed that levels of endogenous MIC19 were too low to be detected under this staining condition (Appendix Fig S4). Consistent with the idea that the rate of failed mitochondrial targeting is intrinsically low (Fig 5C), solely masking the C-degron (+L) did not notably affect MIC19 localization (Fig 5F second row). Moreover, the limited amounts of cytosolic MIC19 may be removed by the N-degron pathway. When we blocked both MIC19 myristoylation and N-degron activity by means of Gly<sup>2</sup> mutation, we observed contrasting distributions for MIC19 with or without a functional C-degron. In the G<sup>2</sup>A or G<sup>2</sup>S background, MIC19 with a C-degron was primarily localized in mitochondria, whereas MIC19 lacking a C-degron was globally distributed (Figs 5F and EV4A). We obtained the same outcome by means of a biochemical fractionation assay (Fig EV4B left), suggesting that a C-degron pathway mediates spatial quality control of MIC19. Mitochondrial enrichment of MIC19 was more apparent by image analysis because the volume of the cytosol is much greater than that of mitochondria (> 10-fold) (Posakony *et al*, 1977), and imaging signals reflect protein “concentrations”, whereas fractionation assays measure “total protein abundance”.

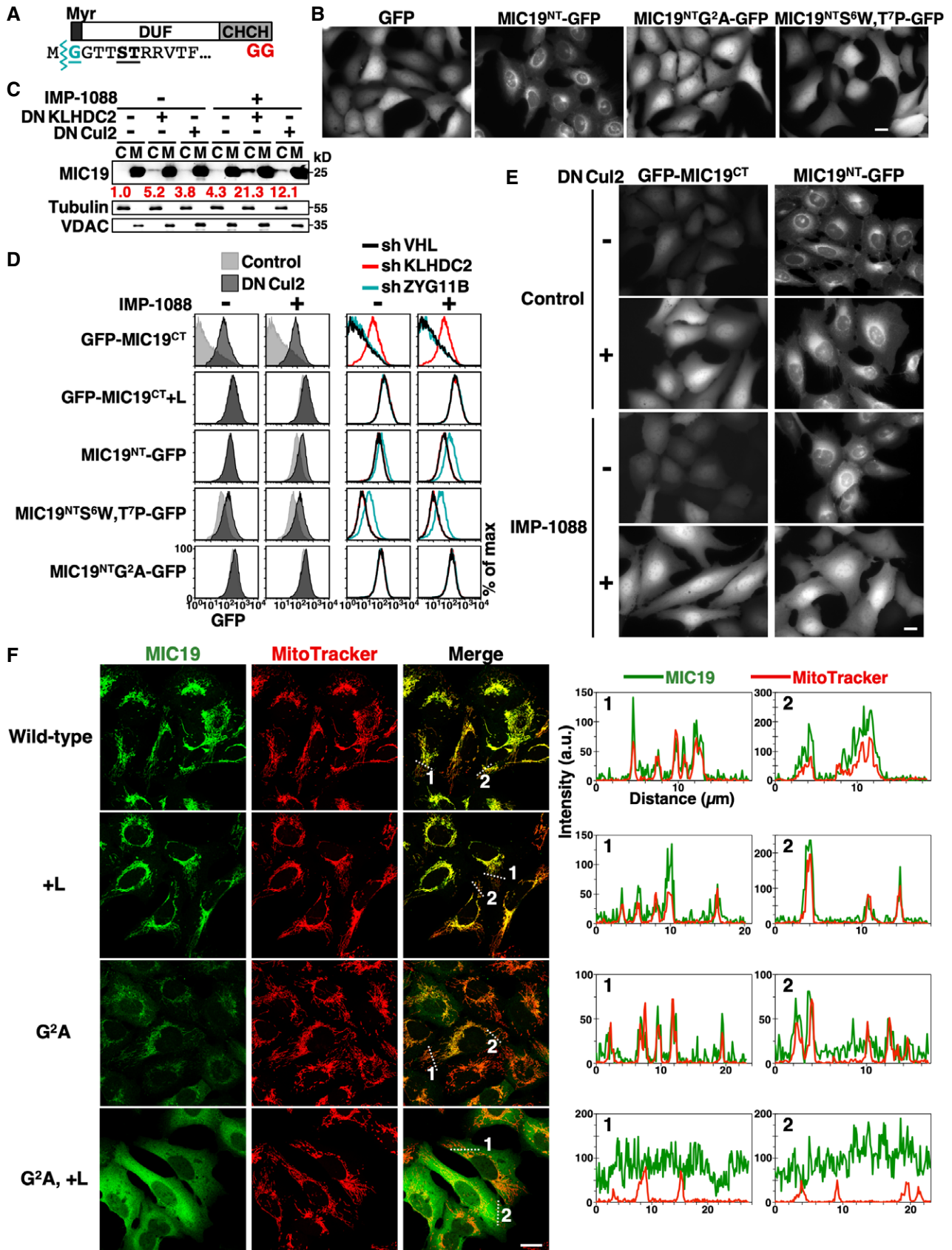


Figure 5.

**Figure 5. MIC19 is regulated by both N-degron and C-degron pathways.**

- A Schematic diagram of the MIC19 protein. MIC19 contains an N-terminal myristoylation motif followed by a DUF37 domain (domain of unknown function) and a CHCH (coiled-coil helix-coiled-coil helix) domain and is terminated by a diGly motif. Myristic acid was co-translationally attached to the N-terminal Gly<sup>2</sup> residue by N-myristoyltransferases (NMTs) after removal of the initiating Met by methionine aminopeptidases. The N-terminal sequence of MIC19 is indicated at bottom with the myristoylation site, Gly<sup>2</sup>, and residues crucial for myristoylation labeled turquoise or underlined, respectively.
- B Live-cell images of U2OS cells stably expressing GFP or GFP N-terminally tagged with the wild-type or mutant N-terminal motif of MIC19 (a.a. 1–24). Scale bar = 20  $\mu$ m.
- C Fractionation and Western blot analysis of endogenous MIC19 in U2OS cells with or without DNKLHDC2, DNCu2, or IMP-1088 treatments. IMP-1088 inhibits NMTs. C and M denote cytosol and mitochondrial fractions, respectively. Tubulin and VDAC serve as fractionation quality controls. The relative abundance of cytosolic MIC19/tubulin was normalized to that of lane 1 and is indicated under each lane in red. Another replicate of this experiment is shown in Fig EV3B.
- D Abundance comparison of GFP N- or C-terminal tagged with the wild-type or mutant N-terminal (a.a. 1–24) or C-terminal (a.a. 218–227) motif of MIC19 in cells with indicated treatments. Data from sh#1 is shown here.
- E Live-cell images of U2OS cells expressing indicated GFP fusion proteins with or without DNCu2 and IMP-1088 treatments. The complete set of images containing degron mutant controls is shown in Appendix Fig S3. Scale bar = 20  $\mu$ m.
- F Representative confocal immunofluorescence images of HA-tagged MIC19 variants with MitoTracker staining in U2OS cells. The fluorescence intensities along the dashed lines are shown as line profile graphs on the right. Scale bar = 20  $\mu$ m.

Source data are available online for this figure.

We applied two approaches to simultaneously appraise the N-degron and C-degron of MIC19. Firstly, we exploited the S<sup>6</sup>W,T<sup>7</sup>P MIC19 mutation that suppresses myristoylation without compromising the activity of its N-degron (Fig 5B and D). In the S<sup>6</sup>W,T<sup>7</sup>P background, MIC19 with a functional N-degron and C-degron remained predominantly concentrated in mitochondria (Fig 6A). Defective N- or C-degrons alone only modestly diminished MIC19 localization, though the effect was more profound for C-degron disruption. Importantly, we observed appreciably more misplaced MIC19 when both degrons were disrupted, as evidenced by both imaging analysis (Fig 6A) and fractionation assays (Fig EV4B right). Secondly, we impaired the N-degron or C-degron pathways *in trans* by inhibiting ZYG11B or KLHDC2, respectively, and we reached the same conclusion. Significantly more cytosolic MIC19 appeared when both BC-box proteins were suppressed in IMP-1088-treated cells (Figs 6B and EV4C). Importantly, IMP-1088 treatment caused fragmented mitochondria due to insufficient mitochondrial MIC19 and this mitochondrial abnormality was recovered by inhibiting N-degron and C-degron pathways-mediated proteolysis (Figs 6C and EV4C). Overall, these data indicate that the two end degron pathways act cooperatively to safeguard proper localization of MIC19 (Fig 6D).

### The diGly/C-degron<sup>KLHDC2</sup> pathway eliminates cleavage products of deubiquitinating enzymes

Apart from diGly-ending full-length proteins, we also identified KLHDC2-binding proteins that do not contain a connate diGly C-terminal (Table 1, Table EV4). We previously found that the N-terminal domain (NTD) of USP1 generated by autocleavage is removed by KLHDC2 (Lin *et al*, 2018). Comparably to USP1, the FAU, SDE2, and TUG proteins contain an ubiquitin-like (UBL) domain with the characteristic C-terminal diGly motif and all these proteins are processed by deubiquitinating enzymes (Dubs) (Olvera & Wool, 1993; Habtemichael *et al*, 2018; Thakran *et al*, 2018). Dubs-mediated cleavage results in their UBL-harboring NTD fragments ending with a diGly tail (Fig 7A; Olvera & Wool, 1993; Jo *et al*, 2016; Habtemichael *et al*, 2018). Therefore, we tested the hypothesis that the diGly/C-degron<sup>KLHDC2</sup> pathway eliminates Dubs cleavage products.

We detected both full-length and NTDs of SDE2 and FAU when HEK293T cells expressed wild-type but not a diGly mutant (GA) version of full-length proteins, suggesting that SDE2 and FAU were

indeed cleaved at the diGly motif by endogenous Dubs (Fig 7B). TUG-NTD was not observed because TUG is processed by a muscle-specific isoform of USP25 (isoform M) (Habtemichael *et al*, 2018). In support of our hypothesis, NTDs but not the full-length forms of all three proteins are KLHDC2 substrates (Figs 7C and EV5A). Mutating the internal diGly of full-length proteins or masking the terminal diGly of NTDs completely abrogated KLHDC2-dependent degradation (Figs 7D and EV5B). Consistently, we observed that wild-type KLHDC2, but not its mutants defective in diGly-peptide binding (R<sup>236</sup>E; R<sup>241</sup>A) (Rusnac *et al*, 2018), physically associated with the NTDs of SDE2, FAU, and TUG but not the respective full-length proteins (Fig 7E). Collectively, these data suggest that the C-degron<sup>KLHDC2</sup> pathway can eliminate diGly-ending cleavage products of Dubs or other proteases.

### Coronaviruses resist proteolytic attack from host C-degron pathways by editing the surrounding sequence of diGly/C-degrons

Coronavirus non-structural proteins (NSPs) are synthesized as long polypeptide chains that are subsequently proteolytically processed into individual NSP proteins by the viral proteases PLpro/NSP3 and 3CLpro. PLpro cleaves at the NSP1-2, 2-3 and 3-4 junctions following a diGly motif (Fig 8A) (Harcourt *et al*, 2004; Báez-Santos *et al*, 2015). We wondered if PLpro-processed NSP proteins are targeted by KLHDC2. Intriguingly, unlike human proteins cleaved by Dubs, diGly-terminating NSP1 and NSP2 from SARS-CoV (severe acute respiratory syndrome coronavirus) escaped the diGly/C-degron pathway (Fig 8B). We characterized an inactivated NSP1 mutant (R<sup>124</sup>A, K<sup>125</sup>A) because wild-type NSP1 expression is lethal to cells (Lokugamage *et al*, 2012; Tanaka *et al*, 2012).

To explain their immunity to the C-degron pathway, we tested if the C-terminal tails of NSP proteins include autonomous KLHDC2 degrons. In contrast to the 12-residue C-terminal motif of SDE2, FAU, or TUG, the C-termini of coronavirus NSP proteins did not or only mildly promoted KLHDC2-mediated degradation (Fig 8C), suggesting that the sequences surrounding coronavirus terminal diGly inactivate the KLHDC2 degrons. Consistent with that notion, the stability of GFP-random peptide libraries ending with the final six residues of SARS-CoV NSP1 was much greater than for those ending with only diGly, and that difference in stability was

completely abolished when KLHDC2 function was blocked (Fig 8D). We found that KLHDC2 degron activity against SARS-CoV NSPs could be drastically improved by changing as little as a single residue at the -3, -4, or -5 positions to Ala (Fig 8C). Collectively,

our data demonstrate differential regulation of cellular and viral diGly-terminating polypeptides by the diGly/C-degron pathway and suggest that coronaviruses have adapted to resist proteolytic attack by host cells.

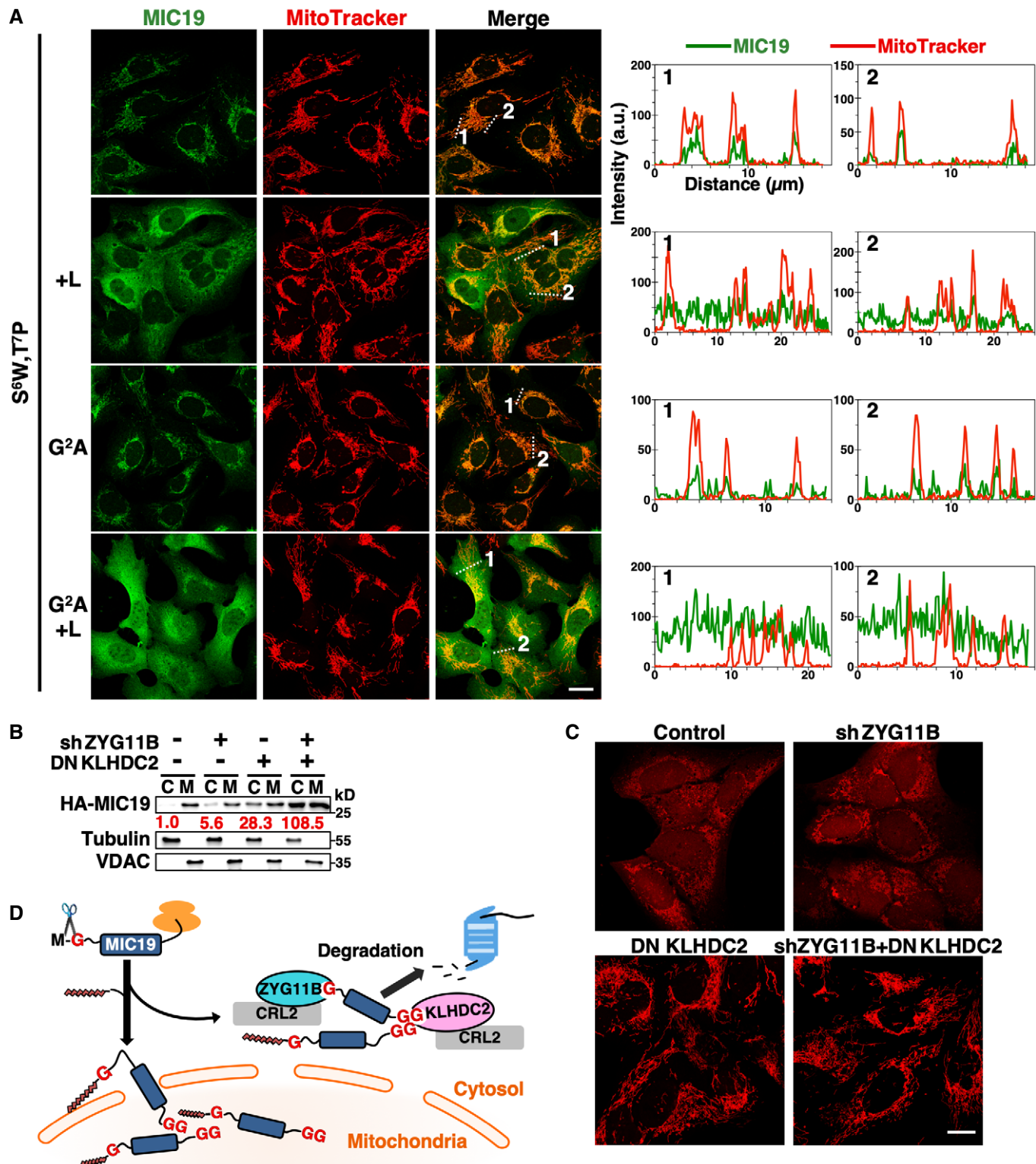
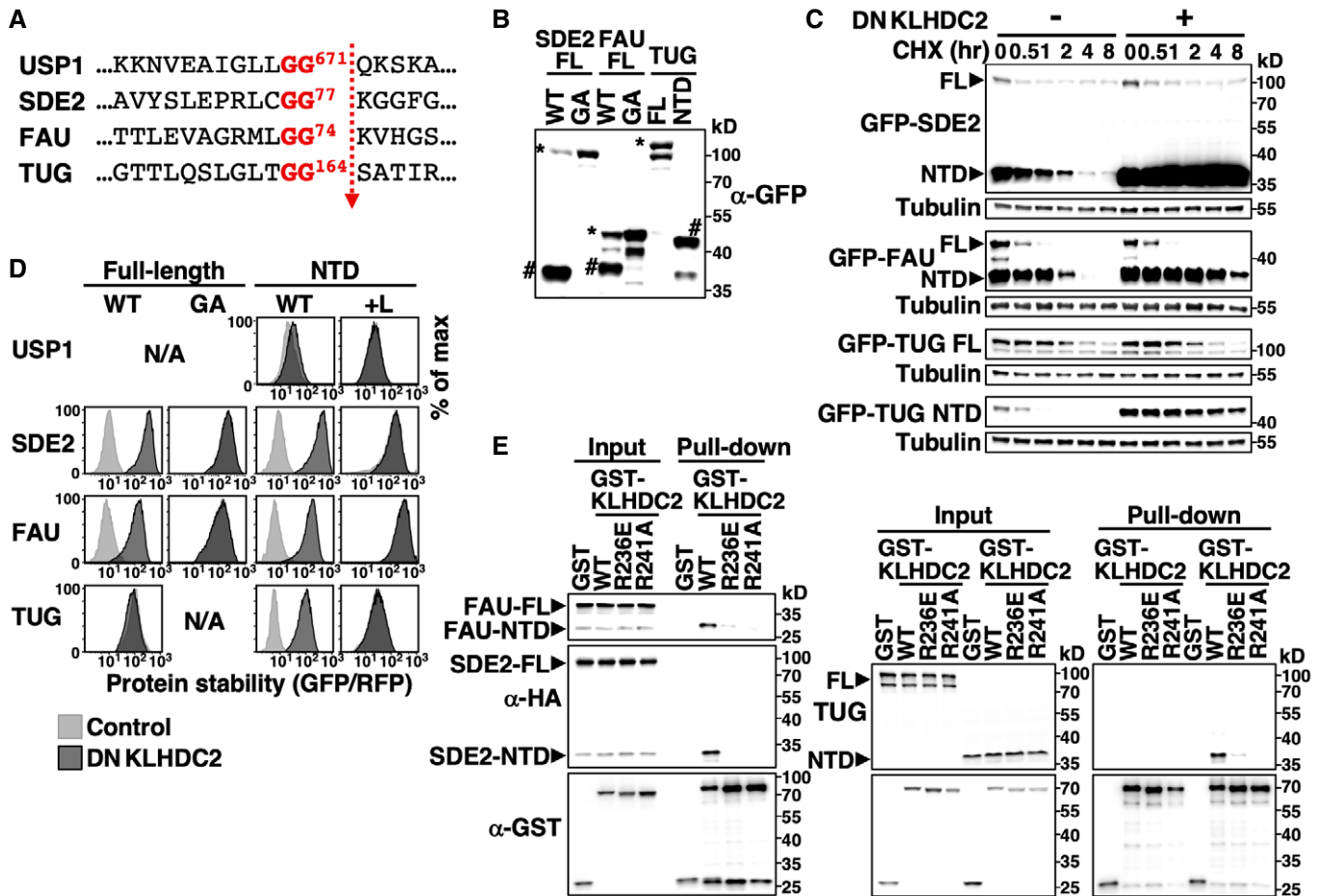


Figure 6.

**Figure 6. Mitochondrial localization of MIC19 is ensured by dual-end inspections.**

- A Representative confocal immunofluorescence images of HA-tagged MIC19 variants with MitoTracker staining in U2OS cells. All MIC19 variants examined carried the S<sup>6</sup>W,T<sup>7</sup>P mutation. The fluorescence intensities along the dashed lines are shown as line profile graphs on the right. Scale bar = 20  $\mu$ m.
- B Fractionation analysis of wild-type HA-MIC19 in U2OS cells treated with IMP-1088 and with or without inhibition of ZYG11B or KLHDC2. The normalized relative abundance of cytosolic HA-MIC19/tubulin is indicated under each lane in red.
- C MitoTracker staining of IMP-1088-treated U2OS cells with or without inhibition of ZYG11B or KLHDC2. The complete set of images containing MIC19 localizations is shown in Fig EV4C. Scale bar = 20  $\mu$ m.
- D Proposed model. CRL2-operated “double-end” inspections safeguard the mitochondrial localization of MIC19.
- Source data are available online for this figure.

**Figure 7. The diGly/C-degron pathway eliminates cleavage products of deubiquitinating enzymes.**

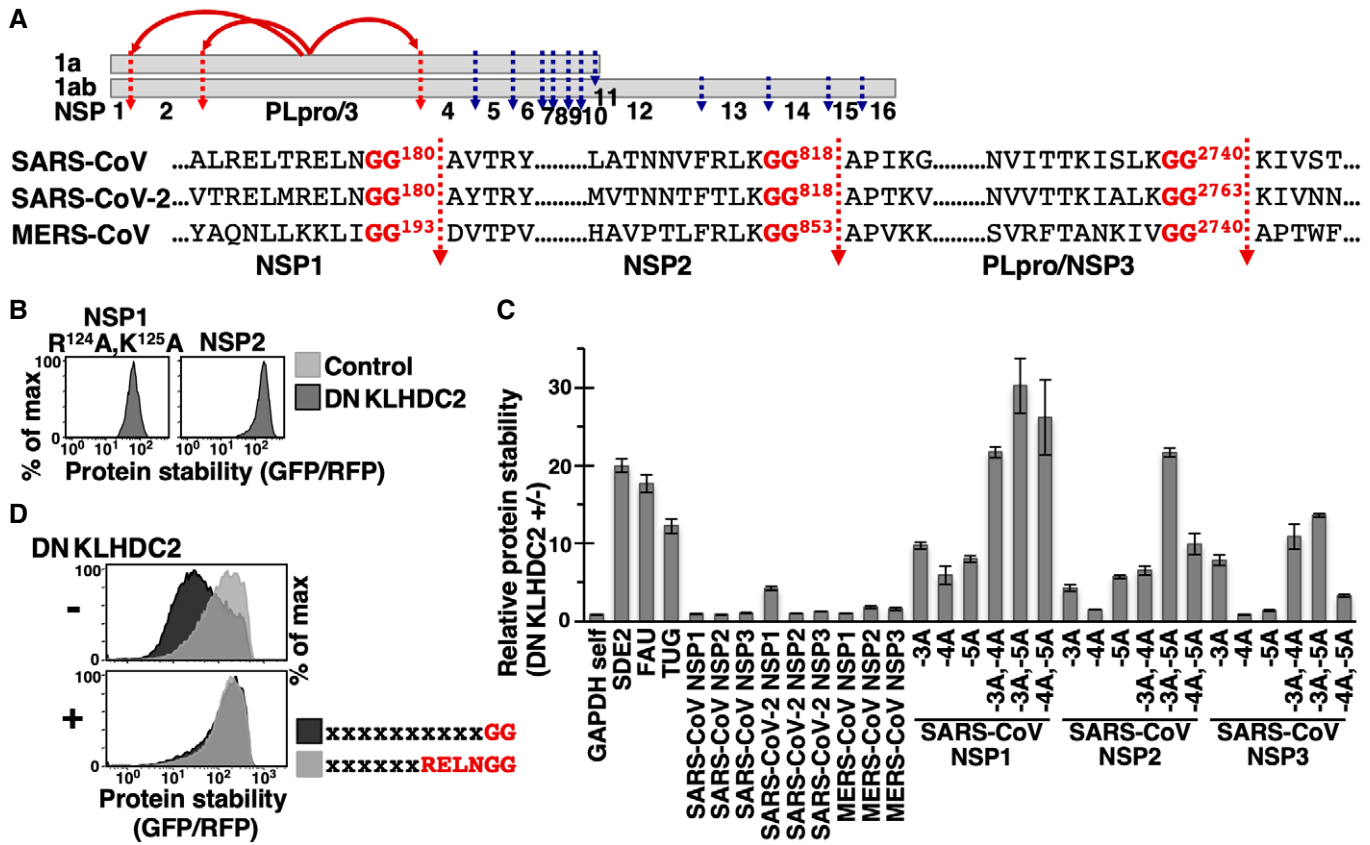
- A The cleavage sequence of indicated proteins.
- B Western blot analysis of HEK293T cells expressing GFP-tagged wild-type (WT) or an uncleavable mutant (diGly to GA) full-length (FL) version or the N-terminal domain (NTD) of indicated proteins. Full-length and NTD proteins are marked by \* or # on the blot, respectively.
- C Cycloheximide (CHX)-chase analysis of GFP-tagged FL and NTD versions of indicated proteins.
- D GPS analysis of wild-type and mutant forms of FL or the NTD of indicated proteins with or without dominant-negative KLHDC2 (DNKLHDC2) treatment. A C-terminal Leu was added (+L) to NTDs to cap the diGly degron.
- E GST pull-down assay using cells expressing GST or GST-tagged wild-type or mutant KLHDC2 and HA-tagged SDE2, FAU or TUG.
- Source data are available online for this figure.

## Discussion

Herein, we report physiological functions of C-degron pathways and demonstrate the importance of protein “ends” in proteolysis-

mediated protein quality inspection. Furthermore, we establish links between the functional proteome and C-degron pathways.

One selective pressure sculpting proteome evolution is the requirement that proteins have sufficient stability to perform their



**Figure 8. Coronaviral diGly-ending polypeptides resist C-degron pathway-mediated proteolysis.**

**A** Schematic depiction of the proteolytic process of coronavirus replicase protein 1a and 1ab by proteases PLpro/NSP3 (labeled red) and 3CLpro (labeled blue). PLpro/NSP3-mediated cleavage generates NSP1, NSP2, and PLpro/NSP3 proteins ending in the diGly motif. The cleavage sequences of indicated viruses are shown.

**B** Stability analysis of FL SARS-CoV NSP1 and NSP2 proteins with or without DNKLHDC2 treatment. We were unable to analyze NSP3 due to its large size (214 kDa).

**C** Relative stability analysis of GAPDH C-terminally tagged with the last 12 residues of indicated proteins with or without DNKLHDC2 treatment by GPS assays. The -3, -4, or -5 residue of SARS-CoV NSP proteins was mutated to Ala (A) as indicated. The experiment was done in triplicate using three independently prepared viruses to express DNKLHDC2. Data are presented as the mean value of GFP/RFP +/- standard deviation from triplicate experiments.

**D** Global protein stability profiling of GPS reporter cells expressing random peptide libraries ending with a diGly motif or the last six residues of SARS-CoV NSP1 protein with or without DNKLHDC2 treatment.

biological functions. Our results indicate that functional proteomes have evolved multiple tactics to cope with the threat of C-degron pathway-mediated proteolysis. Our proteome sequence analyses reveal striking Gly/C-degron scarcity in functional eukaryotic proteomes but not in defective proteins produced from translation errors. Only ~ 50 full-length proteins terminating with diGly occur in the human proteome. It is unclear why terminal Gly is depleted in the yeast proteome given that a CRL2 homolog does not exist in yeast. In line with our expectations, the shortfall in terminal Gly is restricted to intracellular proteins accessible to C-degron pathways. Importantly, mutating the extreme C-terminal residue of human proteins to Gly mostly triggered degradation, supporting a direct causal relationship between Gly ends and protein instability. In addition to excluding Gly/C-degrons, proteins may possess particular folding properties to shield latent C-degrons, such as those we have reported previously for ubiquitin and SUMO2 (Lin *et al.*, 2018). Viral proteins are typically subjected to host protein degradation machineries. Interestingly, our data suggest that the sequences near to the diGly motifs of coronaviral diGly-ending proteins prevent

KLHDC2 attack. These results shed light on the protein quality control systems that support functional proteomes.

Inspired by the significant annotation of secreted and mitochondrial proteins harboring C-degrons and identification of KLHDC2-binding proteins localized at mitochondria or ER membrane, we have revealed the function of C-degron pathways in protein spatial quality control. The targeting signal for protein transport into a given organelle is typically specified by general biophysical properties rather than particular amino acid sequences. For instance, the signal peptide comprises a stretch of primarily hydrophobic helix-compatible residues, and the MTS forms an amphiphilic  $\alpha$ -helix with an alternating pattern of hydrophobic and positively charged amino acids (von Heijne, 1990). Due to their miscellaneous nature, the efficacy of targeting signals in different proteins varies substantially, with overall failure rates of intracellular protein segregation potentially being as high as a few percent (Kim *et al.*, 2002). We have identified a panel of secreted and mitochondrial proteins that are eliminated by C-degron pathways upon mislocalization. The C-termini of aberrant proteins are expected to be mainly stochastic.



24 h. All exogenous proteins and shRNAs were expressed *via* lentivirus-mediated transduction. Constructs and DNAs used in this study are listed in Table EV5.

### Transfection, infection, and lentivirus production

Transfection of HEK293T cells was executed using TransIT-293 transfection reagent (Mirus Bio) according to the manufacturer's instructions. Virus infection was conducted in medium containing 8 µg/ml polybrene (Sigma-Aldrich). To prepare lentiviruses, HEK293T cells at approximately 80% confluency were transfected with pRev, pTat, pHIV gag/pol, pVSVG, and the lentiviral construct of interest. Medium was freshly replaced and viruses were harvested 24 and 48 h post-transfection, respectively.

### Generation of GPS reporter cell lines and GPS assay

To generate GPS reporter constructs, the gene of interest was cloned into the pLenti-GPS vector by Gateway recombination (Invitrogen). To generate GPS reporter cell lines, cells were infected with lentiviruses carrying GPS reporter constructs at a low MOI of ~0.2 and then selected by puromycin (1 µg/ml, Clontech Laboratories). GPS reporter cells were analyzed using a BD LSR Fortessa system (BD Biosciences) operated by BD FACS-Diva™ software. The 488-nm and 561-nm lasers were used to stimulate GFP and RFP, respectively. The flow rate was  $5 \times 10^3 \sim 1 \times 10^4$  cells/s. No background correction or compensation was applied. Cells were gated by doublet discrimination and as RFP-positive in order to only measure the fluorescent signals of individual GPS reporter cells. We typically analyzed 20,000 or 100,000 individual cells for reporter cells carrying a single GPS construct or GPS peptide libraries, respectively. FlowJo software was used for FACS analysis.

### Context-independent GPS random peptide platform

Synthesized oligonucleotides (Life Technologies) with 5' and 3' linkers encoding 12-residue random peptides were amplified by PCR and cloned into pDONR by Gibson assembly, followed by transfer into the pLenti-GPS reporter *via* Gateway recombination to generate GPS random peptide libraries. NNK degenerate codons (N = A/T/G/C; K = G/T) instead of NNN were used to encode random residues in order to reduce amino acid bias and to decrease the incidence of stop codons from 3/64 to 1/32. To avoid uneven representation of libraries due to biases in synthesis of oligonucleotides, two or three independent libraries were constructed for each examined C-terminal residue or motif generated from separately synthesized oligonucleotides. The complexity of each random peptide library is approximately  $10^{10}$ .

GPS random peptide library cells were generated by infecting HEK293T cells with lentiviruses carrying GPS random peptide libraries at a low MOI of ~0.1 to ensure that each cell only expresses one unique GFP-peptide fusion. The percentage of GPS library cells with reduced GFP/RFP ratios, as determined by FACS (and represented as “% degradation”), was used as an index to reflect the ability of peptide mixtures to promote protein degradation. It is unfeasible to analyze cells carrying all ~  $10^{10}$  random peptides for each library. To ensure the reliability of our results, we

performed FACS analyses in triplicate and recorded 100,000 cells each time. Some peptides encoded by random oligonucleotides have fewer than 12 residues due to random insertions of internal TAG stop codons. Thus, the % degradation revealed by FACS ( $D_{\text{obs}}$ ) represents both full-length and undesired truncated peptides. We approximated  $D_{\text{obs}}$  as a linear combination of  $D_{\text{real}}$  and  $D_A$ , i.e., the real % degradation of 12-residue “full-length” peptides terminated with specified C-terminal residues and the observed % degradation from GPS library cells carrying the A-residue random peptide library, respectively

$$D_{\text{obs}} = \left[ \left( \frac{31}{32} \right)^{12-A} * D_{\text{real}} \right] + \left[ \left( 1 - \left( \frac{31}{32} \right)^{12-A} \right) * D_A \right].$$

where “A” denotes the number of random residues coded by NNK codons within the 12 residues.

### Cycloheximide-chase assay

Cells were treated with 100 µg/ml cycloheximide (Calbiochem), followed by sample collection at multiple time-points. Samples were collected by directly lysing cells in 2× Laemmli sample buffer followed by sonication and boiling for 5 min. Protein abundance was analyzed by immunoblotting.

### GST pull-down assay, cell fractionation, and immunoblotting

To detect physical interactions between putative substrates and BC-box proteins, HEK293T cells stably expressing HA-tagged substrates were transfected with plasmids carrying GST-tagged BC-box proteins for 42 h followed by treatment with 1 µM bortezomib and 1 µM MLN4924 for 6 h to block protein degradation. Cells were then lysed on ice using IP-lysis buffer (20 mM Tris pH 7.6, 150 mM NaCl and 0.5% IGEPAL CA-630) supplemented with protease inhibitor cocktails (Roche). Cell extracts were incubated with Glutathione-Sepharose 4B (GE Healthcare Life Sciences) for 16 h at 4°C. The corresponding Sepharose was washed three times in IP-lysis buffer, boiled in 2× Laemmli sample buffer, and subjected to immunoblotting.

Cell fractionation was performed by using a cell fractionation kit (Abcam, ab109719) according to the manufacturer's instructions. In brief, cells were resuspended in buffer A and permeabilized with detergent I. The cell suspension was then spun at 10,000 g for 2 min. The resulting supernatant was collected as the cytosol fraction. The cytosol-depleted pellet was resuspended in buffer A and solubilized with detergent II. Following centrifugation at 10,000 g for 2 min, the supernatant was collected as the mitochondria-enriched fraction. Protein abundance in each fraction was analyzed by immunoblotting. VDAC1 and α-tubulin served as organelle-specific markers for mitochondria and cytosol, respectively.

For immunoblotting, primary antibodies were purchased from the following suppliers: Flag (Sigma-Aldrich, M2), GFP (Takara Bio Clontech, JL-8, 632381), GST (GE Healthcare, 27457701), HA (Abcam, ab130275), α-tubulin (Thermo Scientific™, MS-581-P), GAPDH (GeneTex, 100118), PDP2 (GeneTex, 44905), MRPL28 (GeneTex, 115439), VDAC1 (Abcam, ab14734), and MIC19 (Abcam, ab224565).



## Cell staining

To detect ectopically-expressed MIC19, a HA tag was inserted directly after the myristoylation motif (a.a. 1–14) of MIC19. We have confirmed that the HA tag does not affect mitochondrial targeting of MIC19. U2OS cells stably expressing HA-MIC19 were grown on coverslips coated with poly-L-lysine (Sigma-Aldrich). Mitochondrial staining was performed by incubating cells with growth medium containing 25 nM of MitoTracker probe (Thermo Fisher) for 45 min at 37°C. After rinses with PBS, cells were fixed in 4% paraformaldehyde for 15 min at room temperature, followed by three washes in 0.2% Triton X-100/PBS. The coverslips were blocked with blocking buffer (2% BSA in 0.2% Triton X-100/PBS) for 1 h and then incubated overnight with a MIC19 primary antibody (Abcam, ab224565) (1:100 dilution in blocking buffer) at 4°C. Following three washes in 0.2% Triton X-100/PBS, cells were incubated with Alexa Fluor-conjugated secondary antibody (Thermo Fisher, A27034) (1:100 dilution in blocking buffer) for 1 h at room temperature in the dark. The coverslips were mounted on glass slides using Fluoroshield DAPI mounting medium (Sigma-Aldrich). Images were captured using a scanning laser confocal microscope (ZEISS LSM710). For quantification analysis, all images were acquired using the same device parameters.

## Luciferase assay

Luciferase activities were measured in triplicate using a Pierce™ Firefly Luciferase Glow Assay kit (Thermo Fisher) and BioLux® *Gaussia* Luciferase Assay kit (NEB) according to the manufacturers' instructions. To measure the activity of secreted luciferases, the conditioned media of HEK293T cells stably expressing FLuc or GLuc were collected. To measure cytosolic luciferase activities, HEK293T cells expressing FLuc or GLu were lysed to prepare cell lysates. Chemiluminescence signals were detected using a SpectraMax Paradigm Multi-Mode Microplate Reader (Molecular Devices).

## Affinity purification-mass spectrometry

For mass spectrometry (MS)-based identification of KLHDC2 substrates, HEK293T cells stably expressing GST-tagged DNKLHDC2 (a.a. 1–362) were treated with 1 μM bortezomib and 1 μM MLN4924 for 6 h to block protein degradation before cell harvesting. Cells were lysed on ice in IP-lysis buffer (20 mM Tris pH 7.6, 150 mM NaCl and 0.5% IGEPAL CA-630) supplemented with protease inhibitors (Roche). Cell lysates were spun at 16,000 g and 4°C for 15 min to remove cell debris. Cell compartmentalization was disrupted under this lysis condition. The clarified lysate was then incubated with Glutathione-Sepharose 4B (GE Healthcare Life Science) for 16 h at 4°C. The corresponding Sepharose was washed five times with washing buffer (20 mM Tris pH 7.6, 300 mM NaCl, 1% IGEPAL CA-6300, and 0.5% sodium deoxycholate). KLHDC2-binding proteins were eluted with 2% SDS, reduced with DTT, and alkylated with iodoacetamide, followed by trypsin digestion for 16 h at 37°C. The tryptic peptides were cleaned up using C18 zip-tips and subjected to MS analysis.

MS data were acquired on an Orbitrap Fusion mass spectrometer (Thermo Scientific) equipped with an EASY-nLC 1200 system

(Thermo Scientific) and a nanoelectrospray ion source (New Objective, Inc.). Samples dissolved in 0.1% formic acid were injected onto a self-packed precolumn (150 μm I.D. × 30 mm, 5 μm, 200 Å) at a flow rate of 10 μl/min. Chromatographic separation was performed on a self-packed reverse-phase C18 nanocolumn (75 μm I.D. × 200 mm, 3 μm, 100 Å) using 0.1% formic acid in water as mobile phase A and 0.1% formic acid in 80% acetonitrile as mobile phase B, operated at a flow rate of 300 nl/min. The LC gradient was applied from 2% buffer B at 2 min to 40% buffer B at 40 min. Electrospray voltage was maintained at 1.8 kV and the capillary temperature was set at 275°C. Full MS survey scans were executed in the mass range of  $m/z$  320 to 1,600 (AGC target at  $5 \times 10^5$ ) with lock mass, resolution of 120,000 (at  $m/z$  200), and a maximum injection time of 50 ms. The MS/MS were run in top speed mode with 3 s cycles, and the dynamic exclusion duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. The precursor ion isolation was performed with a mass selecting quadrupole, and the isolation window was set to  $m/z$  2.0. Monoisotopic precursor ion selection was enabled and 1+ charge-state ions were rejected for MS/MS. The MS/MS analyses were carried out with the collision induced dissociation (CID) mode and with a normalization collision energy (NCE) of 30%. The maximum injection time for spectral acquisition was 150 ms and the automatic gain control (AGC) target values for MS/MS scans were set at  $5 \times 10^4$ .

For data analysis, all MS/MS spectra were converted to mgf format from an experiment RAW file using msConvert (version 3.0.18165, ProteoWizard), and then analyzed using Mascot for MS/MS ion search. The search parameters included an error tolerance for precursor ions and the MS/MS fragment ions in spectra of 10 ppm and 0.6 Da, respectively. The enzyme cutting site was customized at the C-terminal of lysine, arginine, and glycine, with missed cleavages numbering two. The variable post-translational modifications in the search parameters were assigned to include oxidation of methionine and carbamidomethylation of cysteine. Peptides were identified with the false discovery rate (FDR) < 1%.

## Targeting sequences for shRNAs

APPBP2: GCCTTCAGTTGTGTACTCT  
 FEM1C: GTAACAGTTGTTTCATAA  
 KLHDC2: CTTGGTGTCTGGGTATATA  
 KLHDC3: TGGAAAAAGATTGAACCGA  
 VHL: TGGCTCAACTTCGACGGCG  
 ZER1: ACAAAGTAAGTGGTGACAA (#1); AACTGCGAGATG  
 TTCCTCA (#2); TCCAGCTACTACAAGTTCA (#3); GGGAT  
 CGAGGTTTCCTACA (#4)  
 ZYG11B: AGCAAGAATCCTTCAAGGT (#1); CTGTTGACC  
 TGATTTCTGA (#2);  
 AGACCGACTCAAGTCTCTA (#3)

## C-degrons used for targeting model substrate degradation

KLHDC2 diGly degron: LRGPSPPPMAGG  
 KLHDC3 R\_\_G degron: ASKERARPCQRG  
 APPBP2 R\_G\_ degron: EMALSPPRSWQQ

### Mitochondria targeting sequences used for targeting GFP to mitochondria

MTS<sup>PDHA1</sup>: MRKMLAAVSRVLSGASQKPASRVLVASRNF  
 MTS<sup>COX8A</sup>: MSVLTPLLLRGLTGSARRLPVPRAK  
 MTS<sup>GLUL</sup>: MTTASSSHNLKGIKQVYMSLPQGEKVQAMYIWI  
 DGTGEGLRCKTRTLTLDSE

### Bioinformatics analysis of anomalous amino acid distribution at protein C-termini

#### Collection and processing of protein sequences

Human protein coding sequences were downloaded from NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>, download date: 18 Feb. 2017) and (<https://www.uniprot.org/>, download date: 05 Aug. 2020). The protein coding sequences of *Mus musculus* (GRCm38.p5), *Drosophila melanogaster* (BDGP6), *Caenorhabditis elegans* (WBcel235), *Arabidopsis thaliana* (Araport11\_genes.201606), *Saccharomyces cerevisiae* (R64-1-1), and *Escherichia coli* (MG1655) were obtained from the Ensembl BioMart data mining tool (<https://www.ensembl.org/biomart/martview/>), the TAIR Arabidopsis Resource (<https://www.arabidopsis.org/>) or NCBI GenBank. Redundant sequences and sequences shorter than 50 amino acids were removed. The total numbers of protein sequences analyzed in human (NCBI), human (Uniprot), mouse, fly, nematode, plant, yeast, and *E. coli* are 36,644, 20,221, 41,504, 21,763, 27,644, 40,366, 6,484, and 4,028, respectively.

Protein sequences produced by NMD were collected from BioMart with the following criteria: gene biotype = protein coding and transcript type = nonsense-mediated decay (GRCh38.p12, release 92). Protein sequences from uORFs were decoded from the TISdb database (<http://tisdb.human.cornell.edu/>) (Wan & Qian, 2014). Since protein products translated from uORFs are relatively small, we kept unique sequences that are equal to or longer than 10 amino acids. The total numbers of NMD and uORF proteins/peptides analyzed are 11,225, and 3,901, respectively.

To prepare abnormal protein sequences, human transcript sequences were downloaded from Ensembl (download date: 3 Jan. 2017) and non-protein transcripts were removed. We considered two types of abnormal protein sequences derived from the normal mRNA transcripts. Protein translation may undergo frameshifts when the reading frame is shifted at the +1 or -1 position, so the downstream amino acid sequences are completely altered. In addition, protein translation may also undergo readthrough error when ribosomes skip the first stop codon they encounter and instead continue peptide elongation until they reach the next stop codon. The procedures for generating these two types of aberrant protein sequences are described in detail below:

#### Frameshifted protein sequences

We retained 36,284 mRNA transcripts satisfying the following criteria as templates for generating frameshifted protein sequences: (1) they possessed RefSeq IDs according to the NCBI database; (2) they encoded proteins, as the RefSeq IDs were converted to Ensembl protein IDs according to the file “gene2ensembl” from NCBI, <ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene2ensembl>; (3) the encoded proteins had  $\geq 50$  amino acids. For each template transcript, we generated 1,000 frameshifted protein sequences by (1) randomly

selecting the positions of the initially disrupted reading frames; (2) randomly selecting the phases of frameshifts (+1 or -1); and (3) discarding sequences shorter than 50 amino acids. One representative was randomly drawn from among the 1,000 frameshifted protein sequences for each transcript. This sampling scheme is preferable to exhaustively generating the frameshifted sequences starting from all positions.

#### Readthrough protein sequences

We selected 39,216 transcripts as templates according to the following criteria: (1) they encode proteins; (2) they contain 3'UTR regions; (3) they have stop codons in their 3'UTR regions; (4) the translated proteins are longer than 50 amino acids; and (5) the translated amino acid sequences are unique. The native stop codons of these transcript templates were skipped, resulting in generation of longer peptide sequences terminating at the first downstream stop codon in the 3'UTR region.

Protein sequences of human transmembrane proteins with different C-terminal tail topologies were downloaded from the Human Transmembrane Proteome database (<http://htp.enzim.hu/>, download date: 7 Mar. 2018). The numbers of sequences corresponding to “C-terminal inside of membrane” and “C-terminal outside of membrane” are 4,052, and 1,360, respectively. Sequences of human secreted proteins were collected from VerSeDa (the vertebrate secretome database) (<http://genomics.cicbiogune.es/VerSeDa/index.php>, download date: 29 Jan. 2018) (Cortazar *et al*, 2017), with the following criteria: “only” extracellular and SignalP  $\geq 0.5$  in order to exclude proteins with multiple subcellular localizations or that are secreted *via* unconventional pathways. Cytosolic, nuclear, and mitochondrial proteins were collected based on the Human Protein Atlas (<https://www.proteinatlas.org/>) and MitoProteome (<http://www.mitoproteome.org/>). Cytosolic proteins do not include those located at endoplasmic reticulum, Golgi, endosomes, lysosomes, peroxisomes, mitochondria, vesicles, or lipid droplets. Proteins with multiple localizations or shorter than 50 amino acids were excluded. The total numbers of sequences from cytosolic, nuclear, mitochondrial, and secreted proteins analyzed are 7,197, 5,983, 3,363, and 2,218, respectively.

#### Statistical assessment of the prominence and significance of amino acid bias at the C-termini of proteins

We extracted the last 50 amino acid sequences of proteins and counted the occurrence frequency of each amino acid residue at each position. Two statistical assessments were performed to quantify the disparity of the amino acid frequencies at the terminal position relative to those of all other positions. First, the occurrence frequencies of amino acid residues at each position were treated as a 20-dimensional vector, and principal component analysis (PCA) was applied to project those vectors onto a two-dimensional space. Second, the *P*-value whereby each amino acid residue was under- or over-represented at the C-terminal compared to a background distribution derived from the protein data over multiple positions was computed. Here, the background (null) model is the multinomial distribution  $F = (F_1, \dots, F_{20})$ , where each  $F_i$  represents the relative occurrence frequency of amino acid residue  $i$  according to the background distribution. Suppose among  $n$  proteins that amino acid residue  $i$  appears at the terminal position in  $n_i$  proteins. Then,  $n_i$  follows a binomial distribution and the probability of under-

representation is the left-tail probability according to the cumulative distribution function (CDF) of  $n_i$ :

$$\Pr(n_i \leq \hat{n}_i) = \sum_{k=0}^{\hat{n}_i} \binom{n}{k} F_i^k (1 - F_i)^{n-k}$$

where  $\hat{n}_i$  denotes the number of proteins where amino acid residue  $i$  appears at the terminal position from the empirical data. The probability of over-representation is the right-tail probability of the same distribution.  $\Pr(n_i \leq \hat{n}_i)$  can be approximated by the CDF function of a Poisson distribution:

$$\Pr(n_i \leq \hat{n}_i) = \sum_{k=0}^{\hat{n}_i} \frac{e^{-\lambda_i} \lambda_i^k}{k!}$$

where  $\lambda_i = nF_i$ .

We computed the  $P$ -values of under- and over-representations for each amino acid at the C-terminal for the following classes of proteins or peptide sequences: (1) functional full-length proteins; (2) functional proteins with specific localizations; (3) frame-shifted protein sequences; (4) stop codon readthrough protein sequences; (5) NMD protein sequences; and (6) peptide products from uORFs. The  $P$ -values of background amino acid frequencies of classes (1)–(5) were derived from all positions of all functional full-length proteins. The  $P$ -values of background amino acid frequencies of class (6) were derived from all positions of the uORF peptides alone due to the differential amino acid usages between functional proteins and uORF peptides. FDR (false discovery rate)-adjusted  $P$ -values were calculated for multi-hypotheses test correction. For the  $P$ -values pertaining to the amino acid frequencies of the last residues of proteins (e.g., Table EV2), there were 20 hypotheses corresponding to all amino acids. For the  $P$ -values pertaining to the amino acid frequencies of the last 50 residues of proteins (e.g., Fig 2), there were  $50 \times 20 = 1,000$  hypotheses corresponding to all combinations of amino acids and positions.

## Data availability

This study includes no data deposited in external repositories.

**Expanded View** for this article is available online.

## Acknowledgements

We thank S. Baker and S. Makino for SARS-CoV *nsp* constructs, S.P. Li for confocal image analysis, M.D. Tsai for mass spectrometry analysis, the Data Science Statistical Cooperation Center of Academia Sinica (AS-IA-108-L02) for statistical support, and H. Cheng, K.L. Hsu, C.L. Lin, S.C. Lin, S.T. Lin, J. O'Brien, and H.N. Wu for suggestions and assistance. Image analyses were performed at the Imaging Core of the Institute of Molecular Biology at Academia Sinica. Mass spectrometric analyses were supported by the Taiwan Protein Project and performed at the Mass Spectrometry Core Facility at the Genomics Research Center, Academia Sinica. This work was supported by MOST grants 107-2321-B-001-021 and 108-2321-B-001-006 from the National Science Council of Taiwan, and an Investigator Award from Academia Sinica.

## Author contributions

C-WY characterized C-degrons, human C-degron-harboring proteins and MIC19. W-CH carried out GST pull-down analyses to search for KLHDC2 substrates and characterized cellular and viral diGly-ending polypeptides. P-HH performed mass spectrometry analyses; K-HY, PW-CH, and Y-NC performed bioinformatic analyses; L-CW and S-CC performed mutagenesis analyses and GPS random peptide assays to characterize C-degrons; H-CL developed GPS random peptide assays and characterized C-degron-bearing model substrates; and C-HY supervised bioinformatic and statistical analyses. H-CSY designed experiments, analyzed data, supervised the project, and wrote the paper.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

- Bachmair A, Finley D, Varshavsky A (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* 234: 179–186
- Báez-Santos YM, St. John SE, Mesecar AD (2015) The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res* 115: 21–38
- Bogan JS, Rubin BR, Yu C, Loffler MG, Orme CM, Belman JP, McNally LJ, Hao M, Cresswell JA (2012) Endoproteolytic cleavage of TUG protein regulates GLUT4 glucose transporter translocation. *J Biol Chem* 287: 23932–23947
- Chen B, Retzlaff M, Roos T, Frydman J (2011) Cellular strategies of protein quality control. *Cold Spring Harb Perspect Biol* 3: a004374
- Cortazar AR, Oguiza JA, Aransay AM, Lavin JL (2017) VerSeDa: vertebrate secretome database. *Database (Oxford)* 2017: baw171
- Darshi M, Mendiola VL, Mackey MR, Murphy AN, Koller A, Perkins GA, Ellisman MH, Taylor SS (2011) ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. *J Biol Chem* 286: 2918–2932
- Darshi M, Trinh KN, Murphy AN, Taylor SS (2012) Targeting and import mechanism of coiled-coil helix coiled-coil helix domain-containing protein 3 (ChChd3) into the mitochondrial intermembrane space. *J Biol Chem* 287: 39480–39491
- Deichaite I, Casson LP, Ling HP, Resh MD (1988) *In vitro* synthesis of pp60v-src: myristylation in a cell-free system. *Mol Cell Biol* 8: 4295–4301
- Dubnikov T, Ben-Gedalya T, Cohen E (2017) Protein quality control in health and disease. *Cold Spring Harb Perspect Biol* 9: a023523
- Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K (2015) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics* 14: 1113–1126
- Giglione C, Fioulaine S, Meinel T (2015) N-terminal protein modifications: bringing back into play the ribosome. *Biochimie* 114: 134–146
- Habtemichael EN, Li DT, Alcazar-Roman A, Westergaard XO, Li M, Petersen MC, Li H, DeVries SG, Li E, Julca-Zevallos O et al (2018) Usp25m protease regulates ubiquitin-like processing of TUG proteins to control GLUT4 glucose transporter translocation in adipocytes. *J Biol Chem* 293: 10466–10486
- Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA, Baker SC (2004) Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol* 78: 13600–13612
- Harper JW, Bennett EJ (2016) Proteome complexity and the forces that drive proteome imbalance. *Nature* 537: 328–338
- von Heijne G (1990) The signal peptide. *J Membr Biol* 115: 195–201

- Hessa T, Sharma A, Mariappan M, Eshleman HD, Gutierrez E, Hegde RS (2011) Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature* 475: 394–397
- Hetz C, Mollereau B (2014) Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases. *Nat Rev Neurosci* 15: 233–249
- Hipp MS, Kasturi P, Hartl FU (2019) The proteostasis network and its decline in ageing. *Nat Rev Mol Cell Biol* 20: 421–435
- Hwang CS, Shemorry A, Varshavsky A (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* 327: 973–977
- Jo U, Cai W, Wang J, Kwon Y, D'Andrea AD, Kim H (2016) PCNA-Dependent cleavage and degradation of SDE2 regulates response to replication stress. *PLoS Genet* 12: e1006465
- Kim SJ, Mitra D, Salerno JR, Hegde RS (2002) Signal sequences control gating of the protein translocation channel in a substrate-specific manner. *Dev Cell* 2: 207–217
- Koren I, Timms RT, Kula T, Xu Q, Li MZ, Elledge SJ (2018) The Eukaryotic proteome is shaped by E3 ubiquitin ligases targeting C-terminal degrons. *Cell* 173: 1622–1635.e1614
- Levine CG, Mitra D, Sharma A, Smith CL, Hegde RS (2004) The efficiency of protein compartmentalization into the secretory pathway. *Mol Biol Cell* 16: 279–291
- Lin HC, Ho SC, Chen YY, Khoo KH, Hsu PH, Yen HC (2015) SELENOPROTEINS. CRL2 aids elimination of truncated selenoproteins produced by failed UGA/Sec decoding. *Science* 349: 91–95
- Lin HC, Yeh CW, Chen YF, Lee TT, Hsieh PY, Rusnac DV, Lin SY, Elledge SJ, Zheng N, Yen HS (2018) C-terminal end-directed protein elimination by CRL2 ubiquitin ligases. *Mol Cell* 70: 602–613.e603
- Lokugamage KG, Narayanan K, Huang C, Makino S (2012) Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *J Virol* 86: 13598–13608
- Matthews GD, Gur N, Koopman WJH, Pines O, Vardimon L (2010) Weak mitochondrial targeting sequence determines tissue-specific subcellular localization of glutamine synthetase in liver and brain cells. *J Cell Sci* 123: 351–359
- Meyer H, Bug M, Bremer S (2012) Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat Cell Biol* 14: 117–123
- Mól AR, Castro MS, Fontes W (2018) NetWheels: a web application to create high quality peptide helical wheel and net projections. *bioRxiv* <https://doi.org/10.1101/416347> [PREPRINT]
- Olvera J, Wool IG (1993) The carboxyl extension of a ubiquitin-like protein is rat ribosomal protein S30. *J Biol Chem* 268: 17967–17974
- Petroski MD, Deshaies RJ (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* 6: 9–20
- Pfanner N, Geissler A (2001) Versatility of the mitochondrial protein import machinery. *Nat Rev Mol Cell Biol* 2: 339–349
- Piatkov KI, Colnaghi L, Bekes M, Varshavsky A, Huang TT (2012) The auto-generated fragment of the Usp1 deubiquitylase is a physiological substrate of the N-end rule pathway. *Mol Cell* 48: 926–933
- Piatkov KI, Oh JH, Liu Y, Varshavsky A (2014) Calpain-generated natural protein fragments as short-lived substrates of the N-end rule pathway. *Proc Natl Acad Sci USA* 111: E817–826
- Pohl C, Dikic I (2019) Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science* 366: 818–822
- Posakony JW, England JM, Attardi G (1977) Mitochondrial growth and division during the cell cycle in HeLa cells. *J Cell Biol* 74: 468–491
- Rageul J, Park JJ, Jo U, Weinheimer AS, Vu TTM, Kim H (2019) Conditional degradation of SDE2 by the Arg/N-End rule pathway regulates stress response at replication forks. *Nucleic Acids Res* 47: 3996–4010
- Rane NS, Chakrabarti O, Feigenbaum L, Hegde RS (2010) Signal sequence insufficiency contributes to neurodegeneration caused by transmembrane prion protein. *J Cell Biol* 188: 515–526
- Rusnac D-V, Lin H-C, Canzani D, Tien KX, Hinds TR, Tsue AF, Bush MF, Yen H-C, Zheng N (2018) Recognition of the diglycine C-end degron by CRL2KLHDC2 ubiquitin ligase. *Mol Cell* 72: 813–822.e814
- Sriram SM, Kim BY, Kwon YT (2011) The N-end rule pathway: emerging functions and molecular principles of substrate recognition. *Nat Rev Mol Cell Biol* 12: 735–747
- Tanaka T, Kamitani W, DeDiego ML, Enjuanes L, Matsuura Y (2012) Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA. *J Virol* 86: 11128–11137
- Thakran P, Pandit PA, Datta S, Kolathur KK, Pleiss JA, Mishra SK (2018) Sde2 is an intron-specific pre-mRNA splicing regulator activated by ubiquitin-like processing. *EMBO J* 37: 89–101
- Timms RT, Zhang Z, Rhee DY, Harper JW, Koren I, Elledge SJ (2019) A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* 365: eaaw4912
- Tsirigos KD, Peters C, Shu N, Kall L, Elofsson A (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 43: W401–407
- Ueda E, Tamura Y, Sakae H, Kawano S, Kakuta C, Matsumoto S, Endo T (2019) Myristoyl group-aided protein import into the mitochondrial intermembrane space. *Sci Rep* 9: 1185
- Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. *Protein Sci* 20: 1298–1345
- Varshavsky A (2019) N-degron and C-degron pathways of protein degradation. *Proc Natl Acad Sci USA* 116: 358–366
- Vitali DG, Sinzel M, Bulthuis EP, Kolb A, Zabel S, Mehlhorn DG, Figueiredo Costa B, Farkas Á, Clancy A, Schuldiner M et al (2018) The GET pathway can increase the risk of mitochondrial outer membrane proteins to be mistargeted to the ER. *J Cell Sci* 131: jcs211110
- Wan J, Qian SB (2014) TISdb: a database for alternative translation initiation in mammalian cells. *Nucleic Acids Res* 42: D845–850
- Wang X, Li S (2014) Protein mislocalization: mechanisms, functions and clinical applications in cancer. *Biochim Biophys Acta* 1846: 13–25
- Williamson A, Werner A, Rape M (2013) The Colossus of ubiquitylation: decrypting a cellular code. *Mol Cell* 49: 591–600
- Yen HC, Xu Q, Chou DM, Zhao Z, Elledge SJ (2008) Global protein stability profiling in mammalian cells. *Science* 322: 918–923
- Zheng N, Shabek N (2017) Ubiquitin ligases: structure, function, and regulation. *Annu Rev Biochem* 86: 129–157



**License:** This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.