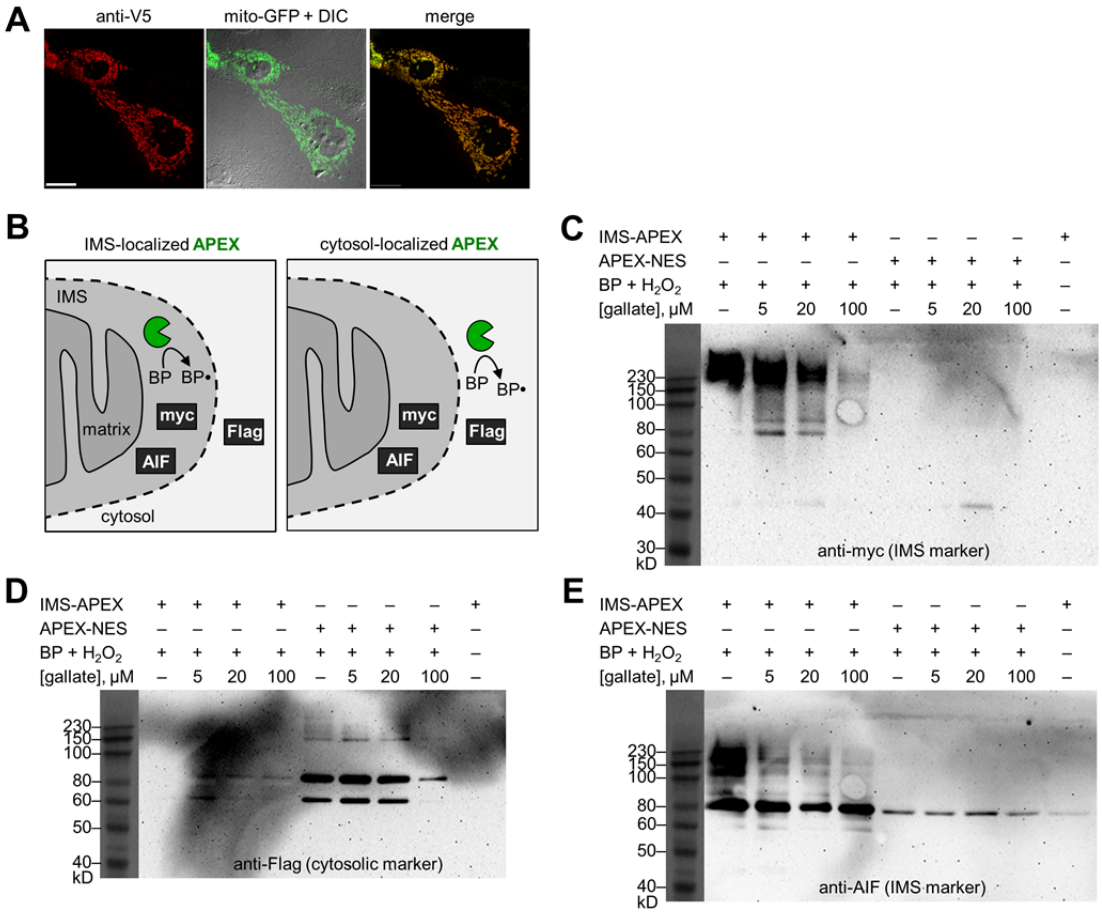
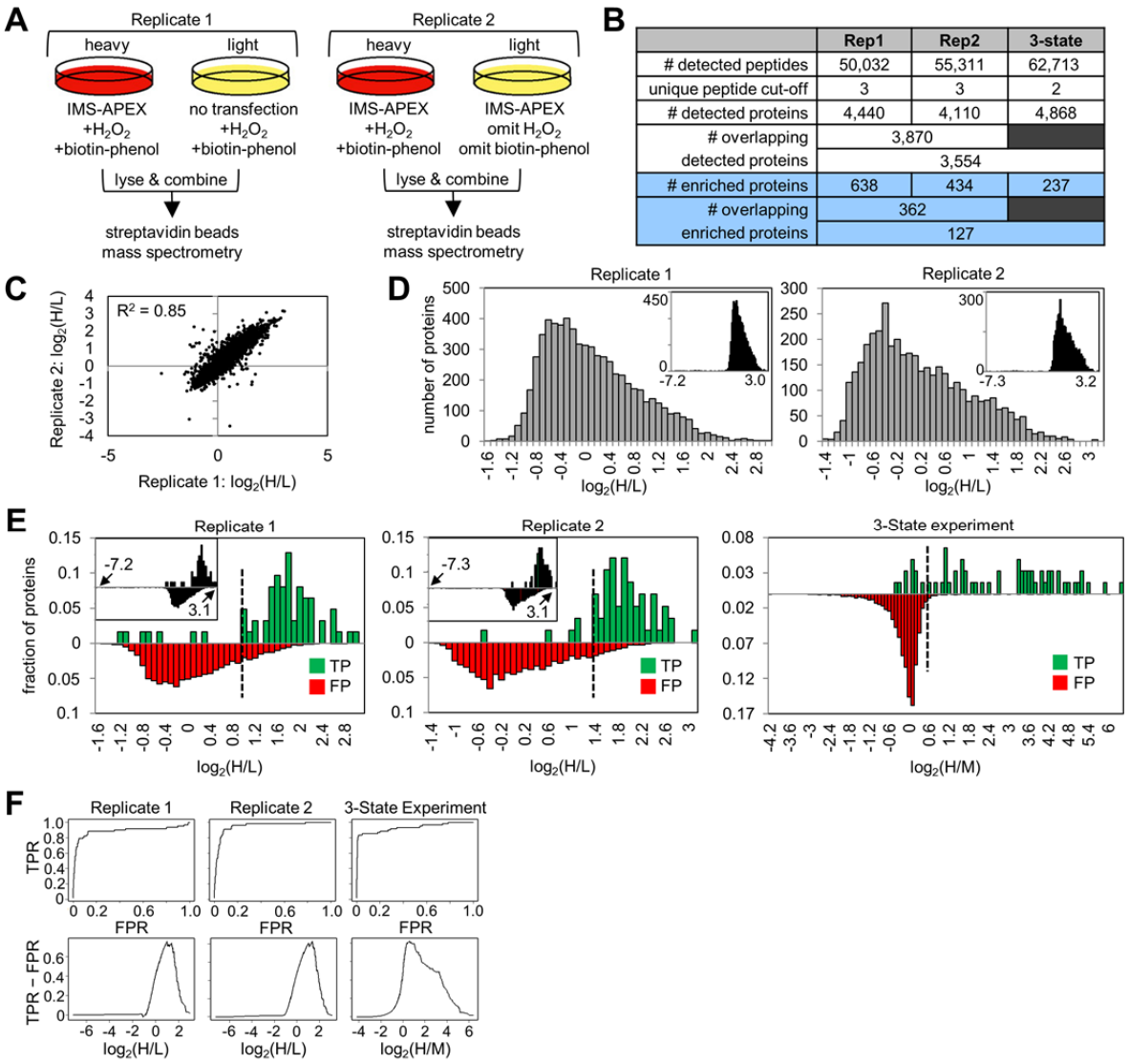


Figure S1. Western blot analysis of IMS-APEX labeling spatial specificity (related to Fig. 1). (**A**) Imaging analysis of IMS-APEX localization with respect to a mitochondrial marker. COS-7 cells were co-transfected with IMS-APEX and a mitochondrial GFP marker. After fixation, IMS-APEX was detected by anti-V5 staining (AlexaFluor568 readout). Scale bar, 20 µm. (**B**) Western blot assay scheme. If APEX labeling is spatially specific, then an IMS-resident protein (endogenous AIF, MW 67 kD or co-transfected LACTB-myc-Y5-YFP-Y5, MW 38 kD) should be biotinylated to a greater extent by IMS-APEX than by cytosol-localized APEX-NES. Similarly, a cytosolic marker protein (Flag-tagged mCherry-LplA, MW 68 kD) should be biotinylated to a greater extent by APEX-NES than by IMS-APEX. To test this, HEK 293T were transfected with either IMS-APEX or APEX-NES, along with a myc-tagged IMS marker and a Flag-tagged cytosolic marker. Biotin-phenol (BP) labeling was performed for 1 minute as in Figure 1C, then cells were lysed, and biotinylated proteins were enriched using streptavidin beads. Eluates were analyzed by western blotting with anti-myc antibody (**C**), anti-Flag antibody (**D**), and anti-AIF antibody (**E**). In these experiments, the effect of adding 0-100 µM of the radical quencher gallate together with biotin-phenol was also tested. Note that for both AIF and the myc-tagged IMS marker, most of the signal exists at high molecular weights. We hypothesize that the labeling chemistry causes these proteins to become cross-linked to other proteins during the 1 minute labeling window, and that the formation of such high molecular weight products is facilitated by the small volume of the IMS.



	Rep1	Rep2	3-state
# detected peptides	50,032	55,311	62,713
unique peptide cut-off	3	3	2
# detected proteins	4,440	4,110	4,868
# overlapping detected proteins	3,870		
	3,554		
# enriched proteins	638	434	237
# overlapping enriched proteins	362		
	127		

Figure S2. 2-state SILAC experimental scheme, and analysis of detected peptides and proteins in both 2-state and 3-state SILAC experiments (related to Fig. 2). (**A**) 2-state SILAC experimental setup. (**B**) Table showing numbers of detected peptides and proteins for the 2-state SILAC and 3-state SILAC experiments. Due to its greater sample complexity, the 3-state SILAC experiment detected fewer known IMS proteins (from our IMS gold+ list, tab 1 of Table S2) than the 2-state SILAC experiment did if the same unique peptide cut-off was used. Therefore, we required only two or more unique peptides for a protein to be "detected" in the 3-state experiment, whereas three or more unique peptides were required for each "detected" protein in the 2-state experiment. Enriched proteins remaining after the data filtering performed in row 5 of Figure 2D are shown in the blue-shaded rows. (**C**) Scatter plot showing correlation between replicate 1 and replicate 2 data in the two-state SILAC experiment. Out of 3,870 proteins, 14 points (~0.36%), all with very negative $\log_2$(H/L) ratios, are not shown in this graph. (**D**) Histograms of H/L ratios for replicates 1 and 2 of the two-state SILAC experiment (bin size 0.1). Complete distributions with expanded SILAC ratio ranges are shown in the insets. (**E**) True positive and false positive analyses on H/L distributions from the 2-state SILAC experiment (left), and H/M distribution from the 3-state SILAC experiment (right). True positive (TP)

histograms in green show detected proteins that are known to be IMS-exposed (i.e., in our IMS gold+ list, tab 1 of Table S2). False positive (FP) histograms in red show detected proteins that lack prior mitochondrial annotation. The dashed lines indicate the SILAC cut-offs applied in row 5 of Figure 2D. (**F**) Receiver operating characteristic and true positive rate (TPR) – false positive rate (FPR) analysis to determine optimal SILAC cut-off values. Top: For every SILAC ratio cut-off, the TPR is plotted against the FPR. Here, the TPR is defined as the fraction of detected IMS gold+ proteins that are above the SILAC cut-off. FPR is defined as the fraction of detected proteins that lack mitochondrial annotation above the SILAC cut-off. Bottom: plots depicting the difference between TPR and FPR at every SILAC cut-off.

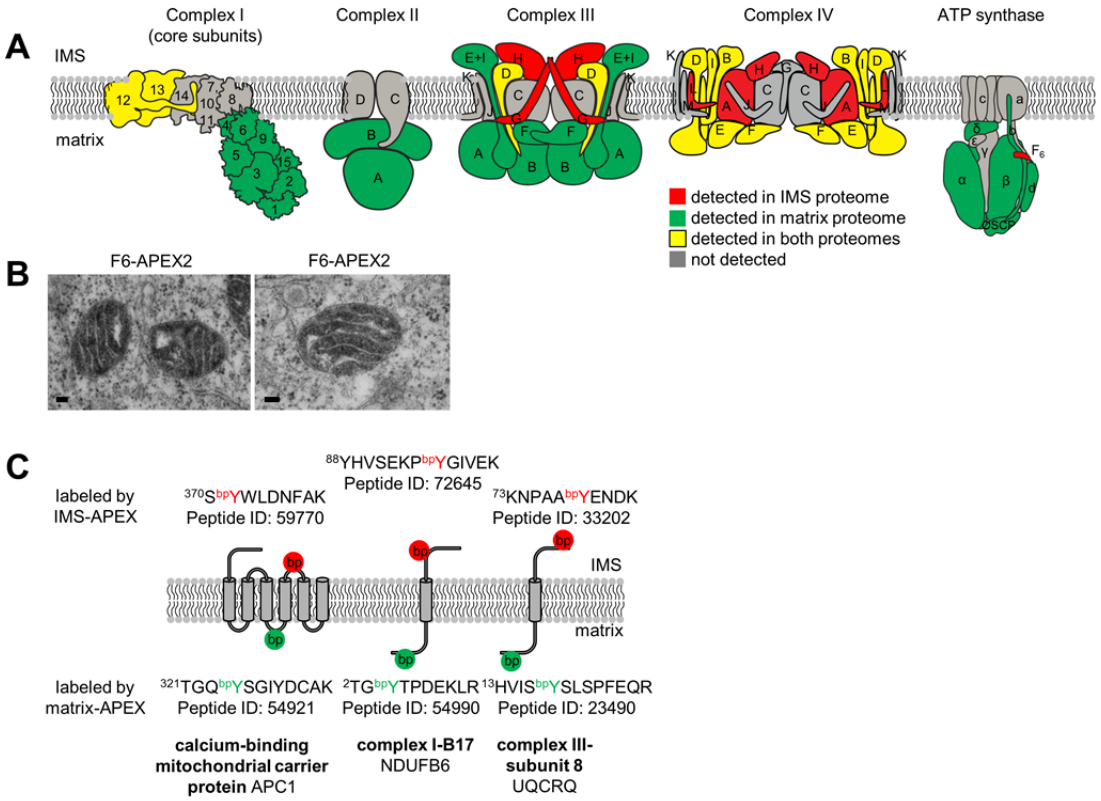


Figure S3. Assessment of IMS-APEX labeling specificity via analysis of inner mitochondrial membrane (IMM) proteins (related to Fig. 3). (**A**) Subunits of Complexes I-IV and F_0-F_1 ATP synthase (Complex V) for which structural information is available are depicted (derived from porcine (PDB ID 1ZOY), bovine (PDB IDs 1L0L and 1OCC), *E. coli* (Walker and Dickson, 2006), and *Thermus thermophilus* (Efremov et al., 2010) structures). Coloring is used to indicate the subunits detected in our IMS proteome (red), previous mitochondrial matrix proteome (Rhee et al., 2013) (green), or both proteomes (yellow). See tab 5 of Table S2 for details. (**B**) Electron microscopy of the F6 subunit of ATP synthase fused to APEX2 (Lam et al., *submitted*) at its C-terminus. The F6 subunit is part of the peripheral stalk that connects the F0 and F1 subunits (Walker and Dickson, 2006). F6 was highly enriched ($\log_2$(H/M) of 6.31 and $\log_2$(H/L) of 3.47 in the 3-state SILAC experiment) in the IMS proteomic experiment and absent from our previous matrix proteome (Rhee et al., 2013). The EM sample was prepared in HEK 293T as in Figure 1B. DAB staining is observed in the mitochondrial matrix rather than in the IMS. Lack of detection of F6 in the mitochondrial matrix proteome may be a result of steric inaccessibility. Scale bars, 100 nm. (**C**) Three IMM proteins produced biotinylated peptides in both the IMS and matrix (Rhee et al., 2013) proteomic datasets. The tyrosine sites labeled by IMS-APEX (red) and matrix-APEX (green) match the predicted membrane topologies of these proteins, shown above.

Table S1: Proteomic data

Tab 1. IMS proteome. 127 entries ranked by 3-state $\log_2$(H/M) SILAC ratio. In addition, 9 more grey-shaded entries are included; these proteins were just below the $\log_2$(H/M) cut-off applied in row 5 of Figure 2D, but still may include true IMS-exposed proteins.

Tab 2. Dual-labeled list. 33 potential IMS/cytosol dual-localized proteins not included in our IMS proteome because their H/M values were too low. These proteins had high extent of biotinylation by IMS-APEX (replicate 1 $\log_2(H/L) \geq 1.01$, replicate 2 $\log_2(H/L) \geq 1.37$, and 3-state SILAC experiment $\log_2(H/L) \geq 0.620$), but their $\log_2$(H/M) values were between -0.620 and 0.620, suggesting that they are not exclusively IMS-localized proteins.

Tab 3. Biotinylated peptides. For 47 proteins in the IMS proteome, biotinylated peptides were detected and sequenced.

Tab 4. Complete 2-state SILAC data. All proteins with 3 or more quantified, unique peptides in either replicate of the 2-state SILAC experiment.

Tab 5. Complete 3-state SILAC data. All proteins with 2 or more quantified, unique peptides in the 3-state SILAC experiment. This table was used to generate the scatter plot in Figure 2C.

Tab 6. Column definitions.

Table S2: Analysis of specificity, coverage, and specific hits

Tab 1. IMS gold+ list. 75 established IMS-exposed proteins curated from literature. This list was used to estimate the coverage of our IMS proteomic dataset.

Tab 2. Mitochondrial orphans. 16 proteins in the IMS proteome that lack prior mitochondrial annotation.

Tab 3. IMS orphans. 21 proteins in the IMS proteome that are known to be mitochondrial but lack IMS, IMM, or OMM annotation.

Tab 4. Sub-mito compartment. 579 human proteins have sub-mitochondrial localization available from GOCC, including 72 proteins detected in the IMS proteome. This table was used to calculate the IMS specificity values shown in Figure 3A.

Tab 5. Membrane complexes. Subunits of Complexes I-V and the TOM/TIM/PAM complex detected in the IMS proteome. Related to Figure 3B and Figure S3A.

Tab 6. IMS x 495 matrix. 33 proteins that appear in both the IMS proteome and the mitochondrial matrix proteome (Rhee et al., 2013).

Tab 7. Ca^{2+} uniporter. SILAC ratio values for the five known components of the mitochondrial calcium uniporter complex. Related to Figure 4C.

Tab 8. Column definitions.

Supplemental Experimental Procedures

Plasmids and cloning

Two strategies were used to generate fusion constructs. To generate constructs where short tags (e.g., V5 or Flag epitope tag) or signal sequences were appended to the protein, the tag was included in the primers used to PCR-amplify the gene. To generate APEX fusion constructs to syntaxin 17 and the F6 subunit of ATP synthase, we used PCR overlap extension to fuse APEX to the gene of interest. PCR products were digested with restriction enzymes and ligated into cut vectors (e.g., pcDNA3, pEGFPN1, and FCPGW). In all cases, the CMV promoter was used to drive expression in mammalian cells. The table below summarizes the plasmids cloned for this study.

Name	**Features**	**Promoter / Vector**	**Details**
matrix-APEX	*NotI*-mito-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pcDNA3	Soybean APEX has 3 mutations relative to wild-type APX: K14D, W41F, and E112K (Rhee et al., 2013). The wild-type soybean APX gene was a gift from Emma Raven (University of Leicester). Mito matrix targeting sequence ("mito"): MLATRVFSLVGKRAISTSVCVR AH (Fornuskova et al., 2010) V5: GKPIPNPLLGLDST
APEX-NES	*NotI*-Flag-APEX-NES-Stop-*XhoI*	CMV/ pcDNA3	Flag: DYKDDDDK NES: LQLPPLERLTLD
IMS-APEX	*NotI*-LACTB$_{(1-68)}$-*BamHI*-V5-APEX-Stop-*XhoI*	CMV/ pcDNA3	LACTB$_{(1-68)}$: MYRLLSSVTARAAATAGPAWD GGRRGAHRRPGLPVLGLGWAG GLGLGLGLALGAKLVVGLRGA VPIQS (Polianskyte et al., 2009)
LACTB-myc-Y5-YFP-Y5	*NotI*-LACTB$_{(1-68)}$-*BamHI*-myc-Y5-YFP-Y5-Stop-*XhoI*	CMV/ pCDNA3	myc: EQKLISEEDL Y5: YGYSGYSYGY
APOO-V5	*BglII*-APOO-V5-Stop-*NotI*	CMV/ pEGFPN1	Human APOO was obtained from OriGene.
V5-Stx17	*NotI*-V5-Stx17-Stop-*XhoI*	CMV/ pcDNA3	Human syntaxin17 was a gift from Noboru Mizushima (University of Tokyo).
A8MTT3-V5	*NotI*-A8MTT3-V5-Stop-*XhoI*	CMV/ pcDNA3	Human A8MTT3 was synthesized by GenScript.
APEX2-Stx17-V5	*BamHI*-APEX2-*AscI*-	CMV/	10 aa linker: KGSGSTSGSG

	10 aa linker-Stx17-V5-Stop-*SbfI*	FCPGW	
F6-APEX2-V5	*NotI*-F6-*AscI*-10 aa linker-APEX2-V5-Stop-*XhoI*	CMV/ pcDNA3	Human F6 was from the Broad hORFeome. 10 aa linker: KGSGSTSGSG

MGST3-V5, SELRC1-V5, and CCDC127-V5 expression plasmids in pLX304 used in Figure 3C were obtained from the Broad Institute hORFeome.

Preparation of APEX2-Stx17-V5 lentivirus (related to Fig. 4A)

Human embryonic kidney (HEK) 293T cells were cultured in Minimum Essential Medium (MEM) supplemented with 10% fetal bovine serum, penicillin, and streptomycin at 37 °C under 5% CO_2. To prepare lentivirus, cells were plated on a 6-well plate. Each well of cells (9.62 cm^2) was transfected with 1 µg of APEX2-Stx17-V5 fusion plasmid, 0.1 µg VSVG, and 0.9 µg dR8.91 using 10 µL Lipofectamine 2000 (Invitrogen) in MEM (without serum or antibiotics) at ~70% confluence. VSVG and dR8.91 are lentiviral packaging plasmids (Pagliarini et al., 2008). The cells were transfected for 3.5 hours. Then, the media was replaced with 2 mL fresh growth media. After 48 hours, the supernatant was collected and filtered through a 0.45 µm syringe filter. The filtered supernatant was used to infect cells immediately.

Electron microscopy of APEX fusion proteins (related to Fig. 1B, 4A, and Fig. S3B)

HEK 293T cells were cultured in MEM supplemented with 10% fetal bovine serum, penicillin, and streptomycin at 37 °C under 5% CO_2. To prepare the samples in Figure 1B and Figure S3B, cells were plated in 6-well plates, and two wells were used to generate each sample. The cells were transfected with 0.1-0.2 µg APEX fusion plasmid using 0.5-1.0 µL Lipofectamine 2000 (Invitrogen) in MEM (without serum or antibiotics) per 0.95 cm^2 of cells at 60-80% confluence. After four hours of transfection, the cell culture media was replaced with fresh growth media, and the cells recovered for 20-24 hours before fixation.

To prepare the lentivirus-transduced samples used for Figure 4A, HEK 293T cells were plated in two wells of a 6-well plate. For each well, the media was replaced with 2.5 mL fresh growth media and 700 µL of the filtered supernatant containing APEX2-Stx17-V5 lentivirus (see "Preparation of APEX2-Stx17-V5 lentivirus" above) and cultured for 48 hours before fixation.

Fixation and staining for EM was performed as previously described (Martell et al., 2012; Rhee et al., 2013). For fixation, room temperature 2% glutaraldehyde (Electron Microscopy Sciences) in sodium cacodylate buffer (100 mM sodium cacodylate with 2 mM $CaCl_2$, pH 7.4) was added to each sample; then the cells were moved to ice for 30 – 60 minutes. The samples remained on ice through the uranyl acetate staining. After fixation, the samples were washed 5 × 2 minutes in cold sodium cacodylate buffer, quenched with 20 mM glycine in sodium cacodylate buffer for 5 minutes, and rinsed again with cold sodium cacodylate buffer for 5 × 2 minutes. The samples were then overlaid with a solution of 1.4 mM 3,3'-diaminobenzidine (DAB) tetrachloride or DAB free base dissolved in HCl and 10 mM H_2O_2 in cold sodium cacodylate buffer to allow APEX to catalyze the polymerization of DAB. For IMS-APEX, the incubation time was 10 minutes; for all other samples, the incubation time was 5 minutes. The cells were then rinsed 5 × 2 minutes with cold sodium cacodylate buffer. The DAB polymers were

subsequently stained with 2% OsO_4 (Electron Microscopy Sciences) for 30 minutes in cold sodium cacodylate buffer. Samples were rinsed 5 × 2 minutes in chilled distilled water. Chilled 2% aqueous uranyl acetate (Electron Microscopy Sciences) was then added, and the samples were incubated overnight at 4 °C. The cells were then brought to room temperature. The samples were rinsed with distilled water. Then, the cells were scraped off, resuspended, and pelleted at 700×g for 1 minute. The supernatant was discarded, and the pellet was dehydrated for 10 minutes each time in 70%, 95%, 100%, and 100% ethanol.

For the samples in Figure 1B, the pellets were infiltrated for 1 hour in Eponate resin (Ted Pella) or EMBED-812 (Electron Microscopy Sciences) using 1:1 (v/v) resin and anhydrous ethanol, then overnight in 2:1 resin:ethanol, and then 2 hours in 100% resin. After infiltration, the samples were placed in fresh resin and polymerized for 48 hours at 60 °C, then cut into 400 nm sections using a diamond knife. The samples were imaged at the Harvard Medical School Electron Microscopy Facility on a FEI-Tecnai™ G² Spirit BioTWIN transmission electron microscope operated at 80 kV.

For all other samples, the pellets were infiltrated in EMBED-812 using 1:1 (v/v) resin and anhydrous ethanol for 1 hour, 2:1 resin:ethanol overnight, and then 100% resin overnight. The pellets were placed in fresh resin and polymerized at 60 °C for 48 hours. Embedded cell pellets were cut into 50 nm sections using a diamond knife on a Leica Ultracut UCT and imaged at the Whitehead Institute Keck Microscopy Facility on a FEI Technai Spirit transmission electron microscope operated at 80 kV.

Characterization of APEX-mediated biotinylation and IMS-APEX localization by fluorescence microscopy (related to Fig. 1D and Fig. S1A)

HEK 293T and COS-7 cells were grown in 1:1 Dulbecco's Modified Eagle's Medium (DMEM) with 4.5g/L glucose and L-Glutamine (L-Gln):MEM media supplemented with 10% fetal bovine serum, penicillin, and streptomycin. Cells were plated on approximately 7 mm×7 mm glass coverslips coated with 50 µg/mL human plasma fibronectin (Millipore) and transfected using 0.1 µg of plasmid and 0.8 µL Lipofectamine 2000 as above. For the sample used for Figure S1A, a mitochondrial GFP marker was co-transfected with the APEX fusions. After 24 hours, the samples used for Figure 1D were incubated with 500 µM biotin-phenol in growth media for 30 minutes at 37 °C. H_2O_2 was added to a final concentration of 1 mM for 1 minute at room temperature, and then the cells were washed three times with quencher solution (formulation given in "SILAC labeling and biotinylation of the IMS proteome"). The sample used for Figure S1A was not labeled with biotin-phenol, and instead was simply washed three times with quencher solution.

The samples were then fixed in 4% paraformaldehyde in "fixation buffer" (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM $MgCl_2$, 0.12 M sucrose, pH 7.4) at room temperature for 15 minutes. The cells were washed three times with Dulbecco's Phosphate Buffered Saline (DPBS) and permeabilized with cold methanol at -20 °C for 5 minutes. After three more washes with DPBS, the cells were blocked with 3% w/v bovine serum albumin (BSA) in DPBS overnight at 4 °C.

APEX expression was detected by staining the cells with anti-Flag (Sigma, 1:500 dilution) or anti-V5 (Invitrogen, 1:1000 dilution) mouse antibodies in 1% w/v BSA in DPBS for 1 hour at room temperature. Samples were then washed 4 × 5 minutes with DPBS. HEK 293T samples were then incubated with secondary goat anti-mouse-AlexaFluor488 (AF488) antibody

(Invitrogen, 1:2000 dilution) in 1% w/v BSA in DPBS. This solution also contained neutravidin protein (Invitrogen) coupled to AF647-NHS esters (Invitrogen) to detect biotinylated proteins at a 1:1000 dilution. The conjugation of AF647 to neutravidin was performed according to the manufacturer's protocol (Invitrogen). The COS-7 sample shown in Figure S1A was incubated with secondary goat anti-mouse-AF568 antibody (Invitrogen, 1:2000 dilution). Samples were washed 4 × 5 minutes with DPBS and then maintained in DPBS on ice for imaging.

The samples were imaged on a Zeiss AxioObserver.Z1 microscope equipped with a Yokogawa spinning disk confocal head, Cascade II:512 camera, and a Quad-band notch dichroic mirror (405/488/568/647 nm). Solid state 491 nm (~20 mW), 561 nm (~20 mW), and 640 nm (~10 mW) lasers were used to excite the samples. YFP/AF488 (528/38 emission filter), AF568 (617/73 emission filter), AF647 (700/75 emission filter), and differential interference contrast (DIC) images were acquired through a 63× oil-immersion lens using Slidebook 5.0 software (Intelligent Imaging Innovations). Acquisition times ranged from 100 to 1000 milliseconds.

Characterization of IMS-APEX and APEX-NES biotinylation by streptavidin blotting of cellular lysate (related to Fig. 1C)

This protocol is adapted from Rhee et al., 2013. HEK 293T were plated in 6-well plates and transfected with 1 µg of the APEX construct of interest using 8 µl of Lipofectamine 2000 for each well as described above. After recovery, the cells were labeled with biotin-phenol and H_2O_2 and quenched as above. After quenching, the cells were harvested by scraping and pelleted by centrifugation at 3,000×g for 10 minutes. The pellet was stored at -80 °C overnight, then lysed with RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% NP-40), 1× protease inhibitor cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF (phenylmethylsulfonyl fluoride), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox. Lysates were centrifuged at 16,000×g for 10 minutes at 4 °C, then run on a 10% SDS-PAGE gel.

Gels were transferred to nitrocellulose membranes. The membranes were stained with Ponceau S (10 minutes in 0.1% w/v Ponceau S in 5% acetic acid/water) and imaged. After destaining in deionized water, the blots were blocked with 3% w/v BSA in 1× TBST (0.1% Tween-20 in Tris-buffered saline) at 4 °C overnight. The blots were rocked in 0.3 µg/mL streptavidin-HRP (Thermo Scientific) in 1% w/v BSA in 1× TBST at room temperature for 1 hour, then washed with 1× TBST four times for 5 minutes each time. Finally, the blots were developed with Clarity Western ECL Substrate (Bio-Rad) and imaged on an Alpha Innotech gel imaging system.

Western blot analysis of APEX labeling radius (related to Fig. S1B-E)

HEK 293T were plated in 6-well plates. Each well was transfected using Lipofectamine 2000 as in "Characterization of IMS-APEX and APEX-NES biotinylation by streptavidin blotting of cellular lysate" with 200 ng of LACTB-Y5-YFP-Y5 plasmid, 320 ng of mCherry-Flag-LplA plasmid (mCherry-Flag-lipoic acid ligase was a gift from Katharine A. White), and 600 ng of either IMS-APEX or APEX-NES plasmid. About 24 hours post-media change, the cells were incubated with 500 µM biotin-phenol and 0 µM, 5 µM, 20 µM, or 100 µM gallate for 30 minutes at 37 °C. The purpose of gallate addition was to test the effect of a biotin-phenoxyl radical quencher on spatial specificity of APEX labeling. Biotinylation was initiated by addition

of H_2O_2 (final concentration 1 mM) for 1 minute. Then, the cells were washed 3 times with quencher solution (formulation given in "SILAC labeling and biotinylation of the IMS proteome"), and scraped and pelleted as above ("Characterization of IMS-APEX and APEX-NES biotinylation by streptavidin blotting of cellular lysate"). The pellets were lysed in 100 µL RIPA lysis buffer with quenchers and clarified by centrifugation as above.

Streptavidin-coated magnetic beads (Pierce catalog no. 88817) were prepared by washing twice with RIPA lysis buffer. Then, 90 µL of whole cell lysate was incubated at room temperature with 30 µL of streptavidin bead slurry and 500 µL of RIPA lysis buffer (without quenchers) for 1.5 hours with gentle rotation. The beads were washed with 2 × 1 mL RIPA lysis buffer, once with 1 mL of 2 M urea in 10 mM Tris-HCl pH 8.0, and again with 2 × 1 mL RIPA lysis buffer. Biotinylated proteins were eluted by boiling the beads in 30 µL 3× protein loading buffer supplemented with 20 mM dithiothreitol (DTT) and 2 mM biotin. The streptavidin eluate (SAE) was collected and run on a 12% SDS-PAGE gel.

Transfer to a nitrocellulose blot and Ponceau S staining was performed as above. Then, the blot was gently rocked at 4 °C in 1% w/v BSA in 1× TBST overnight. The blot was immersed in 1:200 mouse anti-AIF antibody (Santa Cruz Biotechnology) for one hour at room temperature with gentle rocking. After four rinses in 1× TBST for 5 minutes each time, the blot was incubated with secondary HRP-conjugated goat anti-mouse antibody (Invitrogen, 1:2000 dilution) in 1% w/v BSA in 1× TBST for one hour at room temperature. The blot was rinsed with 1× TBST for 4 × 5 minutes, then developed with SuperSignal West Pico reagent (Thermo Scientific) and imaged on an Alpha Innotech gel imaging system.

The blot was then stripped for 15 minutes in stripping buffer (2% w/v SDS, 0.8% β-mercaptoethanol, and 62.5 mM Tris-HCl pH 6.8) warmed to 60 °C. After 7 rinses with deionized water and two 1× TBST rinses, the blot was rocked in blocking buffer at room temperature for 40 minutes. Then the blot was immersed in 1:800 mouse anti-myc antibody (Millipore) in 1% w/v BSA in 1× TBST for one hour at room temperature. Subsequent washes, the 35 minute incubation with secondary antibody, blot development, and imaging were performed as described above.

The blot was stripped and washed for a final time as above, then blocked overnight at 4 °C in 1% w/v BSA in 1× TBST. Then, the blot was immersed in 1:1000 anti-Flag-HRP antibody (Sigma) in 1% w/v BSA in 1× TBST for one hour at room temperature. After four 1× TBST washes for 5 minutes each time, the blot was developed and imaged as above.

<u>In-gel digestion of biotinylated proteins, extraction, liquid chromatography, and mass spectrometry</u>

The table below is a supplement to "SILAC labeling and biotinylation of the IMS proteome" described under Experimental Procedures in the main text. This table summarizes each of the 8 cell culture samples that were prepared (column 1) and how they were combined (columns 2-4) in the three independent SILAC MS experiments.

Cell culture sample	2-State Replicate 1	2-State Replicate 2	3-State	APEX expressed	Biotin-phenol & H_2O_2
H1	+			IMS-APEX	+
H2		+		IMS-APEX	+
H3			+	IMS-APEX	+
M3			+	APEX-NES	+
L1	+			none	+
L2		+		IMS-APEX	–
L3			+	none	+
L4	+	+	+	IMS-APEX	+

All steps from in-gel digestion of biotinylated proteins through the database searching of the mass spectrometry data were performed as previously described (Rhee et al., 2013) but also detailed here.

Biotinylated proteins eluted from streptavidin beads were run on NuPAGE Novex Bis-Tris 4-12% gels for 1 hour at 130 V. The gels were stained overnight with Coomassie G-250 (Invitrogen) to ensure that proteins were indeed eluted. The gels were then destained with water, and the lanes for each SILAC experiment were manually cut into 16 gel bands. Gel bands were further destained in 1:1 acetonitrile:100 mM ammonium bicarbonate pH 8.0 for several hours, dehydrated with acetonitrile, and then swelled for one hour with shaking in 100 µL of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate. The DTT solution was then replaced with 100 µL of 55 mM iodoacetamide in 100 mM ammonium bicarbonate. After 45 minutes in the dark, the iodoacetamide solution was replaced with acetonitrile to dehydrate the bands. Then, 10-50 µL of 20 ng/µL Sequencing Grade Trypsin (Promega) was added to each sample. The samples were shaken at room temperature to digest overnight.

The next day, excess trypsin solution was removed from the samples. Peptides were extracted by incubating the bands in 20 µL 60% acetonitrile/0.1% TFA for 10 minutes, after which the extraction solution was collected. This step was performed three times. The final peptide extraction was carried out in 100% acetonitrile for 1-2 minutes. A vacuum concentrator was used to completely dry the peptides. Dried peptides were reconstituted in 100 µL of 0.1% formic acid and loaded on to C18 StageTips (Rappsilber et al., 2007) conditioned with 50 µL of 50% acetonitrile/0.1% formic acid. Loaded peptides were washed with 2 × 50 µL of 0.1% formic acid. Peptides were then eluted in 50% acetonitrile/0.1% formic acid, dried in a vacuum concentrator, and reconstituted in 0.1% formic acid/3% acetonitrile.

A Q Exactive mass spectrometer (Thermo Scientific) coupled online to an Easy-nLC 1000 UPLC (Proxeon) was used to analyze the peptides. The peptides were separated by reverse-phase chromatography using a self-packed capillary column (360 µm o.d. × 75 µm i.d.) with a 10 µm integrated emitter tip that was packed to 24 cm with 1.9 µm ReproSil-Pur C18-AQ resin (Dr. Maisch GmbH). The UPLC mobile phases A and B were 0.1% formic acid/3% acetonitrile and 0.1% formic acid/90% acetonitrile, respectively. Peptides were separated over a linear gradient that flowed at 200 nL/minute and increased from 6% B to 30% B over 82 minutes for a total acquisition time of 120 minutes.

The Q Exactive mass spectrometer was operated in a data-dependent mode. For the 12 most abundant precursor ions, consecutive high energy collision dissociation MS2 scans (resolution 17,500) were recorded after each MS1 scan (resolution 70,000). The MS1 automatic

gain control target was set to 3×10^6 ions with a maximum ion time of 10 msec. MS2 scans had an advance gain control target of 5×10^4 ions with a maximum ion time of 120 msec, isolation width of 2.5 m/z, normalized collision energy of 25, dynamic exclusion time of 20 seconds, and underfill ratio of 5.0%. Peptide match and isotope exclusion were enabled.

MaxQuant software package (version 1.2.2.5) (Cox and Mann, 2008; Cox et al., 2011) was used to identify and quantify proteins and peptides. The data were searched against the UniProt human database (81,470 entries) with an additional 248 laboratory contaminants provided by the MaxQuant framework. Carbamidomethylation of cysteine was searched as a fixed modification, and N-terminal protein acetylation, oxidation of methionine, and a single biotin-phenol on tyrosine were searched as variable modifications. The trypsin specificity was set to allow cleavages N-terminal to proline and a maximum of two missed cleavages. The maximum precursor ion charge state was set to 6. For the initial mass calibration, the precursor mass tolerance was 20 ppm, and for the main search, 6 ppm. The MS/MS tolerance was 20 ppm. The top MS/MS peaks per 100 Da was set to 10. The minimum peptide length was set to 6. The peptide and protein false discovery rates were set to 1%. Protein quantification was based on unmodified, N-terminally acetylated peptides and peptides with oxidized methionines.

Determination of SILAC cut-offs (related to Fig. 2D and Fig. S2)

Each of the three SILAC experiments was analyzed separately. Contaminants and proteins identified as reverse hits were removed. For the 2-state experiments, only proteins with three or more unique quantified peptides were considered "detected" and retained for further analysis; for the 3-state SILAC experiment, two or more unique peptides were required. Unique peptides are those that are designated by MaxQuant software as peptides not shared with other protein groups (Cox and Mann, 2008; Cox et al., 2011). The number of detected peptides for each SILAC experiment was calculated by counting the number of peptides with unique amino acid sequences (row 1 of Fig. S2B). Therefore, acetylated peptides, peptides with oxidized methionines, and peptides with biotin-phenol modifications were considered redundant with their unmodified counterparts. Protein SILAC ratios were calculated as the median of the corresponding peptide SILAC ratios (Cox and Mann, 2008; Cox et al., 2011). Protein SILAC ratios were normalized (after protein filtering based on the unique peptide count as described above) so that the median of each SILAC ratio distribution is 1.

We classified the detected proteins into three groups: (1) established IMS-exposed proteins that can be found on our IMS gold+ list shown in tab 1 of Table S2; (2) proteins without prior mitochondrial annotation (i.e., proteins not classified as mitochondrial by the Gene Ontology Cell Component (GOCC) database (Ashburner et al., 2000), MitoCarta (Pagliarini et al., 2008), our previous mitochondrial matrix proteome (Rhee et al., 2013), or literature); and (3) all other proteins (i.e., proteins with prior mitochondrial annotation but not on our IMS gold+ list). SILAC ratio cut-offs were determined using class (1) and class (2) proteins only.

For each experiment and for each SILAC ratio value, we calculated the true positive rate (TPR) and false positive rate (FPR) we would obtain if we retained only proteins above that SILAC ratio for our IMS proteome. We defined TPR as the fraction of class (1) proteins (i.e., detected IMS gold+ proteins) above the SILAC ratio in question, and FPR as the fraction of class (2) proteins (i.e., proteins without prior mitochondrial annotation) above the SILAC ratio in question. We selected SILAC ratios that maximize the difference between TPR and FPR as our cut-offs (Figure S2F). These cut-off values are shown in the table in Figure 2D.

Cut-offs in row 2 of Figure 2D were obtained differently. Here, we used the strategy previously employed in our mitochondrial matrix study (Rhee et al., 2013). Separate histograms were plotted for class (1) and (2) proteins binned by their $\log_2$(H/L) ratios in increments of 0.1 centered at 0 (Fig. S2E). For each bin, we calculated the FPR as the ratio of the fraction of class (2) proteins present in that bin vs. the fraction of class (1) proteins present in that bin. We then selected SILAC ratio cut-offs where the FPR falls below 0.1. Proteins above these values are >10 times more likely to be IMS proteins than non-mitochondrial proteins.

As shown in Figure 2D, we found that crossing H/L data from the 2-state experiments with H/M data from the 3-state experiment (row 5) gave a better combination of specificity and coverage than using H/L and H/M data from the 3-state experiment alone (row 4).

Scatter plot analysis of the 3-state SILAC experiment (related to Fig. 2C)

The scatter plot in Figure 2C contains all proteins from the 3-state SILAC experiment that were detected and quantified (these 4,868 proteins are listed in tab 5 of Table S1). As in the SILAC cut-off analysis above, proteins were separated into 3 classes: (1) true positives, which are proteins on our IMS gold+ list (tab 1 of Table S2); (2) false positives, which are proteins that are not classified as mitochondrial by the GOCC database (Ashburner et al., 2000), MitoCarta (Pagliarini et al., 2008), our previous mitochondrial matrix proteome (Rhee et al., 2013), or literature; and (3) all other proteins. In the scatter plot of $\log_2$(H/L) versus $\log_2$(M/L), class (1) proteins are colored green, class (2) proteins are colored red, and class (3) proteins are black.

There is a negative correlation between $\log_2$(H/L) and $\log_2$(M/L) values in the upper left quadrant, which contains most of the known IMS proteins. We hypothesize that this is due to our use of a spike-in sample (L4) in our 3-state SILAC experiment. The L4 sample expressed IMS-APEX and was labeled with with biotin-phenol and H_2O_2 like the H3 sample, but was cultured in light media instead of heavy media. The purpose of the spike-in was to avoid infinite H/L or M/L ratios in the event of proteins with no representation from the negative control light sample (L3, with IMS-APEX omitted). Due to our use of this spike-in, proteins that were strongly biotinylated by IMS-APEX in the heavy state would be expected to have higher absolute intensity values in the light state as well because of the contribution of the spike-in sample. The higher intensity values in the light state lead to lower $\log_2$(M/L) values for these proteins. Therefore, proteins with high $\log_2$(H/L) values will have lower apparent $\log_2$(M/L) values.

Specificity analysis (related to Fig. 3A and 3B)

To calculate the mitochondrial specificity of the IMS proteome (column 2 of Fig. 3A), we first determined if each entry has prior mitochondrial annotation according to the GOCC database (Ashburner et al., 2000), MitoCarta (Pagliarini et al., 2008), our mitochondrial matrix proteome (Rhee et al., 2013), or literature. We then divided the number of entries with prior mitochondrial annotation (column O in tab 1 of Table S1) by the total size of the IMS proteome (127 proteins).

For comparison, we also determined the fraction of the entire human proteome with prior mitochondrial annotation (column 1 of Fig. 3A). To calculate this, we retrieved the human proteome from the Uniprot-GO Annotation database in August 2012. A list of mitochondrial gene ontology (GO) terms was generated from the AmiGo database by searching for “mitochon”. Only GO cell component (GOCC) terms were retained from this list. The list of mitochondrial

GOCC terms were used to identify which genes in the human proteome had a mitochondrial GOCC term. 1,526 proteins of the total 18,863 human proteins had mitochondrial annotation.

To calculate the sub-mitochondrial specificity of the IMS proteome (column 4 of Fig. 3A), we considered the fact that IMS-, OMM-, and IMM-annotated proteins could all potentially have exposure to the IMS space. Therefore, for each entry of the IMS proteome with prior mitochondrial annotation (111 proteins), we checked for previous IMS, IMM, OMM, or matrix annotations using the following GO terms: GO:0005758 for IMS, GO:0005743 for IMM, GO:0005741 for OMM, and GO:0005759 for mitochondrial matrix. The results of this search are shown in columns R-U in tab 1 of Table S1. Any entry with a "1" (indicating "yes") in the IMS, IMM, or OMM columns was considered "IMS-exposed" (column V in tab 1 of Table S1). There were 71 such proteins, which we divided by the total number of IMS proteome proteins with sub-mitochondrial annotation information (72) to obtain our sub-mitochondrial specificity value of 99%.

For comparison, we also determined the fraction of all known mitochondrial proteins with IMS, IMM, or OMM annotation as opposed to matrix annotation (columns I-L in tab 4 of Table S2). Although many proteins have multiple sub-mitochondrial annotations, for the purpose of this analysis, we preferentially selected IMS annotation over IMM, IMM over OMM, and OMM over matrix. For example, any protein with the GO term "GO:0005758" was classified as IMS in column I in tab 4 of Table S2, even if that protein also had terms for IMM, OMM, and matrix. In addition, one protein, HSPE1, was considered an IMS protein rather than a mitochondrial matrix protein even though it is dual-localized (Samali et al., 1999). We found that 74% of the 579 proteins with sub-mitochondrial annotations available (listed in tab 4 of Table S2) were classified as IMS, IMM, or OMM (column 3 of Fig. 3A).

IMS orphans

To generate a list of IMS orphans (tab 3 of Table S2), we focused on the 111 proteins in the IMS proteome with prior mitochondrial annotation. We determined if each entry had prior IMS, IMM, or OMM annotation according to GOCC (columns R-T in tab 1 of Table S1), UniProt, or the literature. Those with such evidence are labeled with "1" in column W in tab 1 of Table S1 ("IMS-exposed") with the supporting evidence listed in column X in tab 1 of Table S1 ("Evidence for IMS exposure"). 21 entries lacked such evidence and are listed in our IMS orphan table (tab 3 of Table S2). In the IMS orphan list, 1 protein has mitochondrial matrix annotation, 3 proteins have mitochondrial membrane annotation, and the remaining 17 proteins lack sub-mitochondrial annotation.

Fluorescence imaging of mitochondrial orphans (related to Fig. 3C)

COS-7 cells were grown at 37 °C in MEM media or 1:1 DMEM with 4.5g/L glucose and L-Gln:MEM supplemented with 10% fetal bovine serum, penicillin, and streptomycin. Cells were plated on approximately 7 mm×7 mm glass coverslips coated with 50 μg/mL human plasma fibronectin. The cells were co-transfected with 0.2 μg of mitochondrial orphan plasmid and 0.1 μg of mitochondrial GFP marker using 0.5 μL Lipofectamine in MEM without FBS or antibiotics for 2.5-3.5 hours at 70% confluence. The media was then replaced with full media supplemented with 10% fetal bovine serum, penicillin, and streptomycin. After 20-44 hours, the cells were fixed with 4% formaldehyde in either 1× phosphate buffered saline or in fixation

buffer (formulation given above) at room temperature for 15- 20 minutes. Cells were permeabilized, blocked, and stained with anti-V5 antibody as described under "Characterization of APEX-mediated biotinylation and IMS-APEX localization by fluorescence microscopy". Goat anti-mouse-AF568 was used as the secondary antibody, and the cells were imaged with a confocal microscope as above.

Quantification of the extent of mitochondrial localization was performed using Slidebook 5.0 software. Fields of view containing multiple transfected cells were cropped so that each cell was in a separate image. For each image, the mean fluorescence intensity plus one standard deviation in each channel was taken as the threshold below which all intensity in that channel was defined as background. Mitochondrial regions were then identified as a binary mask by image segmentation of the mitochondrial GFP (AF488) channel based on the previously calculated threshold. Orphan protein regions were identified by image segmentation in the anti-V5 (AF568) channel by the same process. The sum intensity of the anti-V5 channel in the region of overlap between mitochondrial GFP signal and orphan signal was divided by the total sum intensity of the orphan signal in that cell. For each orphan, percentages were independently calculated for at least ten cells and then averaged.

As negative controls, the V5-tagged endoplasmic reticulum-resident P4HB in pLX304 (obtained from the Broad Institute hORFeome) and APEX-NES were each co-transfected with the mitochondrial GFP marker under similar conditions, imaged, and analyzed. As a positive control, IMS-APEX data used to generate Figure S1A (see "Characterization of APEX-mediated biotinylation and IMS-APEX localization by fluorescence microscopy") was analyzed in the same fashion.

Mitochondrial purification (related to Fig. 3D)

Liver mitochondria were purified from overnight-starved C57BL/6 mice (Pagliarini et al., 2008). After mice were euthanized by CO_2 asphyxiation followed by cervical dislocation, organs were excised and washed in 5 mL of ice-cold buffer A (220 mM mannitol, 70 mM sucrose, 5 mM HEPES-KOH, pH 7.4, 1 mM EGTA, 0.2% BSA, and protease inhibitor cocktail EDTA-free tablets (Roche)). Tissues were subsequently homogenized by hand in 7 mL of buffer A using a 15 mL Teflon pestle tissue grinder (4 strokes for liver). The homogenate was transferred to a 15 mL falcon tube and centrifuged at 600×g for 10 minutes, and the resulting supernatant was transferred to a 30 mL Beckman tube and spun at 8000×g for 10 minutes to pellet mitochondria. The resulting pellet was resuspended in 4 mL of buffer A, transferred to a 15 mL falcon tube, and centrifuged at 600×g for 10 minutes. The supernatant was transferred to a new 30 mL Beckman tube and spun at 8000×g for 10 minutes to pellet mitochondria. The mitochondrial pellet was resuspended in 200-300 µL of buffer A. At this step, a 50 µL sample was set aside (crude mitochondria).

The remaining mitochondria were further purified using a Percoll density gradient as previously described (Pagliarini et al., 2008). Briefly, a suspension of mitochondria was carefully layered on the top of a stepwise density gradient of 0.5 mL of 80%, 1.5 mL of 52%, and 1.5 mL of 26% Percoll prepared in buffer A in a 4 mL Beckman Ultra-Clear centrifuge tube and spun at 21,000×g for 45 minutes at 4 °C. Mitochondria were collected from the 52%-26% interface and placed in a 2 mL Eppendorf tube. Buffer A lacking BSA was added to the tube to 2 mL, and the tube was centrifuged at 21000×g for 10 minutes. The supernatant was removed, and the tube was refilled to 2 mL with buffer A lacking BSA. The sample was again centrifuged at

21000×g for 10 minutes. The resulting pellet was resuspended in 100-200 µL of buffer A lacking BSA to give Percoll purified mitochondria.

Western blot analysis of mitochondrial orphans (related to Fig. 3D)

The protein concentrations of the whole tissue lysate, crude mitochondria, and Percoll purified mitochondria from mouse liver tissue were determined using the BCA assay. 7 µg of each sample was combined with 1× protein loading buffer, boiled, and separated on 10% SDS-PAGE. The gels were transferred to nitrocellulose membranes and stained with Ponceau S as above. The blots were rocked in 3% w/v BSA in 1× TBST overnight at 4 °C. The blots were cut horizontally across using a razor blade. The strips were separated and immersed in 1:1000 rabbit anti-NBR1 (Cell Signaling Technology), 1:500 rabbit anti-FKBP10 (Proteintech), or 1:1667 mouse anti-calreticulin (Calbiochem) antibody solutions in 1% w/v BSA in 1× TBST for 1 hour and 10 minutes at room temperature. Following four 1× TBST washes for 5 minutes each time, the blots were incubated with 1:2000 goat anti-rabbit-HRP (Bio-Rad) or 1:2000 goat anti-mouse-HRP (Bio-Rad) in 1% w/v BSA in 1× TBST for 1.5 hours at room temperature. After four washes in 1× TBST for 5 minutes each time, the blots were developed with Clarity Western ECL Substrate (Bio-Rad) and imaged on an Alpha Innotech gel imaging system.

In a separate western blotting experiment, samples were prepared, separated on a 10% SDS-PAGE gel, and transferred to a nitrocellulose membrane as above. After Ponceau S staining, the blots were rocked in 3% w/v BSA in 1× TBST for 1 hour at room temperature. Then, the blots were incubated with 1:200 mouse anti-ATP5B (Santa Cruz Biotechnology) or 1:200 rabbit anti-CDC25C (Santa Cruz Biotechnology) in blocking buffer overnight at 4 °C with gentle rocking. After four 1× TBST washes for 7 minutes each time, the blots were rocked in 1:2000 goat anti-mouse-HRP or 1:2000 goat anti-rabbit-HRP for 1 hour at room temperature. After four washes in 1× TBST for 5 minutes each time, the blots were developed with SuperSignal West Pico reagent (for anti-ATP5B) or Clarity Western ECL Substrate (for anti-CDC25C) and imaged.

References

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. J. Proteome Res. *10*, 1794–1805.

Efremov, R.G., Baradaran, R., and Sazanov, L.A. (2010). The architecture of respiratory complex I. Nature *465*, 441–445.

Fornuskova, D., Stiburek, L., Wenchich, L., Vinsova, K., Hansikova, H., and Zeman, J. (2010). Novel insights into the assembly and function of human nuclear-encoded cytochrome *c* oxidase subunits 4, 5a, 6a, 7a and 7b. Biochem. J. *428*, 363–374.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Walker, J.E., and Dickson, V.K. (2006). The peripheral stalk of the mitochondrial ATP synthase. Biochim. Biophys. Acta BBA - Bioenerg. *1757*, 286–296.