Comparative Analysis of Mitochondrial N-Termini from Mouse, Human, and Yeast*

Sarah E. Calvo‡§¶¶, Olivier Julien‡‡, Karl R. Clauser¶, Hongying Shen‡§, Kimberli J. Kamer‡§, James A. Wells**§§, and Vamsi K. Mootha‡¶¶

The majority of mitochondrial proteins are encoded in the nuclear genome, translated in the cytoplasm, and directed to the mitochondria by an N-terminal presequence that is cleaved upon import. Recently, N-proteome catalogs have been generated for mitochondria from yeast and from human U937 cells. Here, we applied the subtiligase method to determine N-termini for 327 proteins in mitochondria isolated from mouse liver and kidney. Comparative analysis between mitochondrial N-termini from mouse, human, and yeast proteins shows that whereas presequences are poorly conserved at the sequence level, other presequence properties are extremely conserved, including a length of ~20–60 amino acids, a net charge between +3 to +6, and the presence of stabilizing amino acids at the N-terminus of mature proteins that follow the N-end rule from bacteria. As in yeast, ~80% of mouse presequence cleavage sites match canonical motifs for three mitochondrial peptidases (MPP, Icp55, and Oct1), whereas the remainder do not match any known peptidase motifs. We show that mature mitochondrial proteins often exist with a spectrum of N-termini, consistent with a model of multiple cleavage events by MPP and Icp55. In addition to analysis of canonical targeting presequences, our N-terminal dataset allows the exploration of other cleavage events and provides support for polypeptide cleavage into two distinct enzymes (Hsd17b4), protein cleavages key for signaling (Oma1, Opa1, Htra2, Mavs, and Bcs2l13), and in several cases suggests novel protein isoforms (Scp2, Acadm, Adck3, Hsdl2, Dlst, and Ogdh). We present an integrated catalog of mammalian mitochondrial N-termini that can be used as a community resource to investigate individual proteins, to elucidate mechanisms of mammalian mitochondrial processing, and to allow researchers to engineer tags distally to the presequence cleavage. Molecular & Cellular Proteomics 16: 10.1074/mcp.M116.063818, 512–523, 2017.

Mitochondria are ancient bacterium-derived organelles essential for eukaryotic life. A double membrane divides mitochondria into distinct compartments (matrix, inner membrane, intermembrane space, and outer membrane) that carry out specialized cellular processes, including oxidative phosphorylation, iron-sulfur cluster biogenesis, and a myriad biosynthetic pathways. Mammalian mitochondria contain a tiny genome encoding 13 proteins, whereas the remaining ~1200 proteins (1, 2) are encoded in the nucleus and imported into the organelle. Although there are several different mechanisms that can target proteins to the mitochondrion, the predominant mechanism is via an N-terminal presequence that directs import through the double membrane, which is subsequently cleaved to produce the mature, functional protein (3).

To date, most of our understanding of mitochondrial protein import and processing has been elucidated using bakers’ yeast as a model system. As reviewed recently (3), the canonical import pathway involves an amphiphilic α-helical N-terminal sequence that directs import through the TOMM-TIMM translocase complex spanning both membranes. This presequence is typically cleaved in the mitochondrial matrix by the MPP1 (encoded by Mas1/Mas2 in yeast, Pmpca/Pmpcb in mouse, and PMPCA/PMPCB in human). Other matrix peptidases are known to subsequently cleave one amino acid (Icp55) or eight amino acids (Oct1). Cleavage by MPP typically occurs two amino acids C-terminal to an arginine (R-2); therefore, proteins cleaved by MPP + Icp55 typically have arginine at presequence position –3 (R-3), whereas proteins cleaved by MPP + Oct1 typically have arginine at position –10 (R-10). The inner membrane peptidase (Imp) is known to cleave a handful of proteins destined for the inner membrane or inter-
membrane space, although a specific motif is not known. Approximately 70% of yeast proteins contain an N-terminal presequence that upon import is rapidly degraded (3, 4). However, other mitochondrial import mechanisms exist that do not involve cleavage signals, especially for outer membrane or intermembrane space proteins.

Systematic identification of N-termini of all mitochondrial proteins promises to shed light on mitochondrial protein import/processing mechanisms. Two high-throughput studies have reported the mitochondrial N-proteomes in yeast and humans. In an elegant 2009 study, Vogtle et al. (4) isolated mitochondria from yeast and used combined fractional diagonal chromatography (COFRADIC) to specifically label and identify the N-termini. This method chemically modifies free N-termini of the internal peptides (after first protecting lysine side chains) and separates them from other peptides by chromatography. Then the N-terminal peptides are identified using mass spectrometry. This study identified N-termini for 615 yeast mitochondrial proteins, including 451 with cleaved presequences exceeding 5 aa. Yeast presequences typically are 20–60 aa long and have a net charge between +3 and +6. The study also showed that the N-termini of mature proteins typically begin with stabilizing amino acids such as serine and alanine, consistent with the N-end rule from bacteria (5). In 2015, Vaca Jacome et al. (6) reported the mitochondrial N-proteome in human U937 cells using trimethoxyphenyl phosphonium (TMPP) with C12 and C13 labeling. At exactly pH 8.2, TMPP chemically modifies free N-termini but not lysine side chains, and then the N-termini are determined from peptides with TMPP doublets containing both C12 and C13. This study identified N-termini for 356 human genes, 179 of which are in the MitoCarta2.0 catalog of mitochondrial proteins (2), whereas the remainder may be co-purifying contaminants. These two high throughput studies of mitochondrial N-termini are extremely useful resources to understand individual mitochondrial proteins as well as protein import mechanisms.

In this study, we applied the subtiligase method (7, 8) to systematically identify cleaved N-termini from mitochondria isolated from mouse tissues (Fig. 1). The Wells laboratory has pioneered this method for identifying N-termini and has successfully applied it to elucidate caspase proteolysis of hundreds of substrates (9, 10), identify circulating proteolytic signatures of chemotherapy-induced apoptosis (11, 12), and investigate evolutionary conservation of caspase substrates (13). Briefly, this method uses an engineered peptide ligase called subtiligase to covalently label free N-termini with an ester probe, including a biotin molecule for biochemical enrichment and an α-amino butyric acid (Abu) residue for unambiguous identification using LC-MS/MS. As subtiligase cannot label lysine side chains nor modified N-termini, this method is specific for processed N-termini as most non-processed N-termini are co-translationally acetylated in the cytoplasm prior to mitochondrial import. One known bias of this method is that subtiligase, like many proteases, has a preference for N-termini with small amino acids, specifically serine, alanine, and glycine, although ongoing studies are under way to improve this limitation (14). The subtiligase method uses positive (rather than negative) enrichment, unambiguous identification of free N-termini (Abu tag), and does not require modification of lysine residues prior to labeling. Thus, in this study we chose the subtiligase method to specifically identify mitochondrial N-termini cleaved after import.

Here, we report on mitochondrial protein N-termini encoded by 327 mouse genes, and we compare properties of mouse presequences to those previously reported in yeast and human. This catalog of mitochondrial N-termini across species provides insight into the evolutionary conservation of mitochondrial targeting mechanisms and provides a resource to understand individual mitochondrial proteins.
using TEV protease. The tryptic peptides were fractionated into 11 fractions/sample using high pH reverse phase C18 chromatography and then desalted with using C18 ZipTip (Millipore).

**Mass Spectrometry and Data Analysis—**LC/MS/MS was carried out by reverse phase LC interfaced on two different mass spectrometers. The first dataset (Velos) was acquired on an LTQ Orbitrap Velos (Thermo Fisher Scientific) mass spectrometer featuring a nanoflow HPLC (NanAccuity UPLC system, Waters) equipped with a trap column (180 μm × 20 mm, 5-μm SymmetryC18, from Waters) and an analytical column (100 μm × 100 mm, 1.7-μm BEH130C8, from Waters). Peptides were eluted over a linear gradient over 60 min from 2 to 30% acetonitrile in 0.1% formic acid. The second dataset (QExactive) was acquired on a QExactive Plus (Thermo Fisher Scientific) featuring a nanoflow HPLC (Dionex UltiMate 3000, Thermo Fisher Scientific) equipped with an analytical column (Acclaim PepMap RSLC, 75 μm × 150 mm, Thermo Fisher Scientific). One sample was analyzed from mouse kidney mitochondria (Velos with trypsin protease), and three samples were analyzed from mouse liver mitochondria (Velos with trypsin protease, QExactive with trypsin protease, and QExactive with trypsin and LysC proteases digested separately and then combined prior to injection). For data analysis, peptide sequences were assigned using the ProteinProspector (version 5.13.2) database search engine against the RefSeq mouse proteome database (release 63, 27,886 proteins) (15) retaining all homologs. Search parameters included a precursor mass tolerance of 20 ppm, fragment ion mass tolerance of 20 ppm (and 6 ppm for the QExactive data), up to two missed trypsin cleavages, constant amidomethylation of Cy5, variable modifications of N-terminal addition of Abu amino acid, acetylation of the protein N-terminus, and oxidation of methionine. The identified peptides were searched against a random decoy protein database for evaluating the false-positive rates. The false discovery rate never exceeded 2%. Only peptides containing the N-terminal Abu tag were analyzed (1081 unique peptides mapping to 474 genes). Peptides that mapped to multiple isoforms of the same Entrez gene were assigned to the longest matching RefSeq isoform. Peptides were excluded if they mapped to more than one Entrez gene (149 unique peptides) or if they mapped to non-MitoSeq2.0 genes (191 unique peptides), leading to a final dataset of 855 unique peptides supporting 470 unique cleavages in 327 MitoCarta2.0 genes. Data including peak lists are available at MS-Viewer with search keys jcxjsqyrhvl (liver, Velos), jxjxtrb92xk (kidney, Velos), yjvsp221wg (liver, QExactive with trypsin), and 4mvtyr7ig1 (liver, QExactive tryptic and LysC). Raw files are available in the MassIVE repository: MSV000082006, MSV000080207, MSV000080208, and MSV000080209.

**Mass Spectrometry Half-Tryptic Peptide Search of Published MitoCarta2.0 Spectra—**We reanalyzed existing mass spectra from our previous analysis of mitochondria isolated from 14 mouse tissues (1, 2). These 1.9 million high quality spectra from ion trap collision-induced dissociation LC-MS/MS generated using an LTQ-Orbitrap (Thermo Scientific, San Jose, CA) were recently searched against the mouse RefSeq protein database (release 63) using standard methods to identify all tryptic peptides or those matching the first peptide from the annotated protein start (727,613 matched spectra), as described in the MitoCarta2.0 resource (2). Here, we used Spectrum Mill version 6.0 prerelease (Agilent Technologies, Santa Clara, CA) to search the remaining 1.2 million previously unmatched spectra to identify half-tryptic peptides that had any residue at the N-terminus but lysine or arginine at the C-terminus. Spectrum Mill search parameters included electrospy ionization linear ion trap scoring parameters, mass tolerances of ±25 ppm for precursor and ±0.7 Da for product ions, 35% minimum matched peak intensity, trypsin nonspecific N-terminal enzyme specificity, fixed modification of carbamidomethylation at cysteine, and variable modifications of oxidized methionine with a precursor MH+ shift range of −18 to 64 Da. Spectrum Mill’s autovalidation module was used with a fixed score threshold of 9.0 and a minimum peptide length of 7 aa to produce a draft list of candidate peptide spectrum matches containing 16,471 distinct peptides with a 2.9% peptide level false discovery rate based on forward versus reverse matches. We retained spectra that matched proteins identified by mass spectrometry in MitoCarta2.0 and mapped these to 8554 unique protein cleavages. We then retained only peptides that occurred N-terminal to the first observed tryptic peptide within the published MitoCarta2.0 dataset (457 unique cleavages, supplemental Table S1) with an identification FDR of 2.6%. We further defined a high quality subset, defined as cleavages detected in >1 tissue with a peptide score ≥13 (supplemental Table S1, FDR <0.9%). Comparisons with the subtiligase N-proteome dataset (Fig. 1B) were made at the gene level, between high quality half-tryptic peptides observed in any of the 14 tissues or with the subset observed from analysis of liver mitochondria.

**Comparative Analyses of N-Termini—**The mouse mitochondrial N-terminus subtiligase dataset (327 genes, 470 cleavages) was used to create a master set (327 cleavage sites defined as those with the largest number supporting spectra per gene, using a number of half-tryptic spectra detected in 14 tissues to break ties) and a reference subset of the master set (119 cleavage sites with ≥4 supporting spectra). Multiple spectra supporting the same cleavage site may be derived from the following: (i) identical peptides detected in different experimental samples; (ii) different peptides starting at the same position (e.g. SLCHSDFR and SLCHSDFRK); and/or (iii) different modifications of the same peptide sequence. Human mitochondrial N-termini were downloaded from Vacca Jacome et al. (6). The original dataset (356 UniProt proteins and 425 cleavage sites) was filtered to exclude non-mitochondrial contaminants by retaining only proteins in human MitoCarta2.0 (179 proteins and 229 cleavage sites) and was then used to create a master set (179 cleavages defined as those with the largest number of supporting spectra per gene) and a reference subset (47 cleavage sites with ≥4 supporting spectra). Human UniProt proteins were mapped to NCBI Entrez genes via UniProt Knowledgebase (16) Mapper and mapped to mouse orthologs by best bidirectional hit (BlastP Expect <1e-3). Yeast mitochondrial N-termini were downloaded from Vogtle et al. (4) (615 genes, 1104 N-termini) along with tables containing the master set (279 cleavage sites) and a reference set with ≥6 supporting spectra (84 cleavage sites). Yeast proteins were mapped to 1:1 orthologs in mice (BlastP Expect <1e-3) as described previously (2). For cross-species comparisons, multiple alignments between orthologs were generated by MUSCLE (17), and cleavage sites from each species were mapped to coordinates within the multiple alignment. Note that all three N-proteome datasets contained multiple cleavage sites per gene. Overlaps between species (Fig. 1B and supplemental Fig. S4) were computed at the gene level (i.e. an exact match was counted if any cleavage site exactly matched any orthologous cleavage site).

**Predictions of Mitochondrial Targeting Sequences—**TargetP (18), TPredd3 (19), and MitoFates (20) were run on all MitoCarta2.0 proteins using default parameters. Comparisons with the mouse N-terminal subtiligase dataset were performed using the subtiligase master set (Fig. 3A). Very similar results were obtained if comparisons were performed at the gene level (i.e. an exact match was reported if any of the cleavage sites per gene were correctly predicted by the algorithm).

**Evolutionary Conservation—**Evolutionary conservation (Fig. 3B) was assessed using precomputed multiple alignments for orthologous groups defined by EggNOG (21) version 4.0. RefSeq proteins were mapped to Ensembl proteins (via Ensembl BioMart (22)), and the cleavage site for the subtiligase master set was mapped from RefSeq to Ensembl protein coordinates via MUSCLE (17) multiple alignment.
For each Ensembl protein, EggNOG multiple alignments were downloaded for all vertebrates (veNOG), metazoa (meNOG), and opisthokonts (opiNOG). Percent identity at each position in the precomputed multiple alignment was defined as the percent of homologs sharing the most commonly observed amino acid at that site. These per-
position metrics were averaged across all presequence positions (defined by the mouse cleavage site in the subtiligase master set) or across all remaining positions.

**Amino Acid Frequency**—The background frequency of all amino acids (Fig. 3F) was computed based on all MitoCarta2.0 mouse RefSeq proteins (mouse), all MitoCarta2.0 human RefSeq proteins (human), and all yeast proteins annotated as mitochondria in the Saccharomyces Genome Database (23) (yeast). The frequency of N-terminal amino acids (Fig. 3E) was based on analysis of the first amino acid after the cleavage site using the master sets from each of the three species, defined above.

**Cleavage Site Motifs**—Using the reference sets for each species, motifs were created for the sequences flanking the observed cleavage sites using WebLogo (24), excluding small sample correction. Subsets of the reference sets were used to investigate motifs for sequences matching R-2 (arginine in position 2 N-terminal to cleavage site), R-3 (arginine in the position 3 N-terminal to the cleavage site), or R-10 (arginine in the position 10 N-terminal to the cleavage site).

**Outlier Cleavages >100 Amino Acids**—To identify well supported cleavages >100 aa from the annotated RefSeq protein start, we considered all cleavages with ≥4 spectra (from subtiligase combined with half-tryptic peptides occurring anywhere in the protein, not just before the first tryptic peptide, combined with N-terminal spectra from human orthologs). We excluded the 10 most highly expressed liver proteins (based on MitoCarta2.0 liver proteomics) as these showed many cleavages throughout the length of the protein that were not conserved between mouse and human (proteins Cps, Apit5b, Glud1, Pcx, and Cat). We note that the reported cleavage site refers to cleavage prior to the listed position (e.g. NP_032318:313 refers to cleavage between amino acids 312 and 313). We searched for shorter isoforms by manually inspecting full-length mRNA and ESTs for mouse and human homologs using the UCSC genome browser (25). We searched for nearby translation initiation using existing ribosome profiling data (26) using the GWIPS-viz genome browser (27). For all 17 proteins with outlier cleavage events, we predicted PFAM protein domains (28) using HMMER3 (29) and transmembrane helices using TMHMM (30).

**Mitochondrial Subcompartment Analyses**—For supplemental Fig. S3, mouse MitoCarta2.0 proteins were assigned to subcompartments. Proteins were first classified as membrane (presence of transmembrane helix based on TMHMM (30) or NCBI gene ontology (GO) assignment to “mitochondrial inner membrane” or “mitochondrial membrane”) or soluble. Membrane proteins were classified as internal membrane (if detected in the matrix via APEX (31) or by GO annotation) or outer membrane (GO annotation) or otherwise unclassified. Soluble proteins were classified as matrix (if detected in the matrix via APEX (31) or by GO annotation) or intermembrane space (GO annotation or presence of Cx9C motifs (32)) or otherwise unclassified.

**RESULTS**

**Identification of Cleaved N-Termini from Mouse Mitochondria**—To identify mitochondrial prescence cleavage sites, we used the subtiligase method to label free N-termini in mitochondria from mouse tissues (Fig. 1A). Briefly, we isolated mitochondria from fresh mouse liver and kidney and then incubated protein extracts with purified subtiligase protein and an ester probe with the following three components: a biotin molecule for biochemical enrichment; a TEV site; and an Abu residue. Subtiligase covalently links the ester probe to free N-termini. Labeled proteins were selected using NeutrAvidin beads and then digested with either trypsin or LysC to
Comparative Analysis of Mitochondrial N-Termini

retain only the N-terminal peptide. The ester probe was cleaved at the TEV site, and the Abu-labeled N-terminal peptides were subjected to LC-MS/MS. Peptides were searched against the RefSeq mouse protein database and then filtered for unambiguous matches to MitoCarta2.0 mitochondrial proteins. Identified spectra were highly specific for mitochondrial proteins (supplemental Fig. S1). Only peptides containing the Abu signature were retained for downstream analysis, enabling unambiguous identification of N-termini. In total, three technical replicates were performed from the liver mitochondrial sample (two using trypsin and one using LysC protein digestion), and one replicate was performed from the kidney mitochondrial sample (trypsin protein digestion).

These experiments yielded identification of N-termini for 327 mitochondrial proteins in liver and 158 mitochondrial proteins in kidney, which were a strict subset of those identified in liver (Fig. 1A and supplemental Table S1). This set represents 28% of all MitoCarta proteins, 46% of MitoCarta proteins expressed in the liver, and 58% of mitochondrial matrix proteins expressed in the liver. The 327 identified proteins exhibit a bias toward abundant proteins, consistent with most MS studies (supplemental Fig. S2). As expected, there was strong enrichment for identification of matrix proteins (35% enrichment, \( p = 2e-10 \), Fisher’s Exact test, see supplemental Fig. S3A). Although many methods are biased against membrane proteins, surprisingly, we did not observe a depletion of known inner membrane proteins (supplemental Fig. S3A). As expected, there is depletion of outer membrane proteins that typically are not cleaved after mitochondrial import (33), as these are typically N-acetylated and thus not detectable via subtiligase (supplemental Fig. S3A). Our inventory of 327 mitochondrial genes with the cleaved N-termini identified is more comprehensive than the published N-proteome study in humans (179 genes) and thus substantially improves knowledge of cleavage sites in mammalian mitochondria.

Assessing Sensitivity and Precision—We applied three complementary approaches to assess the quality of the subtiligase N-termini.

First, we compared our dataset to a control set of experimentally determined mitochondrial N-terminal cleavage sites from the literature. The control set included 15 mouse N-termini and 51 rat N-termini, typically determined via Edman degradation, based on literature curation within the UniProt knowledgebase (16) (Table S2). We mapped rat proteins to mouse orthologs via best bidirectional hits (expect \(<1e-3\)) and mapped orthogonal cleavage sites via MUSCLE protein alignments. Our subtiligase dataset contained cleavage sites for 9/15 control mouse proteins and 32/51 rat proteins. Of these, concordant cleavage sites were observed in 7/9 mouse proteins and 29/32 rat proteins (within 2 aa in multiple alignment). Based on this control set, the subtiligase dataset shows 55% sensitivity (36/66) and 88% precision (36/41 matches out of all genes present in both sets). The limited sensitivity can largely be explained by control proteins not expressed in the selected tissues (33%) and by the lack of amenable N-terminal tryptic peptides at the control N-termini (67%) (see Table S2). The five non-concordant N-termini can all be explained by lack of amenable tryptic peptides at the control N-termini (Table S2). Together the control data suggests 55% sensitivity and 88% precision of cleavage site identification.

Second, we compared our subtiligase N-termini to high throughput N-terminal datasets from human (6) and yeast (4) mitochondria (Fig. 1B). Orthologous proteins were mapped between species, and orthologous cleavage sites were identified within MUSCLE protein alignments (supplemental Fig. S4). Of the 116 orthologs with N-termini detected in mouse and human studies, 84% had concordant cleavage sites, including 77% with exact matches and 7% within 2 aa in the multiple alignment (thus may represent alignment artifacts). In contrast, there was very little overlap in cleavage site positions between mouse and yeast orthologs (29% within 2 aa in alignment), likely due to poor conservation of the presequences across the ~400 million years of evolution between these species and the resultant poor protein alignments. The high overlap in experimentally determined N-termini between mammalian species validates the quality of both the subtiligase and TMPP datasets in the mouse and human (6), respectively.

Third, we compared N-termini from our subtiligase dataset to potential cleavage sites that we mined from existing LC-MS/MS spectra of mitochondrial proteins from 14 mouse tissues (Fig. 1C and supplemental Fig. S5). Tissue datasets are particularly useful to increase sensitivity, because although ~50% of mitochondrial proteins are ubiquitously expressed, mRNA and proteomics data show marked tissue specificity for the remainder (e.g. ketogenesis, urea cycle, and steroidogenesis pathways) (1, 34). We previously isolated mitochondria from 14 mouse tissues, digested proteins using trypsin, performed LC-MS/MS, and identified mitochondrial proteins using typical database search criteria that matched tryptic peptides (or half-tryptic peptides beginning at the annotated protein start). We previously published these data as part of the MitoCarta (1) and MitoCarta2.0 (2) catalogs, which include MS/MS support for 1008 mitochondrial proteins and 712 specifically in liver mitochondria. This previous search would fail to detect N-terminal peptides from cleaved proteins, as they would not match the \textit{in silico} tryptic peptide database. Therefore, we now mined the unmatched MS/MS spectra to identify high confidence half-tryptic peptides with lysine or arginine at the C-terminus. To identify N-terminal half-tryptic peptides, we further required that half-tryptic peptides occur N-terminal to any tryptic peptide identified in the original MitoCarta2.0 search (see under “Experimental Procedures”) and filtered for well supported cleavage sites, yielding 216 unique cleavage sites in 137 mitochondrial genes. Finally, we compared the MitoCarta half-tryptic N-terminal set to the
Multiple Cleavage Sites per Protein—Our subtiligase N-terminal dataset, like the published datasets in human and yeast, detects multiple cleavage sites for some proteins—which are likely a mix of technical artifacts and true biological complexity. We detected a total of 704 cleavage sites in 327 genes, and computational predictions using the following three algorithms: TargetP (18), TPpred3 (19), and MitoFates (20).

Multiple Cleavage Sites per Protein—Our subtiligase N-terminal dataset, like the published datasets in human and yeast, detects multiple cleavage sites for some proteins—which are likely a mix of technical artifacts and true biological complexity. We detected a total of 704 cleavage sites in 327 genes, and computational predictions using the following three algorithms: TargetP (18), TPpred3 (19), and MitoFates (20).

Data Integration—We integrated data from all the above methods to create a resource of N-termini for mammalian mitochondrial proteins. This supplemental Table S1 includes our mouse N-termini detected by subtiligase, the cleavage sites detected by half-tryptic peptides across 14 mouse tissues, human N-termini from the Vaca Jacome et al. study (6), and computational predictions using the following three algorithms: TargetP (18), TPpred3 (19), and MitoFates (20).

Multiple Cleavage Sites per Protein—Our subtiligase N-terminal dataset, like the published datasets in human and yeast, detects multiple cleavage sites for some proteins—which are likely a mix of technical artifacts and true biological complexity. We detected a total of 704 cleavage sites in 327 genes, and computational predictions using the following three algorithms: TargetP (18), TPpred3 (19), and MitoFates (20).

We highlight four examples to illustrate single and multiple cleavages in our subtiligase dataset (Fig. 2). Clybl is a protein with a single cleavage observed by the subtiligase method that is also supported by MitoCarta2.0 half-tryptic peptides and by an orthologous cleavage site in humans (6). This cleavage site is predicted by one of three prediction algorithms (TPpred3), and it matches the canonical cleavage site for MPP (R-2). In contrast, we detected four potential cleavage sites for Cox5a, all of which were also supported by MitoCarta2.0 half-tryptic peptides. In both datasets, the most abundant peptide indicates cleavage at position 38 (5 subtiligase spectra, 320 MitoCarta spectra), which matches the canonical MPP+Icp55 cleavage site (R-3). However, there is also support for the MPP cleavage (position 37) as well as two downstream sites (position 40 and 41). Similarly, for protein Akr7a5, both subtiligase and MitoCarta2.0 half-tryptic peptides support cleavage at position 29 consistent with MPP+Icp55 cleavage; however, three additional cleavages are detected (position 21, 28, and 39), all consistent with MPP cleavage. The protein Uqcrh has three potential cleavage sites, none of which match canonical MPP, Icp55, or Oct1 cleavage sites. These examples show broad consistency between the most abundant cleavages detected in two independent mouse N-terminal datasets and may hint that some proteins exist with a spectrum of different N-termini.

To assess whether the observed N-terminal ladders (e.g., Cox5a) are likely to represent true biological complexity versus sample preparation artifact, we sought to determine whether such ladders were enriched in mitochondrial versus non-mitochondrial proteins (supplemental Fig. S6). We compared prevalence of cleavage ladders in mitochondrial proteins versus non-mitochondrial contaminants in our study as well as in mitochondrial studies in human (6) and yeast (4). We additionally analyzed N-terminomics data by the subtiligase method from human blood plasma (36) as well as six cell lines without enrichment for mitochondrial proteins compiled in DegraBase1.0 (14). In each study, ladders were enriched in mitochondria compared with non-mitochondrial proteins (supplemental Fig. S6). Mitochondrial proteins showed the highest prevalence of ladders, followed by secreted proteins in the human plasma, whereas cytoplasmic proteins showed substantially fewer ladders (supplemental Fig. S6A). These data, generated across different platforms and species, together show consistent enrichment of consecutive cleavages in mitochondrial proteins. Because ladders are derived from exoproteolysis, it suggests that more exoproteolysis occurs in the mitochondria relative to the cytosol, consistent with known exoproteolysis via Icp55/Xnpep3. Overall, our data...
indicate that many mitochondrial proteins exist with a spectrum of different N-termini in vivo.

Properties of Presequences across Species—We next explored properties of mitochondrial presequences in our mouse subtiligase dataset compared with published data in human (6) and yeast (4) (Fig. 3). The presequence is computationally defined as the sequence between a protein’s annotated start and the observed cleavage site from the master set. We note that experimental datasets are crucial for this comparison given the low accuracy of existing presequence prediction algorithms (Fig. 3A), as has been previously reported (4). We find that mouse presequences are only half as conserved as the mature protein sequences, based on analysis of precomputed multiple alignments of orthologous proteins across different taxonomic groups in the EggNOG database (21) (Fig. 3B). Despite the low conservation of presequences at the sequence level, other presequence properties are extremely conserved across evolution, including length (Fig. 3C) and net charge (Fig. 3D). In all three species, presequences are typically 20–60 aa long and have net charge ranging from +3 to +6. Consistent with the Vogtle et al. (4) study from yeast, all three species show a preference for stable amino acids at the N-terminus of mature proteins, following the N-end rule from bacteria (Fig. 3, E and F). We note that the mouse subtiligase dataset shows a particular bias for N-terminal serine and glycine, possibly due to subtiligase’s preferential labeling of proteins beginning with these specific residues (14); however, we note that very similar frequencies are observed in the mouse half-tryptic peptide dataset.

In summary, experimentally determined N-termini from mouse, human, and yeast mitochondrial proteins show that whereas mitochondrial targeting signals are not well conserved at the sequence level, other properties show striking conservation, including length, net charge, and presence of stabilizing amino acids at the N-terminus of mature proteins.

Cleavage Site Recognition Motifs—Next, we aimed to assess whether mammalian mitochondrial cleavage sites have the same recognition motifs as those determined previously in yeast studies (3, 4). In yeast presequences, the predominant cleavage site motif is the presence of arginine upstream of the cleavage site at position R-2 (MPP), R-3 (MPP/Icp55), or R-10 (MPP/Oct1) (3, 4, 20). As in the published yeast study (4), we compiled a reference set of well supported cleavage sites in mouse and human (see under “Experimental Procedures”). Analysis of the position of the first arginine relative to the cleavage site (Fig. 4A) shows clear enrichment for the R-2, R-3, and R-10 motifs in all three species. However, in all three species, 15–25% of cleavage sites do not match any of these motifs (Fig. 4B) pointing to additional peptidases yet to be determined or pleiotropy of existing peptidases. Alignment of all cleavage sites showed broadly consistent patterns in each species, albeit with low information content (Fig. 4C, top row). Therefore, we also separately aligned cleavage sites matching R-2, R-3, or R-10 (Fig. 4C) and again observe consistency of recognition motifs across species. We note that the R-10 motif shows the least conservation between yeast and mice (Fig. 4C); specifically the mouse motif shows a preference for serine at position P4 and lacks the phenylalanine preference...
at position $P_8$, consistent with a recent study using a synthetic peptide library for human MIPEP substrates (37).

We then asked whether the substrates of secondary peptidases Icp55 and Oct1 are also conserved across different species (Fig. 4, D and E). Yeast secondary peptidases Icp55 and Oct1 have mouse orthologs, Xpnpep3 and Mipep, respectively; however their substrates are not characterized in the mouse. We compiled the set of experimentally determined targets of Icp55 and Oct1 in yeast and identified the subset having 1:1 mouse orthologs with N-termini determined via subtiligase. Of the 13 yeast substrates of Icp55, 77% of mouse orthologs shared the same R-3 motif, supporting that these substrates are indeed strongly conserved across evolution (Fig. 4D). In striking contrast, 0/9 mouse orthologs to yeast Oct1 substrates had the canonical R-10 motif and instead were consistent with MPP or Icp55/Xpnpep3 cleavage (Fig. 4E). Thus, although the mouse data clearly show an R-10 motif, presumably the result of Mipep cleavage, the sub-

**Fig. 4. Comparative analysis of cleavage site motifs.** A, histogram of position of the first arginine relative to the reported cleavage site from the stringent reference set of N-termini from this study (mouse, $n = 119$, Vaca Jacome et al. (6); human, $n = 47$, and Vogtle et al. (4); yeast, $n = 94$). B, barplot shows percent of all reference set N-termini that have an arginine at position $-2$, $-3$, $-10$, or none. Note: a cleavage site may be counted in multiple categories. C, cleavage site motifs generated by WebLogo (24) are shown for all reference set N-termini (1st row), the subset with the first arginine at position $-2$ (2nd row), the subset with first arginine at position $-3$ and not $-2$ (3rd row), and the subset excluding R-2 and R-3 (4th row), and the subset with arginine at position $-10$ (5th row). D, shown are all known targets of Icp55 in yeast (3) that have orthologous mouse N-termini detected in this study. Gene names are shown with the 10 aa preceding the cleavage site in yeast ($\text{red text}$ highlights known motifs for MPP and Icp55). E, shown are all known targets of Oct1 in yeast (4) that have orthologous mouse N-termini detected in this study. Gene names are shown with the 10 aa preceding the cleavage site (residues in $\text{red}$ highlights the canonical motifs for Oct1 in yeast and for MPP and Icp55 in mouse).
strates of yeast Oct1 and mouse Mipep are likely to be quite distinct.

Given that the substrates of Oct1/Mipep are not conserved between yeast and mouse, we next sought to assess whether the putative Mipep cleavage sites (R-10) were conserved across animals. Thus, we selected 12 metazoan species and aligned orthologs of all mouse proteins with observed R-10 cleavage sites. We observed the mouse P10 arginine is typically well conserved across animals in 11/12 cases, with one exception (Prodh) (supplemental Fig. S7). Moreover, in all five instances where human orthologs were present in the Vaca Jacome study (6), the human N-termini showed the same P10 arginine. Together, these comparative data suggest that whereas Oct1/Mipep substrates are not conserved between yeast and mouse (Fig. 4E), within animals the substrates appear well conserved (supplemental Fig. S7).

Investigation of Outlier Cleavage sites—Although most observed N-termini are consistent with a cleaved presequence 20–60 aa long, we observed 17 proteins with well supported cleavage sites >100 aa from the annotated RefSeq protein start (Fig. 5) (see “Experimental Procedures”). We note that such distant cleavages would not have been identified in the yeast study, which mapped mass spectra to a database consisting only of the first 100 aa of all mitochondrial proteins. Investigation of these 17 proteins revealed four with existing experimental support for the orthologous cleavage site in other animals, including known cleavage of a polyprotein into two functional enzymes (Hsd17b4/NP_032318:313) (38), and three sensor proteases that are cleaved and activated in response to stress conditions (Htra2/NP_062726:134 (39), Oma1/NP_080185:140 (40), and Opa1/NP_598513:195 (41)). We additionally observe a novel cleavage site in Opa1 with evidence in liver, cerebellum, and spinal cord based on half-trypptic peptides (Opa1/NP_598513:184). For two mitochondrial outer membrane proteins known to be cleaved by cytosolic proteases (42–44), we report the first evidence of exact
cleavage sites in mouse (Mavs/NP_001193314:330 and Bcl2l13/NP_705736:327). These examples demonstrate the utility of our outlier N-termini to highlight protein cleavages important for mitochondrial pathways and stress response.

Among the remaining outlier cleavage events, we identified three genes with evidence of a shorter mRNA isofrom not present in RefSeq but consistent with our observed N-termini (Acadm, Adck3/Coq8a, and Scp2; see supplemental Fig. S8). Such novel isoforms are further supported by nearby translation start sites detected by ribosome profiling experiments in mouse liver that used lactamidomycin to halt translation at the initiating codon (supplemental Fig. S8) (26).

Strikingly, we observed internal cleavages in two subunits of a key Krebs cycle enzyme complex, 2-oxoglutarate dehydrogenase, Ogdh (NP_001239211:112) and Dist (NP_084501:202). Furthermore, we observed experimental support for a novel protein isoform of DSLT in the Human Protein Atlas (45), where Western blottings of liver extracts using a C-terminal targeting antibody show the following two bands: a 49-kDa band consistent with the full-length protein and an ~26-kDa band consistent with a shorter isoform produced by the observed N-terminus. Thus existing antibody validation experiments are consistent with one of our novel protein isoforms in liver.

Together, these 17 outlier cleavages highlight known polypeptide and regulated cleavages, novel mRNA isoforms, and suggest new protein isoforms for key metabolic enzymes. Further experiments are required to validate all these results and to decipher roles of shorter protein isoforms in mitochondrial pathways.

DISCUSSION

Here, we present the largest experimental catalog of N-termini for cleaved mammalian mitochondrial proteins. Our evolutionary analysis comparing cleavages between 327 mouse genes to published data from 179 human genes and 451 yeast genes shows that although mitochondrial precursors are poorly conserved at the sequence level, properties such as length and charge are extremely well conserved, as is the presence of stabilizing N-terminal amino acids consistent with the N-end rule (5). Mammalian precursors contain highly similar cleavage site motifs to those known in yeast, namely R-2 (MPP), R-3 (MPP+Lcp55), and R-10 (MPP+Oct1); however, 15–25% of cleavage sites do not match any known peptidase motif. By integrating experimental data from mouse and human, as well as targeting signal predictions by three algorithms, we created a resource to enable investigations into individual mitochondrial proteins or protein classes (supplemental Table S1).

We note that our inventory of mitochondrial cleavage sites has several distinctions from similar resources determined using different technologies in human and yeast. Most importantly, our subtiligase method cannot label acetylated N-termini and thus our dataset is nearly specific for cleaved mitochondrial proteins. Indeed, our master set of 327 mouse N-termini contain only six uncleaved RefSeq proteins (beginning with methionine at position 1), which may represent un-acetylated N-termini. In contrast, the COFRADIC methodology applied to yeast mitochondria (4) identified both uncleaved and cleaved N-termini, and these authors estimated 70% of mitochondrial proteins contain a cleaved precursor exceeding 5 aa. Both subtiligase and COFRADIC methods include a step to enrich either N-terminally labeled proteins or N-terminal peptides (positive enrichment for subtiligase and negative enrichment for COFRADIC), and this enrichment avoids interference from the unlabeled proteins or internal peptides and increases sensitivity for detecting N-termini by LC-MS. In comparison, the TMPP labeling used in the human study lacks an enrichment step, which might have contributed to the lower coverage (179 human MitoCarta proteins detected). We note that the subtiligase method utilizes an enzyme-based modification compared with the other two chemical modifications. Finally, as our analysis pipeline did not limit peptide identification to the first 100 aa of proteins, we detect 17 outlier cleavages >100 aa from the annotated protein start that confirm existing knowledge of polyprotein and signaling cleavage events, as well as identify new protein isoforms for key metabolic enzymes.

We estimate our dataset has 55% sensitivity for detecting true cleaved mitochondrial proteins. Increased sensitivity could be obtained in future studies by increasing the number of mouse tissues analyzed, which we estimate could identify an additional 33% of cleaved proteins. Moreover, additional proteases could be used to identify N-termini that lack amenable tryptic peptides. Although subtiligase methods are not compatible with Glu-C (which would cleave the TEV site), future studies could increase sensitivity using Asp-N or chymotrypsin proteases (which we estimate would each identify ~60 additional proteins (supplemental Fig. S9)). We note that both subtiligase and other technologies have difficulty detecting low abundance proteins, but the subtiligase approach improves the chance of detecting low abundance proteins due to positive enrichment for N-termini (14).

Together, our experimental catalog of mouse mitochondrial N-termini, supporting evidence from half-tryptic peptides across 14 mouse tissues, and orthologous N-termini from published human data provides the most comprehensive catalog of mammalian cleavage sites. These data can help researchers studying individual proteins to determine where the functional processed protein begins, which can be important for protein biochemistry or in engineering N-terminal selectable tags distal to cleavage sites. Additionally, these data can help elucidate mechanisms of mitochondrial protein import and processing and identify regulated cleavages such as those we observed in known sensors (Htra2, Oma1, Opal, Mavs, and Bcs2l13). Given the large fraction of cleaved proteins not matching known proteases, there are likely additional biological processing pathways yet to be elucidated.
Acknowledgment—We thank C. Savojardo for providing TPpred3 predictions on all human and mouse proteins.

* This work was supported by National Institutes of Health Grants R01 GM0077465 (to V. K. M.), R01 GM081051 (to J. A. W.), R01 GM097316 (to J. A. W.), R01 CA154802 (to J. A. W.), and F32 GM114905 (to H. S). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. [S] This article contains supplemental material.

To whom correspondence should be addressed. E-mail: scalvo@broadinstitute.org.

‡‡ Recipient of a Banting Postdoctoral Fellowship funded by the Canadian Institutes of Health Research and the Government of Canada and a fellowship from the UCSF Program for Breakthrough Biomedical Research, which is funded in part by the Sandler Foundation.

¶¶ Investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A., and Mootha, V. K. (2008) A mitochondrial protein compendium elucidates complex I disease biology. Cell 134, 112–123

2. Calvo, S. E., Clauser, K. R., and Mootha, V. K. (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44, D1251–D1257

3. Mossmann, D., Meisinger, C., and Vogtle, F. N. (2012) Processing of mitochondrial presequences. Biochim. Biophys. Acta 1819, 1098–1106

4. Vogt, F. N., Woltrekamp, S., Zahedi, R. P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J. V., Voos, W., Sickmann, A., Pfanner, N., and Meisinger, C. (2008) Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. Cell 139, 428–439

5. Tasaki, T., Siriram, S. M., Park, K. S., and Kwon, Y. T. (2012) The N-end rule pathway. Annu. Rev. Biochem. 81, 261–289

6. Vaca Jacome, A. S., Rabilloud, T., Schaeffer-Reiss, C., Rompais, M., Ayoub, D., Lane, L., Bairoch, A., Van Dorssevelaer, A., and Carapito, C. (2015) N-termine analysis of the human mitochondrial proteome. Proteomics 15, 2519–2524

7. Mehrus, S., Trinidad, J. C., Barkan, D. T., Sali, A., Burlinge, A. L., and Wells, J. A. (2008) Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N-termini. Cell 134, 866–876

8. Witta, A. P., Seaman, J. E., and Wells, J. A. (2014) Global analysis of cellular proteolysis by selective enzymatic labeling of protein N-terminals. Methods Enzymol. 544, 327–358

9. Julien, O., Zhuang, M., Wiita, A. P., O’Donoghue, A. J., Knudsen, G. M., and Wells, J. A. (2016) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol. Cell. Proteomics 14, 1113–1126

10. Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldon, T., Rattie, T., Creevey, C., Kuhn, M., Jensen, L. J., von Mering, C., and Bork, P. (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res. 42, D231–D239

11. Cunningham, F., Amode, M. R., Barrett, D., Beal, K., Bills, B., Saim, S., Xenarios, I., Lalev, A., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Grön, C. G., Gordon, L., Hourlier, T., Hunt, S. E., et al. (2015) Ensembl 2015. Nucleic Acids Res. 43, D662–D669

12. Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Kornelák, K., Krieger, et al. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 40, D700–D705

13. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. Genome Res. 14, 1188–1190

14. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002) The human genome browser at UCSC. Genome Res. 12, 996–1006

15. Mao, X., Lan, J., Liu, B., Ma, M., Shen, B., and Qian, S. B. (2015) Quantitative profiling of initiating ribosomes in vivo. Nat. Methods 12, 147–153

16. Michel, A. M., Ahern, A. M., Donohue, C. A., and Baranov, P. V. (2015) GWIPS-viz as a tool for exploring ribosome profiling evidence supporting the synthesis of alternative proteoforms. Proteomics 15, 2410–2416

17. Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Quek, S., Sagrador-Vega, A., Salazar, G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285

18. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41, e121

19. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580

20. Rhee, H. W., Zou, P., Udesi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., and Ting, A. Y. (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science 339, 1328–1331

21. Gabriel, K., Milenkovic, D., Chacinska, A., Muller, J., Guiard, B., Pfanner, N., and Meisinger, C. (2007) Novel mitochondrial intermembrane space proteins as substrates of the MIA import pathway. J. Mol. Biol. 365, 612–620

22. Dukanovic, J., and Rapaport, D. (2011) Multiple pathways in the integration of proteins into the mitochondrial outer membrane. Biochim. Biophys. Acta 1808, 971–980

23. Mootha, V. K., Bunkenborg, J., Olsen, J. V., Hjerrild, M., Wisniewski, J. R., Stahl, E., Beloumi, M. S., Ray, H. N., Shihag, S., Kamal, M., Patterson, N., Lander, E. S., and Mando, M. (2003) Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. Cell 115, 629–640

24. Timmer, J. C., Enoksson, M., Wildfang, E., Zhu, W., Igarashi, Y., Denault, J. B., Ma, Y., Dummitt, B., Chang, Y. H., Mast, A. E., Eroshkin, A., Smith, J. W., Tao, W. A., and Salvesen, G. S. (2007) Profiling constitutive proteolytic events in vivo. Biochem. J. 407, 41–48
36. Wildes, D., and Wells, J. A. (2010) Sampling the N-terminal proteome of human blood. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4561–4566

37. Marcondes, M. F., Alves, F. M., Assis, D. M., Hirata, I. Y., Juliano, L., Oliveira, V., and Juliano, M. A. (2015) Substrate specificity of mitochondrial intermediate peptidase analysed by a support-bound peptide library. *FEBS Open Bio* **5**, 429–436

38. Leenders, F., Husen, B., Thole, H. H., and Adamski, J. (1994) The sequence of porcine 80-kDa 17β-estradiol dehydrogenase reveals similarities to the short chain alcohol dehydrogenase family, to actin binding motifs and to sterol carrier protein 2. *Mol. Cell. Endocrinol.* **104**, 127–131

39. Suzuki, Y., Imai, Y., Nakayama, H., Takahashi, K., Takio, K., and Takahashi, R. (2001) A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. *Mol. Cell* **8**, 613–621

40. Baker, M. J., Lampe, P. A., Stojanovski, D., Korwitz, A., Anand, R., Tatsuta, T., and Langer, T. (2014) Stress-induced OMA1 activation and autocatalytic turnover regulate OPA1-dependent mitochondrial dynamics. *EMBO J.* **33**, 578–593

41. Ishihara, N., Fujita, Y., Oka, T., and Mihara, K. (2006) Regulation of mitochondrial morphology through proteolytic cleavage of OPA1. *EMBO J.* **25**, 2966–2977

42. Li, X. D., Sun, L., Seth, R. B., Pineda, G., and Chen, Z. J. (2005) Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17717–17722

43. Yang, Y., Liang, Y., Ou, L., Chen, Z., Yi, M., Li, K., and Lemon, S. M. (2007) Disruption of innate immunity due to mitochondrial targeting of a picornaviral protease precursor. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7253–7258

44. Julien, O., Zhuang, M., Wiita, A. P., O'Donoghue, A. J., Knudsen, G. M., Craik, C. S., and Wells, J. A. (2016) Quantitative MS-based enzymology of caspases reveals distinct protein substrate specificities, hierarchies, and cellular roles. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E2001–E2010

45. Uhlen, M., Fagerberg, L., Hallström, B. M., Lundberg, J., Copley, R. R., Petersen, O., Zwahlen, S., Steen, H., Liljegren, S., von Mering, C., and Ponting, C. (2015) Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419