Published in final edited form as:

Biochemistry. 2018 July 17; 57(28): 4225–4235. doi:10.1021/acs.biochem.8b00565.

Dissecting substrate specificities of the mitochondrial AFG3L2 protease

Bojian Ding[†], Dwight W. Martin^{‡,§}, Anthony J. Rampello[†], and Steven E. Glynn^{†,*}

[†]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY.

[‡]Department of Medicine, Stony Brook University, Stony Brook, NY.

[§]Proteomics Center, Stony Brook University, Stony Brook, NY.

Abstract

Human AFG3L2 is a compartmental AAA+ protease that performs ATP-fueled degradation at the matrix face of the inner mitochondrial membrane. Identifying how AFG3L2 selects substrates from the diverse complement of matrix-localized proteins is essential for understanding mitochondrial protein biogenesis and quality control. Here, we create solubilized forms of AFG3L2 to examine the enzyme's substrate specificity mechanisms. We show that conserved residues within the pre-sequence of the mitochondrial ribosomal protein, MrpL32, target the subunit to the protease for processing into a mature form. Moreover, these residues can act as a degron, delivering diverse model proteins to AFG3L2 for degradation. By determining the sequence of degradation products from multiple substrates using mass spectrometry, we construct a peptidase specificity profile that displays constrained product lengths and is dominated by the identity of the residue at the P1' position, with a strong preference for hydrophobic and small polar residues. This specificity profile is validated by examining the cleavage of both fluorogenic reporter peptides and full polypeptide substrates bearing different P1' residues. Together, these results demonstrate that AFG3L2 contains multiple modes of specificity, discriminating between potential substrates by recognizing accessible degron sequences, and performing peptide bond cleavage at preferred patterns of residues within the compartmental chamber.

Graphical Abstract

*Corresponding Author: Steven E. Glynn, Department Biochemistry and Cell Biology, 450 Life Sciences Building, Stony Brook University, Stony Brook, NY, 11794-5215, USA; steven.glynn@stonybrook.edu, tel: 1-631-632-1055; fax: 1-631-632-9730.

Author Contributions

All authors have given approval to the final version of the manuscript.

Supporting Information

The supporting information pdf file contains: additional details on the expression and size exclusion chromatography of the AFG3L2 constructs used in this study; full uncropped representative SDS-PAGE images for all degradation reactions; sequence logos for all individual substrates; additional details on the cleavage of fluorogenic reporter peptides.

INTRODUCTION

Mitochondria utilize complex systems of proteostasis to maintain energy production, calcium signaling, and fatty acid oxidation, among other activities in eukaryotic cells^{1–4}. The m-AAA protease is a significant contributor to this regulation, performing surveillance and biogenesis of proteins within the mitochondrial matrix^{5–7}. Subunits of m-AAA are inserted into the mitochondrial inner membrane through dual transmembrane helices separated by a small intermediate domain⁵. The orientation of the transmembrane domains projects fused AAA+ ATPase and M41-family zinc metallopeptidase domains into the matrix where they can access both soluble and membrane-embedded substrates⁵. Hexameric assembly of m-AAA proteases creates a ring of ATPase domains that sit atop a compartmental peptidase chamber. Substrates are thought to initially bind to the external surface of the protease before being fed through a narrow central pore into the proteolytic chamber by processive cycles of ATP-driven unfolding and translocation. Once inside, unfolded substrates can access six peptidase active sites for cleavage into small peptide fragments. Both yeast and human m-AAA proteases assemble into heterohexamers of alternating subunits (Yta10 and Yta12 in yeast; AFG3L2 and Paraplegin in humans)^{8–11}. However, as in other related membrane anchored AAA+ proteases such as FtsH and mitochondrial i-AAA, the human m-AAA protease can also assemble into homohexamers of AFG3L2^{5, 12–14}. Loss of AFG3L2 results in severe pleiotropic phenotypes including developmental defects, mitochondrial fragmentation, and impaired mitochondrial transport^{15–18}. Furthermore, at least 16 amino acid substitutions located in the catalytic domains of AFG3L2 are linked to the development of spinocerebellar ataxia type 28 (SCA28), a severe neurodegenerative disease characterized by progressive gait, limb ataxia and dysarthria^{19–28}. Although m-AAA proteases bearing these mutations have been demonstrated to have impaired proteolytic activity of *in vivo*, the lack of an *in vitro* system for biochemical and biophysical characterization has prevented an examination of the molecular mechanisms linking mutation to malfunction¹⁹.

Established functions of m-AAA proteases include the biogenesis and maturation of diverse mitochondrial proteins, including non-assembled respiratory chain subunits and ATP synthase subunit A^{29, 30}. Recently, m-AAA proteases in mice were shown to secure the gatekeeping function of the mitochondrial Ca²⁺ uniporter (MCU) by turning over the MCU

subunit, EMRE³¹. While the majority of identified m-AAA substrates are membrane-anchored, the mitochondrial ribosomal subunit, MrpL32, is a well-established substrate in the mitochondrial matrix³². MrpL32 precursors are imported into the matrix bearing a long unstructured N-terminal pre-sequence that is removed by m-AAA prior to assembly into the large subunit³³. Proper processing is required for respiratory growth in yeast and is dependent on the integrity of a zinc-binding motif within the tightly folded C-terminal domain of MrpL32³⁴. Current evidence supports a model where the N terminal pre-sequence is recognized and translocated into the peptidase chamber and degraded until steric clashes between the C-terminal domain and the protease surface prompt release of the mature subunit³⁴. However, how MrpL32 precursors are selected by the protease is not known.

Proteases frequently exhibit tight regulation to limit the threat of uncontrolled protein degradation to cell viability. Substrate specificity is a key mechanism used to constrain proteolysis to a precise collection of substrate proteins. Specificity in AAA+ proteases has been shown to occur at multiple levels. Many family members specifically recognize accessible sequence degrons present on the substrate³⁵. Indeed, a 15 residue degron was recently identified that is sufficient to target proteins to the mitochondrial intermembrane space protease, Yme1³⁶. Moreover, some compartmental proteases selectively cleave polypeptides only within certain sequences of residues, controlled by the fine structure of the peptidase active sites^{37–40}. Unlike the 20S proteasome core particle, which contains three distinct peptidase subunits exhibiting different cleavage specificities, AFG3L2 homohexamers must efficiently degrade proteins containing highly diverse sequences using only a single type of peptidase active site^{5, 41}. Cleavage specificity for the M41-family of zinc metalloproteases has not yet been characterized. Therefore, how AFG3L2 both identifies its substrates and efficiently cleaves them in the peptidase chamber are open questions.

To investigate the mechanisms of substrate selection by AFG3L2, we designed solubilized versions of the homohexameric enzyme suitable for *in vitro* study. We use these proteases to probe specificity at both the substrate degron and cleavage site levels. These experiments show that an MrpL32 precursor is targeted to the protease by a degron sequence within the N-terminal pre-sequence. Moreover, this sequence is sufficient to direct model proteins to the protease for degradation. By comparing the degradation products from a diverse set of protein substrates, we show that substrate sequence strongly influences the pattern of cleavages within the peptidase chamber and that the length of peptide products is constrained. Together, these results demonstrate that AFG3L2 operates using multiple modes of substrate specificity to select and process proteins for the maintenance of mitochondrial integrity.

EXPERIMENTAL PROCEDURES

Construct design and cloning

The ^{core}AFG3L2 and ^{cchex}AFG3L2 constructs were produced by PCR amplifying a sequence encoding residues 272–797 of human AFG3L2 from cDNA (Accession # BC065016), and sub-cloning into the 2G-T vector (Addgene ID 29707) or a previously described modified 2G-T vector (2-GT-cc-hex) containing a sequence encoding cc-hex

followed by a ten-residue linker⁴². All ^{core}AFG3L2 and ^{cchex}AFG3L2 variants were produced by site-directed mutagenesis using plasmids encoding ^{core}AFG3L2 or ^{cchex}AFG3L2 as templates. Plasmids containing sequences encoding ^{SF}GFP-10/11^{A226G}, cp7-^{SF}GFP- β 20, I27^{CD}, I27- β 20, and β 20- λ CI^N were gifts from Prof. Robert Sauer (MIT)^{43–45}. Purified *Y. pestis* HspQ^{Y20} variants and Y2853 proteins were gifts from Dr. Neha Puri (Stony Brook University)⁴⁶. A plasmid encoding *S. cerevisiae* Tim9^N was produced as previously described³⁶. ^{mut}GFP- β 20 was produced by addition of the β 20 sequence to the C-terminus of ^{SF}GFP-10/11^{A226G} by PCR. DNA encoding human MrpL32 was synthesized (Genewiz) and sequences were added to the N-terminus of I27^{CD} and Tim9^N by PCR. MrpL32 truncations were produced by inserting sequences encoding human MrpL32 residues 31–188 or 51–188 into a modified 2S-U vector³⁶. All MrpL32 variants were altered by site-directed mutagenesis with the template of 2S-U-³⁰MrpL32.

Protein expression and purification

All ^{core}AFG3L2 and ^{cchex}AFG3L2 variants were expressed in *E. coli* BL21-CodonPlus cells (Agilent) using an identical procedure to that previously described for ^{cchex}YME1³⁶. Cells containing ^{cchex}AFG3L2 were harvested and re-suspended in buffer L1 (20 mM Tris-HCl (pH 7.8), 300 mM NaCl, 0.1 mM EDTA, 10% glycerol, 10 mM β -mercaptoethanol) supplemented with 1 mM PMSF and lysed by sonication. Cell lysate was clarified by centrifugation and applied to a Glutathione Superflow Agarose column (Pierce). Unbound proteins were removed by washing with buffer L1 and bound proteins were eluted by addition of buffer L1 supplemented with 10 mM reduced glutathione. 1 mg of His-tagged TEV protease was added per L of cell culture and incubated at 4 °C for 16 hours to remove the N-terminal His₆-GST tag. Digested proteins were applied to a Ni-NTA column (Thermo Scientific) to separate His₆-GST and TEV protease from ^{cchex}AFG3L2. Flow through was collected, concentrated and applied to a Superose 6 10/300 GL Increase column (GE Healthcare) equilibrated with buffer S1 (20 mM Tris-HCl (pH 7.8), 100 mM NaCl, 0.1 mM EDTA, 10% glycerol and 1 mM DTT). Fractions corresponding to the hexamer were pooled, concentrated and flash-frozen in liquid nitrogen. The ^{core}AFG3L2 variants were purified using a similar protocol to ^{cchex}AFG3L2 with the following modifications. Buffer L1 was replaced with buffer L2 (50 mM HEPES-HCl (pH 7.5), 300 mM KCl, 0.1 mM EDTA, 10% glycerol, and 10 mM β -mercaptoethanol). Size exclusion chromatography was performed using a Superdex 200 10/300 GL Increase column equilibrated with buffer S2 (25 mM HEPES-HCl (pH 7.5), 100 mM KCl, 10% glycerol and 1 mM DTT).

All substrate proteins were expressed in *E. coli* BL21-CodonPlus cells at 37 °C in LB supplemented with 100 μ g/ml ampicillin or 50 μ g/ml kanamycin as appropriate, and 34 μ g/ml chloramphenicol until OD₆₀₀ = 0.6. Expression was induced by an addition of 1 mM isopropyl- β -D-thiogalactopyranoside (IPTG), followed by growth at 16 °C for 16 hours. All MrpL32 variants were expressed in LB supplemented with 500 μ M zinc acetate and zinc sulfate. All GFP variants, I27- β 20, β 20- λ CI^N, Tim10, and Tim9 variants, were purified using previously described protocols^{36, 44, 45}. For all MrpL32 and I27^{CD} variants, cells were harvested, re-suspended and lysed in buffer L3 (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10% glycerol, 10 mM imidazole, and 10 mM β -mercaptoethanol) supplemented with 1 mM PMSF. Cell lysate was clarified by centrifugation and supernatant was applied to a Ni-NTA

column. Unbound proteins were removed by washing with buffer L3 supplemented with 50 mM imidazole and bound proteins eluted by addition of buffer L3 supplemented with 250 mM imidazole. 0.3 mg His-tagged Ulp1 protease was added per L of cell culture, and the mixture was incubated at 4 °C for 16 hours. Digested proteins were buffer exchanged into buffer L4 (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10% glycerol, and 10 mM β -mercaptoethanol) and applied to a Ni-NTA column. Flow through was collected and applied to a Superdex 200 10/300 GL Increase column equilibrating in buffer S3 (20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10% glycerol, and 1mM DTT). Fractions were pooled, concentrated and flash frozen in liquid nitrogen.

Biochemical Assays

ATPase assays were performed as previously described with the following modifications³⁶. The reaction buffer contained 2 mM ATP and adjusted to pH 7.5. All reactions were conducted at 37 °C containing 1 μ M enzyme in a 384-well clear bottom plate (Corning) using a SpectraMax M5 plate reader (Molecular Devices). All degradation assays were conducted at 37 °C using 0.5 μ M or 1 μ M enzyme in PD buffer (25 mM HEPES-KOH (pH 7.5), 100 mM KCl, 5 mM MgCl₂, 10% glycerol, 1 mM DTT) supplemented with an ATP regeneration system (2 mM ATP, 18.75 U ml⁻¹ PK, and 20 mM PEP). Steady-state ATPase data were fit to the Hill version of the Michaelis-Menten equation [$v = k_{\text{ATPase}}/(1 + K_{0.5}/[\text{ATP}]^n$]. Gel based degradation reactions (70 μ l total) contained 5 μ M (MrpL32 variants) or 20 μ M substrate (all other proteins). Time point aliquots were removed and quenched by addition of Laemlli sample buffer containing 2 % SDS at 90 °C for 4 min prior to application to an SDS-PAGE gel. SDS-PAGE band intensities were quantified as previously described³⁶. Full uncropped representative SDS-PAGE images for all degradation reactions are shown in their respective supporting figure. Fluorescence-based degradation reactions (30 μ l) were carried out in a 384-well black plate (Corning) using a SpectraMax M5 plate reader (ex = 467 nm; em = 511 nm). Initial degradation rates were calculated from the loss of fluorescence over early linear time points. Fluorogenic peptide cleavage assays were carried out at 37 °C using 1 μ M enzyme and 50 μ M peptide (GenScript) in PD buffer. All reactions (60 μ l) were measured in a 384-well plate (Corning) using a SpectraMax M5 plate reader (ex = 320 nm; em = 420 nm). Initial cleavage rates were determined by measuring the loss of fluorescence over early linear time points. Values shown for all kinetic experiments are means of independent replicates (n=3) $\pm$ s.d. **P* 0.05, ***P* 0.01, ****P* 0.001, *****P* 0.0001 as calculated using the Student's two-tailed t-test.

Mass Spectrometry and Data Analysis

All degradation reactions for mass spectrometry were performed using a modified protocol as that described above. Reactions (80 μ l total) containing 1 μ M ^{cchex}AFG3L2 and 20 μ M substrate were incubated at 37 °C for 4 hours and quenched by addition of 10 mM EDTA. Completion of each reaction was confirmed by SDS-PAGE. For identification of accumulated products from MrpL32 processing experiments, excised gel pieces were destained, reduced, alkylated and digested with trypsin (Promega Gold, Mass Spectrometry Grade) using a previously described protocol⁴⁷. All peptide extracts were dried under vacuum and dried peptides were dissolved into buffer A (2% Acetonitrile (ACN), 0.1% Formic Acid (FA)) prior to LC/MS/MS analysis. For identification of cleavage site

specificities, peptides were isolated from degradation reactions described above. To one volume of quenched reaction solutions, 0.25 volumes of BSA (10 mg/ml) were added and the solutions were vortexed briefly. To these mixtures, 0.375 volumes of TCA (72 %) were added and the mixtures were vortexed and incubated on ice for 30 min. The resulting suspensions were centrifuged for 15 min at 16000 g and 4 °C. Supernatants containing the peptides were brought to 2 % formic acid and desalted with Pierce C18 micropipette tips using stepped elutions with 0.1% FA in 20–60% ACN. The solvent was removed from the eluted peptides under vacuum and the resultant dried peptides stored at –80 °C. The dried peptides were dissolved in buffer A for analysis by LC/MS/MS.

Fused-silica capillaries (100 μ m inner diameter (i.d.)) were pulled using a P-2000 CO₂ laser puller (Sutter Instruments, Novato, CA) to a 5 μ m i.d. tip and packed with 10 cm of 5 μ m ProntoSil 120–5-C18H (Bischoff Chromatography, Leonberg, Germany) using a pressure bomb. The samples were loaded via a Dionex WPS-3000 autosampler (Germering, Germany). The column was installed in-line with a Dionex LPG-3000 Chromatography HPLC pump running at 300 nL min^{–1}. The peptides were eluted from the column by applying gradients of buffer B (98 % ACN, 0.1 % FA) (5 min 2–10 %; 60 min 10–45%; 10 min 45–80%; 2 min 80–2%; 20 min 2 %). The application of a 2.2 kV distal voltage electrosprayed the eluting peptides directly into an LTQ Orbitrap XL ion trap mass spectrometer (Thermo Fisher, San Jose, CA) equipped with a nano-liquid chromatography electrospray ionization source. Full mass spectra (MS) were recorded on the peptides over a 400 to 2000 *m/z* range at 60,000 resolution, followed by top-five MS/MS scans in the ion-trap. Charge state dependent screening was used to analyze peptides with a charge state of +2 or higher. Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (Thermo Fisher, San Jose, CA). MS/MS spectra were extracted from the RAW file with ReAdW.exe (<http://sourceforge.net/projects/sashimi>). The resulting mzXML data files were searched with Inspect without protease specificity against a custom database composed of a EColi_K12 proteome with added sequences for ^{cchex}AFG3L2, protease substrates and common contaminants⁴⁸. Peptide mass searches included the option for methionine oxidation (+15.994915). For in-gel trypsin digest samples, search parameters were protease=trypsin, fixed C+57.021464 and optional M+15.994915.

Post-processing of Inspect output was performed using Inspect subroutines PValue.py and Summary.py applying thresholds for protein identification of 2 peptides, 2 spectra, and *p*=0.02 confidence level. Post-processing produced a list of peptides corresponding to all assignable MS/MS spectra. There were >2 spectra for most peptides. The six amino acids making up the C-terminus of each mass spectrometry identified peptide was designated as a P segment of a cleavage site and the six amino acids from the N-terminal was designated as a P' segment. Theoretical P and P' segments were generated by fractionating substrate protein sequences into a list of overlapping 12 amino acid fragments using PeptGen (www.hiv.lanl.gov/content/sequence/PEPTGEN/peptgen.html) stepped sequentially by 1 amino acid. The 12mer of the cleavage site for each observed peptide in the mass spectrometry data was identified by searching against the theoretical P and P' segments. Sequence logos were generated for each individual substrate as well as from a master list of all 12mers from all substrates using WebLOGO (weblogo.berkeley.edu) to generate the specificity profile. Oxidized Methionine (M-ox) was not treated as a distinct amino acid

during the 12mer identification process. Subpocket cleavage entropy (S_i) values for all 12 positions of the specificity profile were calculated from the master list of 12mers as previously described⁴⁹.

RESULTS

Creating a solubilized active AFG3L2 protease

To examine the mechanisms of proteolysis by AFG3L2, we sought to develop a solubilized homohexameric variant of the protease for *in vitro* study. Previous crystallographic and biochemical studies of the homologous bacterial FtsH protease have shown that the catalytic ATPase and peptidase domains can assemble into active enzymes in the absence of the N-terminal transmembrane regions^{12, 50}. Therefore, we first generated a construct comprising the ATPase and peptidase domains of human AFG3L2 (^{core}AFG3L2; residues 272–797) (Figure 1A). However, during expression in *E. coli*, the levels of ^{core}AFG3L2 dropped rapidly under continuous induction (Figure S1A). Incorporation of inactivating substitutions within the ATPase Walker-B motif (^{core}AFG3L2^{E408Q}), which serves as an ATP trap by abolishing ATP hydrolysis but not binding, or the peptidase active site (^{core}AFG3L2^{E575Q}) significantly increased protein levels during expression (Figure S1A). In size exclusion chromatography, ^{core}AFG3L2^{E408Q} largely migrated as a hexamer compared to ^{core}AFG3L2^{E575Q}, which formed unassembled lower molecular weight species, suggesting that hexamers are stabilized by ATP binding (Figure S1B). Incubation of ^{core}AFG3L2^{E575Q} with a non-hydrolysable analogue, AMP-PNP, shifted the migration profile towards the hexameric species (Figure S1B). From these results, we conclude that the wild-type ^{core}AFG3L2 enzyme is auto-degraded during expression and that, although ATPase-inactive hexameric catalytic cores can be isolated, alternative protein engineering methods were required to produce fully active AFG3L2 enzymes.

Recently, we have reconstituted active solubilized AAA+ proteases by replacing the transmembrane domains with a hexamerizing sequence (cchex) that promotes oligomerization of the catalytic domains^{36, 42, 51}. Fusion of cchex to the N-terminus of ^{core}AFG3L2 separated by a flexible linker produced a protein (^{cchex}AFG3L2) that was well expressed in *E. coli* and migrated as a hexamer by size exclusion chromatography (Figure S1A and S1C). Moreover, ^{cchex}AFG3L2 rapidly hydrolyzed ATP demonstrating proper formation of the ATPase active sites between subunit interfaces ($k_{\text{ATPase}} = 91 \text{ ATPs min}^{-1} \text{ enz}_6^{-1}$; $K_{0.5} = 44.1 \mu\text{M}$) (Figure 1B). To demonstrate that ^{cchex}AFG3L2 is capable of performing all aspects of ATP-dependent protein degradation, we incubated the protease with a low stability GFP variant described by Sauer and co-workers (mutGFP) bearing a C-terminal $\beta 20$ degron (^{mut}GFP- $\beta 20$)⁴³. $\beta 20$ is a hydrophobic 20-residue sequence that has been shown to target proteins for degradation by multiple AAA+ proteases, including human YME1L and *E. coli* Lon^{42, 45}. Degradation of ^{mut}GFP- $\beta 20$ by ^{cchex}AFG3L2 was observed in the presence but not absence of ATP ($k_{\text{deg}} = 0.28 \text{ GFPs min}^{-1} \text{ enz}_6^{-1}$) with an affinity ($K_M = 6.2 \mu\text{M}$) comparable to a $\beta 20$ tagged circularly permuted ^{SF}GFP variant by human ^{cchex}YME1L⁴² (Figure 1C–D; Figure S1D). Furthermore, no degradation was observed by an inactivated Walker-B variant (^{cchex}AFG3L2^{E408Q}) in the presence of ATP, confirming the degradation activity is specific to ^{cchex}AFG3L2 (Figure 1C; Figure S1D). These results

confirm that ^{cchex}AFG3L2 can be employed to study the mechanism of ATP-dependent degradation in solution.

MrpL32 processing is dependent on a sequence in the unstructured N-terminus.

We first used the *in vitro* protease to understand how AFG3L2 selectively recognizes the precursor form of ribosomal subunit MrpL32. Based on a cryoEM structure of the assembled yeast mitochondrial ribosome, we estimate that the MrpL32 precursor contains ~86 residues in the unstructured N-terminal region⁵². Attempts to purify full-length human MrpL32 precursor failed as the protein was insoluble. However, removal of residues 1–30 produced a soluble protein (³⁰MrpL32) that underwent proteolysis by ^{cchex}AFG3L2 (Figure 2A). A smaller molecular weight product accumulated in the reactions concurrent with the loss of ³⁰MrpL32 (Figure 2A–B; Figure S2A). LC/MS/MS confirmed the accumulating product as residues 78–188 of MrpL32, indicating that residues 31–77 had been removed by the protease (Figure 2C). Further truncation of MrpL32 to residue 51 (⁵⁰MrpL32) significantly reduced the rate of precursor loss implying a crucial role for residues 31–50 in targeting the subunit to the protease (Figure 2A; Figure 2D; Figure S2A). Moreover, no obvious accumulation of a smaller molecular weight fragment was observed for ⁵⁰MrpL32, likely indicating that this variant is fully degraded rather than undergoing limited processing (Figure 2A).

To examine the ability of the pre-sequence to direct proteins to the AFG3L2, fragments of approximately twenty-residues spanning residues 1–80 of MrpL32 were fused to the N-terminus of the model unfolded protein I27^{CD} and tested for degradation by ^{cchex}AFG3L2 (Figure 3A; Figure S3). All I27^{CD} variants were expressed bearing a SUMO tag that was subsequently removed to yield a scar-less N-terminus. As with full-length MrpL32, I27^{CD} fusions bearing sequences derived from residues 1–30 were insoluble. Addition of residues 31–50 (^{31–50}I27^{CD}) dramatically increased the degradation rate compared to untagged I27^{CD}, with a smaller but still significant increase observed after addition of residues 40–60 (^{40–60}I27^{CD}) (Figure 3B). In this construct, the presence of a proline at residue 41 mandated the inclusion of Ser-40 to enable removal of the SUMO tag. Only a very small increase in rate was observed after addition of 61–80 (^{61–80}I27^{CD}) and addition of residues 51–70 (^{51–70}I27^{CD}) did not produce any increase in rate over I27^{CD} alone (Figure 3B). The most effective targeting sequence was further narrowed down by testing smaller fragments from within the 31–50 region (Figure 3A). Residues 40–50 (^{40–50}I27^{CD}) largely replicated the large increase in degradation rate observed in ^{31–50}I27^{CD} (Figure 3B; Figure S3). However, no notable increase was observed after addition of residues 31–40 (^{31–40}I27^{CD}) (Figure 3B; Figure S3). Addition of residues 40–50 to the N-terminus of a truncated variant of the mitochondrial chaperone Tim9 (^{40–50}Tim9^N) also significantly increased degradation rate compared to Tim9^N alone, demonstrating that this sequence can target diverse proteins to ^{cchex}AFG3L2 for degradation (Figure 3C; Figure S3). To discern the importance of residues 40–50 in MrpL32 processing, this region was replaced within ³⁰MrpL32 with a sequence identical to residues 51–61 (³⁰MrpL32_{swap40–50}), which did not induce degradation of I27^{CD} (Figure S2C). A dramatic reduction in the rate of precursor loss was observed for ³⁰MrpL32_{swap40–50} (Figure 3D; Figure S2B), whereas an alternative substitution of residues 31–39 with a sequence identical to residues 51–59 had no apparent effect on

proteolysis (Figure 3D; Figure S2B; Figure S2C). Together, these results demonstrate that AFG3L2 can discriminate between substrates on the basis of terminal sequence and suggest that residues 40–50 of MrpL32 play a significant role in targeting the precursor to the protease for processing.

Substrate sequence determines cleavage within the peptidase chamber

In addition to the recognition of substrate degron sequences, proteases can achieve specificity by selectively cleaving peptide bonds based on the pattern of residues proximal to the scissile bond. Such cleavage site preferences have been identified in some AAA+ proteases despite the need for these enzymes to degrade substrates with highly diverse sequences^{37–40}. A close examination of the products of ³⁰MrpL32 processing by ^{cchex}AFG3L2 revealed twenty-five unique cleavage sites within the precursor N-terminus with varying frequencies. Little information exists on the cleavage specificity of the M41 family of zinc-metalloproteases, to which AFG3L2 belongs. We sought to identify the enzyme's peptidase specificity profile to understand how a single type of peptidase active site can cleave highly diverse sequences. Eight well-degraded proteins were independently digested with ^{cchex}AFG3L2 at equal concentration (20 μM) in the presence of ATP (Figure 4A; Figure S4A). In addition to ³⁰MrpL32, seven model proteins bearing both N- or C-terminal degrons confirmed in prior AAA+ protease studies were included to increase sequence diversity and provide information on degradation from either terminus. Reactions were quenched after 4 hours, and peptide products isolated by trichloroacetic acid precipitation. Peptide sequences were determined by LC/MS/MS mass spectrometry and aligned to their respective full-length proteins to identify cleavage sites. From a total group of 3151 peptides, the sequences of the 12 residues surrounding each scissile bond were used to generate the specificity profile (Figure 4B; Figure S4B). Placing the scissile bond in the center of the profile, residues are designated P1→P_n to the N-terminus and P1'→P_n' to the C-terminus⁵³. The most striking feature in the specificity profile is a pronounced preference for either hydrophobic or small polar residues in the P1' position. The most commonly found residue in this position was Phe followed by Leu>Ser>Ala>Val>Ile>Thr. Less pronounced preferences were identified for hydrophobic residues in the P2 position (Leu>Val>Ala>Ile) and for a diverse collection of residues in the P2' position (Gln>Val>Glu>Ser). Thus, peptide bond cleavage in AFG3L2 appears to be dominated by the identity of the residue in the P1' position. Based on the distribution of different residues within the mass spectrometry data, we calculated the subpocket cleavage entropy (S_i) of the P1' position to be 0.832, indicating a specific cleavage site⁴⁹ (Figure 4B). No other position within the profile reached the threshold for specificity ($S_i < 0.85$). Analysis of peptide lengths revealed a broad distribution of products with 14-mers occurring with the greatest frequency and 90 % of peptides containing between 9 and 24 residues (Figure 4C).

To validate the importance of the P1' residue in promoting cleavage, we examined the cleavage of small fluorogenic reporter peptides by AFG3L2. To ensure that reporter cleavage was governed by peptidase activity alone, we removed any contribution from ATP hydrolysis by employing the ^{core}AFG3L2^{E408Q} variant that forms hexamers in the absence of cchex. Moreover, lack of the cchex domain above the central ATPase pore should allow the reporter peptides to more freely diffuse into the peptidase chamber. Short pentapeptide

reporters were derived from the sequence Leu-Tyr-Phe-Gln identified from a highly abundant cleavage in the proteomic analysis between the Tyr (P1) and Phe (P1') residues. Each peptide contained a 3-nitrotyrosine (3-NO₂-Tyr) quencher at position 2 and an additional modified lysine residue bearing a sidechain-linked 2-aminobenzozyl fluorophore (Lys-Abz) at the C-terminus. Proteolytic cleavage at any position between these two residues releases the Lys-Abz fluorophore to yield a detectable increase in fluorescence. Based upon the specificity profile, the Phe residue at position 3 should occupy the P1' site, placing 3-NO₂-Tyr in the P1 site, where little sequence preference was detected (Figure 5A). As expected, this Leu-(3-NO₂-Tyr)-Phe-Gln-(Lys-Abz) reporter was rapidly cleaved by coreAFG3L2^{E408Q} (Figure 5A–B). Substitution of Phe with either Leu or Ser reduced the cleavage rate in agreement with their lower abundance at this position of the specificity profile (Figure 5A–B). Furthermore, substitution with Gly abolished cleavage consistent with the very low abundance of Gly residues in the P1' position (Figure 5A–B). Interestingly, unassembled coreAFG3L2^{E408Q} subunits isolated during size exclusion chromatography did not cleave the reporter peptides, comparable to a peptidase-inactive variant (coreAFG3L2^{E408Q/E575Q}) (Figure S5A). Thus, it appears that hexameric assembly is required to properly activate the peptidase sites.

Contributions from residues distal to the scissile bond could be lost when examining cleavage of short peptide reporters. Therefore, we tested whether the cleavage preferences observed in the fluorogenic peptides could be replicated in the context of a whole protein substrate. Analysis of the cleavage products from one substrate, a variant of the *Y. pestis* HspQ protein (HspQ^{Y20}), revealed a frequently cleaved site between Tyr-14 (P1) and Phe-15 (P1'). This site is located in a linker region between a His₆ affinity tag and the folded HspQ domain and thus should be amenable to substitution without altering the protein stability. We reasoned that substituting Phe-15 for less preferred residues would reduce the number of peptides detected by LC/MS/MS resulting from cleavage at this site. Indeed, substitution to either Leu or Gly progressively reduced the number of peptides generated from this cleavage site with no notable change to the total number of peptides detected from the degradation (Figure 5C; Figure S5B). In both cases the reduction in cleavage between residues 14–15 correlated with increased cleavage between residues 13–14 (Leu-Tyr) (Figure 5D). The result of this compensatory cleavage is to maintain the length of peptide products produced from this region close to the most commonly observed length of 14 residues. No change was detected in the number of peptides generated from a second frequently cleaved site located within the folded HspQ domain implying that the substitutions do not impair degradation of the protein (Figure S5C). Together, these results provide a strong validation of the specificity profile and demonstrate that the pattern of cleavage events within the peptidase chamber is heavily influenced by the sequence of the substrate protein.

DISCUSSION

The capacity of AFG3L2 to regulate both protein biogenesis and quality control relies upon the selection of specific substrates from among a diverse complement of mitochondrial proteins. Our results demonstrate that at least two forms of substrate specificity can occur within AFG3L2: (1) discrimination between potential substrates on the basis of their accessible sequences and (2) preferred cleavage of peptide bonds within the proteolytic

chamber dependent on the pattern of residues close to the scissile bond. Moreover, we demonstrate that a precursor form of human MrpL32 is targeted to AFG3L2 by a region within the pre-sequence, suggesting that the protease is capable of recognizing patterns of residues at internal positions.

Proteolysis of $^{30}\text{MrpL32}$ by $^{\text{cchex}}\text{AFG3L2}$ resulted in the accumulation of a N-terminally truncated product beginning at Met78. In a recent cryoEM structure of the assembled human mitochondrial ribosome, the N-terminus of the mature MrpL32 subunit was modeled as Ala79, identical to that of the yeast homolog as determined by mass spectrometry (Ala72)^{32, 52}. The close similarity in these positions suggests that the partial proteolysis of $^{30}\text{MrpL32}$ in our assays is functionally similar to the maturation of the subunit *in vivo*. The single residue difference may be due to additional steric clashes between the cchex domain and the folded C-terminal domain of MrpL32, which do not occur in the membrane-embedded protease. Importantly, multiple peptides corresponding to different fragments of the MrpL32 N-terminus were identified within the proteomic study, indicating that the unstructured pre-sequence is progressively degraded until the enzyme meets the folded C-terminal domain, rather than undergoing a single cleavage event to generate the mature N-terminus. In our experiments, MrpL32 processing appears to be dependent on the presence of residues 40–50. This sequence is highly conserved in mammals and to a lesser extent in yeast, suggesting that these residues may serve as an internal degron to allow AFG3L2 to specifically recognize or engage the precursor (Figure 2C). The ability of this sequence to target unrelated model proteins to the protease supports this model. However, we note that our inability to purify constructs containing residues 1–30 leaves open a functional role for these residues. Moreover, while constructs containing residues 40–50 displayed the highest degradation rate, the rate of $^{31-50}\text{I27}^{\text{CD}}$ was greater than $^{40-60}\text{I27}^{\text{CD}}$, despite them sharing the key residues. This difference could signify a role for less prominent contributions arising from other residues in the pre-sequence or residues within I27^{CD} . A smaller difference between $^{31-50}\text{I27}^{\text{CD}}$ and $^{40-50}\text{I27}^{\text{CD}}$ was not statistically significant but could also arise from the absence of these additional contributions. The ability of residues 40–50 to specifically target proteins to the protease suggests that AFG3L2 can discriminate between potential substrates by surveying patterns of accessible residues. Recent studies have demonstrated that the mitochondrial calcium uniporter subunit, EMRE, is degraded by m-AAA *via* engagement of a matrix-localized N-terminus. In contrast to our results, replacement of the EMRE N-terminus with a random sequence had no effect on its degradation, suggesting that recognition of the substrate is sequence-independent⁵⁴. Given the diverse roles that m-AAA plays in mitochondrial proteostasis, it is possible that the protease contains multiple modes of substrate recognition, both engaging accessible unstructured regions independent of sequence for house-keeping degradation, and recognizing residue patterns to select specific protein targets.

Based on our mass spectrometry data, the peptidase sites of AFG3L2 appear to sample substrate polypeptides at the P1' residue with a strong preference for phenylalanine and other hydrophobic residues, as well as small polar residues such as serine and threonine. By preferentially cleaving sequences containing hydrophobic residues, the enzyme may be well placed to degrade the many membrane-associated proteins identified as m-AAA substrates. Interestingly, we observed no obvious difference in cleavage site preference for either N-

and C-terminally tagged proteins, suggesting that the orientation of the substrate in the peptidase active site is not determined by the direction of translocation into the chamber (Figure S4). AFG3L2 belongs to the M41 family of zinc metallopeptidases based on similarity to the archetypal bacterial FtsH protease⁹. Numerous studies have identified the P1' residue as the critical specificity determinant for many other zinc metalloprotease families^{55–57}. Moreover, the subpocket cleavage entropy value we determined for the P1' position of AFG3L2 is consistent with S_i values calculated for a wide variety of metalloproteases, suggesting comparable degrees of specificity and potentially related modes of substrate binding. Structures of these proteases reveal a key S1' binding pocket close to the active site that is tailored to accommodate the P1' residue^{58–61}. For example, modifying the S1' pocket of thermolysin-like protease dramatically altered the enzyme's residue preference at P1'⁶². In contrast, many serine proteases, including other AAA+ proteases containing serine peptidase activities, display a dominant preference for the P1 residue, across the scissile bond from the P1' site^{37–40}. Again, this preference matches well with the identification of S1 binding pockets in these enzymes^{37, 63, 64}. We expect that the P1' preferences seen in AFG3L2 reflect the structure of the enzyme's peptidase binding pocket that can easily accommodate bulky hydrophobic residues but may also form interactions with polar residues. Indeed, in our recent structure of the related Yme1 protease, we identified a hydrophobic patch in close proximity to the peptidase active site, which could potentially serve as such a binding site⁵¹. Some degree of preference is also observed at the P2 and P2' positions, consistent with studies on thermolysin and matrix metalloproteases identifying S2 binding pockets that contribute less significantly to cleavage preference^{65, 66}. High-resolution structural information is required to fully understand how the AFG3L2 captures substrates at the peptidase active sites. Assembly of membrane-embedded AAA+ proteases is required for activity. In both human and yeast i-AAA proteases, the transmembrane domains are needed to drive hexamerization^{42, 67}. However, the observed oligomerization of ^{core}AFG3L2^{E408Q} suggests that the catalytic domains can assemble in the absence of the transmembrane domains only when ATP is bound. Interestingly, the heterooligomeric yeast m-AAA protease, Yta10/Yta12, can form active assembled proteases when transmembrane domains are present in only one of the two subunits⁶⁸. These observations may indicate that m-AAA proteases contain greater subunit-subunit interactions between the ATPase and peptidase domains than i-AAA. A comparison of ATP hydrolysis activity between similarly reconstituted human m-AAA (^{cchex}AFG3L2) and i-AAA (^{cchex}YME1L) proteases reveal that the maximal rate for ^{cchex}AFG3L2 is 2-fold higher than ^{cchex}YME1L with a 30-fold lower $K_{0.5}$ culminating in a near 70-fold difference in catalytic efficiency. In the matrix, AFG3L2 competes with a variety of nucleotide associating proteins, such as additional AAA+ enzymes and ATP transporters. Conversely, YME1L has evolved in a subcompartment where ATP is not utilized as frequently thus reducing the need for tight nucleotide association. Indeed, fluctuations in ATP concentration within the intermembrane space have been proposed to modulate YME1L activity through its degradation by OMA1⁶⁹.

By successfully reconstituting active AFG3L2 homohexamers, we have established a system that can be used to study the mechanisms of ATP-driven degradation by a major isoform of the human m-AAA protease. In addition to providing a useful tool for structural biology, the

ability to parse the individual activities of ATP hydrolysis, translocation, peptide bond cleavage, and product release by AFG3L2 may shed light on the precise mechanistic defects caused by mutations of human neurodegenerative disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J. Guo (Stony Brook) for assistance with proteomic data analysis. We thank A.W. Karzai for helpful discussions and S. Smith (Stony Brook) for access to instrumentation. We thank R. Sauer (MIT) and N. Puri (Stony Brook) for generously providing plasmids and purified proteins.

Funding Sources

The authors declare no competing financial interests. This work was supported by NIH grant R01GM115898 (S.E.G.) and NIH/NCRR 1 S10 RR023680 (Proteomics Center).

ABBREVIATIONS

DTT	dithiothreitol
ETDA	ethylenediaminetetraacetic acid
GST	glutathione-S-transferase
PMSF	phenylmethylsulfonyl fluoride

REFERENCES

- [1]. Diaz F, and Moraes CT (2008) Mitochondrial biogenesis and turnover, *Cell Calcium* 44, 24–35. [PubMed: 18395251]
- [2]. Neupert W, and Herrmann JM (2007) Translocation of proteins into mitochondria, *Annu Rev Biochem* 76, 723–749. [PubMed: 17263664]
- [3]. Koppen M, and Langer T (2007) Protein degradation within mitochondria: versatile activities of AAA proteases and other peptidases, *Critical reviews in biochemistry and molecular biology* 42, 221–242. [PubMed: 17562452]
- [4]. Patron M, Sprenger HG, and Langer T (2018) m-AAA proteases, mitochondrial calcium homeostasis and neurodegeneration, *Cell Res* 28, 296–306. [PubMed: 29451229]
- [5]. Leonhard K, Herrmann JM, Stuart RA, Mannhaupt G, Neupert W, and Langer T (1996) AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria, *The EMBO journal* 15, 4218–4229. [PubMed: 8861950]
- [6]. Tatsuta T, Augustin S, Nolden M, Friedrichs B, and Langer T (2007) m-AAA protease-driven membrane dislocation allows intramembrane cleavage by rhomboid in mitochondria, *The EMBO journal* 26, 325–335. [PubMed: 17245427]
- [7]. Glynn SE (2017) Multifunctional Mitochondrial AAA Proteases, *Front Mol Biosci* 4, 34. [PubMed: 28589125]
- [8]. Arlt H, Tauer R, Feldmann H, Neupert W, and Langer T (1996) The YTA10–12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria, *Cell* 85, 875–885. [PubMed: 8681382]

- [9]. Banfi S, Bassi MT, Andolfi G, Marchitelli A, Zanotta S, Ballabio A, Casari G, and Franco B (1999) Identification and characterization of AFG3L2, a novel paraplegin-related gene, *Genomics* 59, 51–58. [PubMed: 10395799]
- [10]. Casari G, De Fusco M, Ciarmatori S, Zeviani M, Mora M, Fernandez P, De Michele G, Filla A, Coccozza S, Marconi R, Durr A, Fontaine B, and Ballabio A (1998) Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease, *Cell* 93, 973–983. [PubMed: 9635427]
- [11]. Augustin S, Gerdes F, Lee S, Tsai FT, Langer T, and Tatsuta T (2009) An intersubunit signaling network coordinates ATP hydrolysis by m-AAA proteases, *Molecular cell* 35, 574–585. [PubMed: 19748354]
- [12]. Suno R, Niwa H, Tsuchiya D, Zhang X, Yoshida M, and Morikawa K (2006) Structure of the whole cytosolic region of ATP-dependent protease FtsH, *Molecular cell* 22, 575–585. [PubMed: 16762831]
- [13]. Bieniossek C, Schalch T, Bumann M, Meister M, Meier R, and Baumann U (2006) The molecular architecture of the metalloprotease FtsH, *Proceedings of the National Academy of Sciences of the United States of America* 103, 3066–3071. [PubMed: 16484367]
- [14]. Koppen M, Metodiev MD, Casari G, Rugarli EI, and Langer T (2007) Variable and tissue-specific subunit composition of mitochondrial m-AAA protease complexes linked to hereditary spastic paraplegia, *Mol Cell Biol* 27, 758–767. [PubMed: 17101804]
- [15]. Maltecca F, Aghaie A, Schroeder DG, Cassina L, Taylor BA, Phillips SJ, Malaguti M, Previtali S, Guenet JL, Quattrini A, Cox GA, and Casari G (2008) The mitochondrial protease AFG3L2 is essential for axonal development, *J Neurosci* 28, 2827–2836. [PubMed: 18337413]
- [16]. Maltecca F, Magnoni R, Cerri F, Cox GA, Quattrini A, and Casari G (2009) Haploinsufficiency of AFG3L2, the gene responsible for spinocerebellar ataxia type 28, causes mitochondria-mediated Purkinje cell dark degeneration, *J Neurosci* 29, 9244–9254. [PubMed: 19625515]
- [17]. Almajan ER, Richter R, Paeger L, Martinelli P, Barth E, Decker T, Larsson NG, Kloppenburg P, Langer T, and Rugarli EI (2012) AFG3L2 supports mitochondrial protein synthesis and Purkinje cell survival, *J Clin Invest* 122, 4048–4058. [PubMed: 23041622]
- [18]. Kondadi AK, Wang S, Montagner S, Kladt N, Korwitz A, Martinelli P, Herholz D, Baker MJ, Schauss AC, Langer T, and Rugarli EI (2014) Loss of the m-AAA protease subunit AFG(3)L(2) causes mitochondrial transport defects and tau hyperphosphorylation, *The EMBO journal* 33, 1011–1026. [PubMed: 24681487]
- [19]. Di Bella D, Lazzaro F, Brusco A, Plumari M, Battaglia G, Pastore A, Finardi A, Cagnoli C, Tempia F, Frontali M, Veneziano L, Sacco T, Boda E, Brussino A, Bonn F, Castellotti B, Baratta S, Mariotti C, Gellera C, Fracasso V, Magri S, Langer T, Plevani P, Di Donato S, Muzi-Falconi M, and Taroni F (2010) Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28, *Nature genetics* 42, 313–321. [PubMed: 20208537]
- [20]. Mariotti C, Brusco A, Di Bella D, Cagnoli C, Seri M, Gellera C, Di Donato S, and Taroni F (2008) Spinocerebellar ataxia type 28: A novel autosomal dominant cerebellar ataxia characterized by slow progression and ophthalmoparesis, *Cerebellum* 7, 184–188. [PubMed: 18769991]
- [21]. Qu J, Wu CK, Zuzuarregui JR, and Hohler AD (2015) A novel AFG3L2 mutation in a Somalian patient with spinocerebellar ataxia type 28, *J Neurol Sci* 358, 530–531. [PubMed: 26454370]
- [22]. Lobbe AM, Kang JS, Hilker R, Hackstein H, Muller U, and Nolte D (2014) A novel missense mutation in AFG3L2 associated with late onset and slow progression of spinocerebellar ataxia type 28, *J Mol Neurosci* 52, 493–496. [PubMed: 24293060]
- [23]. Zuhlke C, Mikat B, Timmann D, Wiczorek D, Gillesen-Kaesbach G, and Burk K (2015) Spinocerebellar ataxia 28: a novel AFG3L2 mutation in a German family with young onset, slow progression and saccadic slowing, *Cerebellum Ataxias* 2, 19. [PubMed: 26677414]
- [24]. Svenstrup K, Nielsen TT, Aidt F, Rostgaard N, Duno M, Wibrand F, Vinther-Jensen T, Law I, Vissing J, Roos P, Hjermand LE, and Nielsen JE (2017) SCA28: Novel Mutation in the AFG3L2 Proteolytic Domain Causes a Mild Cerebellar Syndrome with Selective Type-1 Muscle Fiber Atrophy, *Cerebellum* 16, 62–67. [PubMed: 26868664]

- [25]. Cagnoli C, Stevanin G, Brussino A, Barberis M, Mancini C, Margolis RL, Holmes SE, Nobili M, Forlani S, Padovan S, Pappi P, Zaros C, Leber I, Ribai P, Pugliese L, Assalto C, Brice A, Migone N, Durr A, and Brusco A (2010) Missense mutations in the AFG3L2 proteolytic domain account for approximately 1.5% of European autosomal dominant cerebellar ataxias, *Human mutation* 31, 1117–1124. [PubMed: 20725928]
- [26]. Charif M, Roubertie A, Salime S, Mamouni S, Goizet C, Hamel CP, and Lenaers G (2015) A novel mutation of AFG3L2 might cause dominant optic atrophy in patients with mild intellectual disability, *Front Genet* 6, 311. [PubMed: 26539208]
- [27]. Colavito D, Maritan V, Suppiej A, Del Giudice E, Mazzarolo M, Miotto S, Farina S, Dalle Carbonare M, Piermarocchi S, and Leon A (2017) Non-syndromic isolated dominant optic atrophy caused by the p.R468C mutation in the AFG3 like matrix AAA peptidase subunit 2 gene, *Biomed Rep* 7, 451–454. [PubMed: 29181157]
- [28]. Edener U, Wollner J, Hehr U, Kohl Z, Schilling S, Kreuz F, Bauer P, Bernard V, Gillessen-Kaesbach G, and Zuhlke C (2010) Early onset and slow progression of SCA28, a rare dominant ataxia in a large four-generation family with a novel AFG3L2 mutation, *Eur J Hum Genet* 18, 965–968. [PubMed: 20354562]
- [29]. Atorino L, Silvestri L, Koppen M, Cassina L, Ballabio A, Marconi R, Langer T, and Casari G (2003) Loss of m-AAA protease in mitochondria causes complex I deficiency and increased sensitivity to oxidative stress in hereditary spastic paraplegia, *J Cell Biol* 163, 777–787. [PubMed: 14623864]
- [30]. Richter U, Lahtinen T, Martinen P, Suomi F, and Battersby BJ (2015) Quality control of mitochondrial protein synthesis is required for membrane integrity and cell fitness, *J Cell Biol* 211, 373–389. [PubMed: 26504172]
- [31]. König T, Troder SE, Bakka K, Korwitz A, Richter-Dennerlein R, Lampe PA, Patron M, Muhlmeister M, Guerrero-Castillo S, Brandt U, Decker T, Lauria I, Paggio A, Rizzuto R, Rugarli EI, De Stefani D, and Langer T (2016) The m-AAA Protease Associated with Neurodegeneration Limits MCU Activity in Mitochondria, *Molecular cell* 64, 148–162. [PubMed: 27642048]
- [32]. Grohmann L, Graack HR, Kruff V, Choli T, Goldschmidt-Reisin S, and Kitakawa M (1991) Extended N-terminal sequencing of proteins of the large ribosomal subunit from yeast mitochondria, *FEBS Lett* 284, 51–56. [PubMed: 2060626]
- [33]. Nolden M, Ehses S, Koppen M, Bernacchia A, Rugarli EI, and Langer T (2005) The m-AAA protease defective in hereditary spastic paraplegia controls ribosome assembly in mitochondria, *Cell* 123, 277–289. [PubMed: 16239145]
- [34]. Bonn F, Tatsuta T, Petrucci C, Riemer J, and Langer T (2011) Presequence-dependent folding ensures MrpL32 processing by the m-AAA protease in mitochondria, *The EMBO journal* 30, 2545–2556. [PubMed: 21610694]
- [35]. Baker TA, and Sauer RT (2006) ATP-dependent proteases of bacteria: recognition logic and operating principles, *Trends in biochemical sciences* 31, 647–653. [PubMed: 17074491]
- [36]. Rampello AJ, and Glynn SE (2017) Identification of a Degradation Signal Sequence within Substrates of the Mitochondrial i-AAA Protease, *J Mol Biol* 429, 873–885. [PubMed: 28214511]
- [37]. Gersch M, Stahl M, Poreba M, Dahmen M, Dziedzic A, Drag M, and Sieber SA (2016) Barrel-shaped ClpP Proteases Display Attenuated Cleavage Specificities, *ACS Chem Biol* 11, 389–399. [PubMed: 26606371]
- [38]. Maurizi MR (1987) Degradation in vitro of bacteriophage lambda N protein by Lon protease from *Escherichia coli*, *J Biol Chem* 262, 2696–2703. [PubMed: 2950089]
- [39]. Nishii W, Suzuki T, Nakada M, Kim YT, Muramatsu T, and Takahashi K (2005) Cleavage mechanism of ATP-dependent Lon protease toward ribosomal S2 protein, *FEBS Lett* 579, 6846–6850. [PubMed: 16337203]
- [40]. Nishii W, Maruyama T, Matsuoka R, Muramatsu T, and Takahashi K (2002) The unique sites in SulA protein preferentially cleaved by ATP-dependent Lon protease from *Escherichia coli*, *Eur J Biochem* 269, 451–457. [PubMed: 11856303]
- [41]. Orlowski M, and Wilk S (2000) Catalytic activities of the 20 S proteasome, a multicatalytic proteinase complex, *Arch Biochem Biophys* 383, 1–16. [PubMed: 11097171]

- [42]. Shi H, Rampello AJ, and Glynn SE (2016) Engineered AAA+ proteases reveal principles of proteolysis at the mitochondrial inner membrane, *Nat Commun* 7, 13301. [PubMed: 27786171]
- [43]. Iosefson O, Nager AR, Baker TA, and Sauer RT (2015) Coordinated gripping of substrate by subunits of a AAA+ proteolytic machine, *Nat Chem Biol* 11, 201–206. [PubMed: 25599533]
- [44]. Wohlever ML, Nager AR, Baker TA, and Sauer RT (2013) Engineering fluorescent protein substrates for the AAA+ Lon protease, *Protein engineering, design & selection : PEDS* 26, 299–305.
- [45]. Gur E, and Sauer RT (2008) Recognition of misfolded proteins by Lon, a AAA(+) protease, *Genes & development* 22, 2267–2277. [PubMed: 18708584]
- [46]. Puri N, and Karzai AW (2017) HspQ Functions as a Unique Specificity-Enhancing Factor for the AAA+ Lon Protease, *Molecular cell* 66, 672–683 e674. [PubMed: 28575662]
- [47]. Shevchenko A, Wilm M, Vorm O, Jensen ON, Podtelejnikov AV, Neubauer G, Shevchenko A, Mortensen P, and Mann M (1996) A strategy for identifying gel-separated proteins in sequence databases by MS alone, *Biochem Soc Trans* 24, 893–896. [PubMed: 8878870]
- [48]. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M, Pevzner PA, and Bafna V (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra, *Anal Chem* 77, 4626–4639. [PubMed: 16013882]
- [49]. Fuchs JE, von Grafenstein S, Huber RG, Margreiter MA, Spitzer GM, Wallnoefer HG, and Liedl KR (2013) Cleavage entropy as quantitative measure of protease specificity, *PLoS Comput Biol* 9, e1003007. [PubMed: 23637583]
- [50]. Bieniossek C, Niederhauser B, and Baumann UM (2009) The crystal structure of apo-FtsH reveals domain movements necessary for substrate unfolding and translocation, *Proceedings of the National Academy of Sciences of the United States of America* 106, 21579–21584. [PubMed: 19955424]
- [51]. Puchades C, Rampello AJ, Shin M, Giuliano CJ, Wiseman RL, Glynn SE, and Lander GC (2017) Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing, *Science* 358.
- [52]. Brown A, Rathore S, Kimanius D, Aibara S, Bai XC, Rorbach J, Amunts A, and Ramakrishnan V (2017) Structures of the human mitochondrial ribosome in native states of assembly, *Nat Struct Mol Biol* 24, 866–869. [PubMed: 28892042]
- [53]. Schechter I, and Berger A (1967) On the size of the active site in proteases. I. Papain, *Biochem Biophys Res Commun* 27, 157–162. [PubMed: 6035483]
- [54]. Tsai CW, Wu Y, Pao PC, Phillips CB, Williams C, Miller C, Ranaghan M, and Tsai MF (2017) Proteolytic control of the mitochondrial calcium uniporter complex, *Proceedings of the National Academy of Sciences of the United States of America* 114, 4388–4393. [PubMed: 28396416]
- [55]. Overall CM (2002) Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites, *Mol Biotechnol* 22, 51–86. [PubMed: 12353914]
- [56]. Fields GB, Van Wart HE, and Birkedal-Hansen H (1987) Sequence specificity of human skin fibroblast collagenase. Evidence for the role of collagen structure in determining the collagenase cleavage site, *J Biol Chem* 262, 6221–6226. [PubMed: 3032960]
- [57]. Netzel-Arnett S, Sang QX, Moore WG, Navre M, Birkedal-Hansen H, and Van Wart HE (1993) Comparative sequence specificities of human 72- and 92-kDa gelatinases (type IV collagenases) and PUMP (matrilysin), *Biochemistry* 32, 6427–6432. [PubMed: 8390857]
- [58]. Bogdanovic X, Palm GJ, Schwenteit J, Singh RK, Gudmundsdottir BK, and Hinrichs W (2016) Structural evidence of intramolecular propeptide inhibition of the aspzincin metalloendopeptidase AsaP1, *FEBS Lett* 590, 3280–3294. [PubMed: 27528449]
- [59]. Welch AR, Holman CM, Huber M, Brenner MC, Browner MF, and Van Wart HE (1996) Understanding the P1' specificity of the matrix metalloproteinases: effect of S1' pocket mutations in matrilysin and stromelysin-1, *Biochemistry* 35, 10103–10109. [PubMed: 8756473]
- [60]. Lovejoy B, Cleasby A, Hassell AM, Longley K, Luther MA, Weigl D, McGeehan G, McElroy AB, Drewry D, Lambert MH, and et al. (1994) Structure of the catalytic domain of fibroblast collagenase complexed with an inhibitor, *Science* 263, 375–377. [PubMed: 8278810]

- [61]. Stams T, Spurlino JC, Smith DL, Wahl RC, Ho TF, Qoronfleh MW, Banks TM, and Rubin B (1994) Structure of human neutrophil collagenase reveals large S1' specificity pocket, *Nat Struct Biol* 1, 119–123. [PubMed: 7656015]
- [62]. de Kreij A, van den Burg B, Veltman OR, Vriend G, Venema G, and Eijssink VG (2001) The effect of changing the hydrophobic S1' subsite of thermolysin-like proteases on substrate specificity, *Eur J Biochem* 268, 4985–4991. [PubMed: 11559368]
- [63]. Steitz TA, Henderson R, and Blow DM (1969) Structure of crystalline alpha-chymotrypsin. 3. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypsin, *J Mol Biol* 46, 337–348. [PubMed: 5360043]
- [64]. Hedstrom L (1996) Trypsin: a case study in the structural determinants of enzyme specificity, *Biol Chem* 377, 465–470. [PubMed: 8922280]
- [65]. Hangauer DG, Monzingo AF, and Matthews BW (1984) An interactive computer graphics study of thermolysin-catalyzed peptide cleavage and inhibition by N-carboxymethyl dipeptides, *Biochemistry* 23, 5730–5741. [PubMed: 6525336]
- [66]. Overall CM, and Kleifeld O (2006) Towards third generation matrix metalloproteinase inhibitors for cancer therapy, *Br J Cancer* 94, 941–946. [PubMed: 16538215]
- [67]. Leonhard K, Stiegler A, Neupert W, and Langer T (1999) Chaperone-like activity of the AAA domain of the yeast Yme1 AAA protease, *Nature* 398, 348–351. [PubMed: 10192337]
- [68]. Korbel D, Wurth S, Kaser M, and Langer T (2004) Membrane protein turnover by the m-AAA protease in mitochondria depends on the transmembrane domains of its subunits, *EMBO Rep* 5, 698–703. [PubMed: 15205678]
- [69]. Rainbolt TK, Saunders JM, and Wiseman RL (2015) YME1L degradation reduces mitochondrial proteolytic capacity during oxidative stress, *EMBO Rep* 16, 97–106. [PubMed: 25433032]

Figure 1. Designing an active soluble AFG3L2 protease.

(A) AFG3L2 contains a short N-terminal domain (N), dual transmembrane passes (1 and 2), an intermembrane space domain (IMS), AAA+ ATPase domain, M41 peptidase domain. Key catalytic residues are present in the Walker-A (WA), Walker-B (WB) and peptidase motifs (HExxH). (B) Rate of ATP hydrolysis against ATP concentration for cchex AFG3L2. Data were fit to the Hill version of the Michaelis-Menten equation ($k_{\text{ATPase}} = 91 \text{ min}^{-1} \text{ enz}_6^{-1}$; $K_{0.5} = 44.1 \text{ μM}$). (C) Degradation of the model substrate mutGFP-β20 by cchex AFG3L2 in the presence and absence of ATP and for cchex AFG3L2^{E408Q} in the presence of ATP. (D) Rate of mutGFP-β20 degradation against substrate concentration by cchex AFG3L2. Data were fit to the Michaelis-Menten equation ($k_{\text{deg}} = 0.28 \text{ min}^{-1} \text{ enz}_6^{-1}$; $K_M = 6.2 \text{ μM}$).

(A) Proteolysis of $^{30}\text{MrpL32}$ and $^{50}\text{MrpL32}$ by $^{\text{cchex}}$ AFG3L2. Loss of $^{30}\text{MrpL32}$ (p) occurs concurrently with accumulation of a smaller molecular weight fragment (m). (B) Loss of $^{30}\text{MrpL32}$ (p) over time and accumulation of the smaller molecular weight fragment (m). (C) Sequence alignment of MrpL32 from mammalian and yeast sources. Identical (orange) and highly conserved residues (yellow) are highlighted. N-termini of the accumulated fragment from proteolysis of $^{30}\text{MrpL32}$ (black arrow) and previously identified mature form of the yeast homolog (red arrow) are shown. (D) Initial rates of precursor loss for $^{30}\text{MrpL32}$ and $^{50}\text{MrpL32}$.

Figure 3. Residues within the MrpL32 pre-sequence can target proteins to AFG3L2.

(A) Schematic showing fragments of the MrpL32 N-terminus fused to the model unfolded protein I27^{CD}. Position of the N-terminus of the processed form of ³⁰MrpL32 is shown in red. (B) Initial degradation rates of fusion proteins bearing residues from MrpL32 appended to I27^{CD}. Statistical significances calculated using the Student's t-test are shown relative to I27^{CD} except where indicated. (C) Initial degradation rates of Tim9^{ΔN} and 40-50Tim9^{ΔN}. (D) Initial rates of precursor loss for ³⁰MrpL32, ³⁰MrpL32_{swap40-50}, and ³⁰MrpL32_{swap31-39}.

Figure 4. Identifying the AFG3L2 peptidase specificity profile.

(A) Schematic showing the mass spectrometry approach to determining cleavage site preferences in coreAFG3L2E408Q. (B) Cleavage site preferences of coreAFG3L2E408Q identified by LC/MS/MS from degradation of all eight substrates. Subpocket cleavage entropy values (S_i) are listed for each position. (C) Distribution of peptide product lengths identified by LC/MS/MS from degradation of all eight substrates.

Figure 5. Validation of the AFG3L2 peptidase specificity profile.

(A) Fluorogenic reporter peptides used in cleavage reactions. Representative data is shown from incubation of reporter peptides with ^{core}AFG3L2^{408Q} at 37 °C. (B) Plot showing initial rates for cleavage of reporter peptides by ^{core}AFG3L2^{408Q}. (C) Percentage of total peptides identified by LC/MS/MS resulting from cleavage between residues 14–15 of HspQ^{Y20} and HspQ^{Y20} variants bearing substitutions of Phe-15. (D) Percentage of total peptides identified by LC/MS/MS resulting from cleavage between Leu-13 and Tyr-14 of HspQ^{Y20} and HspQ^{Y20} variants bearing substitutions of Phe-15.