Proteomic mapping in live Drosophila tissues using an engineered ascorbate peroxidase

Citation


As Published

http://dx.doi.org/10.1073/pnas.1515623112

Publisher

National Academy of Sciences (U.S.)

Version

Final published version

Accessed

Mon Nov 26 13:44:33 EST 2018

Citable Link

http://hdl.handle.net/1721.1/102096

Terms of Use

Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.

Detailed Terms

Please share how this access benefits you. Your story matters.
Proteomic mapping in live *Drosophila* tissues using an engineered ascorbate peroxidase

Chiao-Lin Chen\(^{a,b}\), Yanhui Hu\(^a\), Namrata Udeshi\(^b\), Thomas Y. Lau\(^a\), Frederik Wirtz-Peitz\(^a\), Li He\(^a\), Alice Y. Ting\(^b,c\),
Steven A. Carr\(^b\), and Norbert Perrimon\(^a,d\)\(^1\)

\(^{a}\)Department of Genetics, Harvard Medical School, Boston, MA 02115; \(^{b}\)Broad Institute, Cambridge, MA 02142; \(^{c}\)Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; and \(^{d}\)Howard Hughes Medical Institute, Boston, MA 02115

Contributed by Norbert Perrimon, August 12, 2015 (sent for review April 16, 2015; reviewed by Jeffrey D. Axelrod and Bernard Mathey-Prevot)

Characterization of the proteome of organelles and subcellular domains is essential for understanding cellular organization and identifying protein complexes as well as networks of protein interactions. We established a proteomic mapping platform in live *Drosophila* tissues using an engineered ascorbate peroxidase (APEX). Upon activation, the APEX enzyme catalyzes the biotinylation of neighboring endogenous proteins that can then be isolated and identified by mass spectrometry. We demonstrate that APEX labeling functions effectively in multiple fly tissues for different subcellular compartments and maps the mitochondrial matrix proteome of *Drosophila* muscle to demonstrate the power of APEX for characterizing subcellular proteomes in live cells. Further, we generate “MitoMax,” a database that provides an inventory of *Drosophila* mitochondrial proteins with subcompartmental annotation. Altogether, APEX labeling in live *Drosophila* tissues provides an opportunity to characterize the organelle proteome of specific cell types in different physiological conditions.

AMS-APEX | *Drosophila*

Specialized biological processes are carried out in specific organelles and subcellular compartments. For example, mitochondria are the site of oxidative respiration, neurons pass electrical or chemical signals to others through synapses, and apical and basolateral domains of epithelial cells are critical for their polarized functions. Understanding how these structures underlie specialized functions requires the comprehensive identification of proteins within spatially defined cellular domains.

A common strategy to study the localization of a particular protein is to generate green fluorescent protein (GFP) fusion proteins. However, it is time-consuming and labor-intensive to investigate protein localization at a large scale using GFP tagging, especially in vivo. Therefore, highly sensitive mass spectrometry (MS) approaches have been developed to systematically characterize the proteome of subcellular compartments. However, using MS approaches to characterize the proteome of subcellular domains has been limited by purification methods and is commonly associated with numerous false positives and false negatives due to contamination and loss of components during purification, respectively. For example, mitochondria are composed of an outer membrane and an inner membrane, generating two subcompartmental regions: the intermembrane space and the matrix located within the inner membrane. Because the ultrastructure of mitochondria is often disrupted during isolation processes, the isolation of specific subcompartmental regions of mitochondria is prone to contamination.

Recently, a method based on an engineered ascorbate peroxidase (APEX) has been developed and shown to function in cultured mammalian cells for proteomic mapping (1). Upon activation, the APEX enzyme turns a biotin-phenol substrate into a highly reactive radical that covalently tags neighboring proteins on electron-rich amino acids such as tyrosine. Biotinylated endogenous proteins can then be isolated and identified by MS. Thus, APEX labeling can be applied to bypass organelle purification steps, offering an alternative approach for systematic proteomic characterization in live cells. Here we report that the approach can be applied to characterize the subcellular proteome in live tissues and map the mitochondrial matrix proteome of *Drosophila* muscle. In addition to characterizing a number of uncharacterized putative mitochondrial proteins, we establish MitoMax, a database that provides an inventory of *Drosophila* mitochondrial proteins with subcompartmental annotation.

**Results**

**Expressing APEX in Different Subcellular Compartments of *Drosophila* Cells.** To express APEX in different *Drosophila* tissues at specific developmental stages, we used the UAS/Gal4 system (2) and generated flies with APEX fused to various signal peptides, including a nuclear localization signal (NLS) (3), a nuclear export signal (NES) (4), and the mitochondrial targeting sequence of human COXVIII (mito) (5) (Fig. 1). In addition, to validate the expression levels and patterns of different APEX constructs, APEX was fused with either a Flag tag or GFP at the C terminus. These constructs were expressed and examined in the body-wall muscle cells of third-instar larvae using the *DmeF-2-Gal4* driver. As expected, NLS-APEX localizes to nuclei, as identified by DAPI staining (Fig. 1B and Fig. S1). In contrast, mito-APEX expression tightly overlaps with ATP5a, also known as ATP5A1 in human, a known mitochondrial marker (6) (Fig. 1C), whereas

**Significance**

We use a protein labeling technique based on an engineered ascorbate peroxidase (APEX) to map the proteome of the mitochondrial matrix in live tissues. The approach allows us to establish MitoMax, a comprehensive database providing a high-quality inventory of *Drosophila* mitochondrial proteins with subcompartmental annotation. We demonstrate that APEX labeling is effective in vivo and provides an opportunity to characterize subcellular proteomes in specific cell types and in different physiological conditions. Given the interest in defining the mitochondrial proteome in different physiological conditions and tissues, our analysis provides a resource for systematic functional analyses of mitochondria that will in particular facilitate investigation of mitochondrial diseases.


Reviewers: J.D.A., Stanford University; and B.M.-P., Duke University.

The authors declare no conflict of interest.

Data deposition: The raw mass spectrometry data and the sequence database used for searches may be downloaded from MassIVE (massive.ucsd.edu) using the identifier MSV000079107. Download this dataset directly from ftp://MSV000079107@massive.ucsd.edu.

\(^1\)To whom correspondence may be addressed. Email: cichen@genetics.med.harvard.edu or perrimon@receptor.med.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515623112/-/DCSupplemental.
Chen et al.

APEX shows a nonnuclear expression pattern different from that of ATP5α (Fig. 1D).

APEX Labeling Functions Effectively in Various Fly Tissues. Next, we explored whether the APEX method could be applied to live tissues, as many biological processes cannot be studied in tissue-culture cells. Labeling of live tissues with APEX presents a number of challenges, because it requires that biotin-phenol and heme, the APEX cofactor, is present in a sufficient amount; catalase activity, which may quench H$_2$O$_2$, is low; and APEX transgenes are generated and expressed at the appropriate levels. Following a series of tests both in Drosophila S2R+ cells and dissected tissues, we found that for muscle studies, incubation with the biotin-phenol substrate for 30 min followed by 1-min incubation with H$_2$O$_2$ to activate APEX resulted in consistent biotinylation without additional heme supplementation (Fig. 2A).

Following covalent biotinylation of nearby endogenous proteins by APEX, the dissected tissues were stained with streptavidin to reveal the presence of biotinylated proteins (Fig. 2B–D and Figs. S1 and S2). Mito-APEX effectively labeled neighboring proteins, as shown by the overlap between the expression of the tag and the streptavidin staining in larval muscles (Fig. 2B and Fig. S1E). Further, we examined the activity of APEX in larval imaginal discs and found that APEX functions effectively in the subcellular regions in which it is expressed (Fig. 2C and D and Figs. S1A–C). When mito-APEX was expressed along the anterior–posterior compartmental boundary of the wing imaginal disc using ptc-gal4, the streptavidin staining overlapped with the mito-APEX expression and showed no background in the region where APEX was not expressed (Fig. 2C and D). Similarly, NLS-APEX and NES-APEX were able to catalyze biotinylation of proteins in the proper cellular compartment (Fig. 1B and Fig. S1). In addition to muscle cells and imaginal discs, similar results were also observed in the salivary gland (Fig. S1D–F). In

**Fig. 1.** Expressing APEX constructs in different subcellular compartments of fly muscle cells. (A) Strategy for targeting APEX to different cellular compartments. To express APEX in different Drosophila tissues at specific developmental stages, we used the UAS/Gal4 system and generated transgenic flies with APEX constructs under the regulation of UAS promoters. To target APEX to the nucleus, cytoplasm, or mitochondrial matrix, APEX is fused with a nuclear localization signal peptide, a nuclear export signal peptide, and a 29-aa peptide from human COX VIII, respectively. To detect the expression of APEX, GFP, or Flag, tags are added at the C termini of the constructs. (B–D) Expression and labeling activity of APEX targeted to different subcellular localizations in larval muscles. Immunostaining of muscle cells expressing NLS-APEX-GFP (B), mito-APEX-Flag (C), and NES-APEX-Flag (D) using Dmef2-Gal4. Nuclei are detected by DAPI (blue in B–D; gray in B). APEX is detected by either GFP expression or Flag staining (green in B–D; gray in B, C, and D). Mitochondrion is marked by ATP5α staining (red in C and D; gray in C′ and D′). Biotinylated proteins are visualized by streptavidin-coupled fluorescent staining (red in B; gray in B′).
summary, APEX functions effectively in multiple fly tissues (Figs. 1B and 2B–D and Fig. S1).

In addition to immunostainings (Fig. 2B and Fig. S3), we also analyzed the biotinylation status of fly muscles by Western blotting (Fig. S2). Few endogenously biotinylated proteins were present in negative control cells, whereas providing both substrate and H$_2$O$_2$ to cells without APEX expression was associated with low background. Incubating cells that express mito-APEX with substrate alone led to a weak background, and supplementing H$_2$O$_2$ alone to cells expressing mito-APEX had negligible effect on biotinylation. In contrast, providing both substrate and H$_2$O$_2$ to APEX-expressing cells generates specific and strong biotinylation.

Finally, we took advantage of the application of APEX for electron microscopy (EM) to examine the localization of mito-APEX at high resolution. It has been shown that APEX can catalyze diaminobenzidine (DAB) precipitation to generate contrast after OsO$_4$ fixation. Indeed, mito-APEX generates contrast specifically in the mitochondrial matrix (compare darker regions in Fig. 2F and H with controls in Fig. 2E and G), allowing us to confirm by EM that mito-APEX localizes specifically to the mitochondrial matrix.

**Mapping of the Mitochondrial Matrix Proteome by APEX Tagging.** To demonstrate the use of APEX to label the proteome of organisms in vivo, we mapped the proteome of the mitochondrial matrix in *Drosophila* muscle cells by quantitative MS (Fig. S4A). To distinguish proteins specifically biotinylated by APEX from endogenous biotinylated proteins and to subtract background or false positives caused during the process, such as nonspecific binding to streptavidin or beads, we used iTRAQ (isobaric tags for relative and absolute quantification) followed by liquid chromatography–mass spectrometry (LC-MS/MS) (8–11). Because iTRAQ allows chemically labeled peptides with ion reporters of different mass from four different samples to be simultaneously analyzed by MS, third-instar larval muscle of two different controls (wild-type and *Dmef2-Gal4* flies) and two replicates of mitochondrial APEX labeling (*Dmef2-mito-APEX-Flag* flies) were prepared. The dissected body-wall muscle samples from all four different groups were processed for APEX labeling, and the biotin-tagged proteome was affinity-purified using streptavidin-coupled beads. For quality control, biotinylation of endogenous proteins was confirmed by Western blotting using streptavidin-HRP (Fig. 3B). Furthermore, consistent with mitochondrial labeling, both the mitochondrial matrix protein ATP5α and mito-APEX-Flag were enriched by streptavidin beads (Fig. 3B). To retrieve enriched proteins, on-bead tryptic digestion was performed to generate proteolytic peptides. Collected peptides from wild-type and *Dmef2-Gal4* flies and two replicates of flies with mitochondrial APEX labeling were labeled with reporter ion tags 114, 115, 116, and 117, respectively. From LC-MS/MS, we retrieved 18,600 unique peptides that resulted in 2,222 genes with unique peptides >1 and iTRAQ ratio >1 for further analysis (Fig. S4A). Notably, the expression levels of each protein show high correlations between the mito-APEX replicates and between the two controls (Fig. 3C and Fig. S4C).

**Determination of the Mitochondrial Matrix Proteome.** To define the mitochondrial matrix proteome, iTRAQ ratios between experimental and control samples were calculated for each protein, giving rise to four different datasets of iTRAQ ratios (116/114, 117/114, 116/115, and 117/115). To maximize the recovery of mitochondrial matrix proteins with high specificity, we set the threshold of the false positive rate (FPR) to <0.1 as in previous studies (1) (Fig. 4A and Fig. S4B), which means that a protein is 10 times more likely to be a true mitochondrial protein than a false positive. The FPR is calculated based on the assembled lists.
of positive and negative controls that are representing proteins that are predicted to be localized either to mitochondria or to other structures, respectively, based on data and annotation in human or fly. At this threshold, 389 genes passed the cutoff for all four datasets and were selected as our final "mitochondrial matrix proteome" (Fig. 4A and B, Fig. S4A, and Dataset S1). The list contains both soluble matrix proteins and inner mitochondrial membrane proteins that are exposed to the matrix lumen.

To analyze the specificity of our mitochondrial proteome, we cross-referenced our data with positive and negative control lists. Compared with the fly genome, our dataset is indeed enriched with mitochondrial genes; 80.2% of our identified proteins have prior mitochondrial annotation, whereas 2.1% are annotated with other locations and thus are potential false positives (Fig. 4B and Dataset S1). The other 17.7% (69 proteins) potentially represent previously unidentified mitochondrial proteins (Fig. 4B and Dataset S1).

To analyze the depth of coverage, five established groups of functionally related mitochondrial proteins were analyzed (Fig. 4C and Dataset S2); 53–92% of proteins in each group were detected. Because this analysis relies heavily on the human mitochondrial annotation due to the lack of annotation for subcompartmental localization of mitochondria in the fly genome, bias and noise may be introduced during orthologous mapping. Alternatively, the core subcomplexes of mitochondria, which are more likely to share the same sublocalization, were

Fig. 4. Identification and analysis of the mitochondrial matrix proteome using APEX. (A) The threshold of the iTRAQ ratio is determined and based on a false positive rate of 0.1 (dashed line), which means that a protein is 10 times more likely to be a true mitochondrial protein than a false positive. (B) Analysis of specificity. Fraction of proteins in the entire fly proteome or in the matrix proteome. Discovery specificity; 80.2% of identified proteins have prior mitochondrial annotations, whereas 2.0% are false positives and 17.7% potentially represent previously unidentified mitochondrial proteins. (C) Analysis of discovery rate: Five groups of established mitochondrial proteins were analyzed, and 53–92% of proteins in each group were detected. (D) An example of COMPLEAT analysis. Each node represents one component of the complex, and node color reflects the average log2 ratio of each protein (red, positive log2; gray, proteins not detected; see more examples in Fig. S5). Solid lines represent binary protein–protein interactions identified in Drosophila, whereas dotted lines represent binary protein–protein interactions identified in other species. (E) Comparison of the specificity of identified fly mitochondrial proteins in different studies; 83% of the APEX-labeled matrix proteome corresponds to validated mitochondrial proteins (note that for this analysis, we included the 11 mitochondrial proteins validated in Fig. S5). In contrast, proteins identified from isolation-based approaches (16, 17) were 49% (345 out of 698) and 57% (624 out of 1,087), respectively. (F) Comparison of the results obtained from mitochondrial isolation methods and the APEX labeling method. Yin et al. (16) identified 718 proteins corresponding to 698 genes based on the FlyBase release 5.54 gene annotation. Lotz et al. (17) identified 1,089 genes; 210 genes are identified in all three studies, and 325 genes are identified both by the APEX labeling method and by at least one isolation-based experiment.
also examined using COMPLEAT (12), a bioinformatics tool for analyzing protein complex enrichment. For example, 13 out of 21 components (61.9%) of respiratory chain complex I were identified in our mitochondrial matrix proteome (Fig. 4D). In total, 65.7% of proteins in all enriched COMPLEAT complexes were discovered in our dataset (Fig. S5 and Dataset S3). Increasing the amount of input sample may slightly improve the coverage of proteins with low expression levels, although the recovery of mitochondrial proteins using iTRAQ ratios does not correlate with RNA expression levels (Fig. S4D).

To confirm that the identified proteins represent mitochondrial proteins, we examined their localization in S2R+ cells. Twenty-three genes were overexpressed in S2R+ cells by transfection of available constructs with C-terminal HA tags (13) and the proteins encoded by these genes were examined (Fig. 5, Fig. S6, and Dataset S1). Eleven of them clearly showed specific mitochondrial localization. Notably, we were able to discover previously uncharacterized mitochondrial genes, such as CG34140, that have not been previously identified by isolation-based approaches, illustrating the power of the APEX labeling method. Although the other 12 genes showed various expression patterns, including ubiquitous expression, cytoplasmic localization, or nuclear localization, it is possible that they also localize to mitochondria, as many mitochondrial proteins do not exclusively reside in mitochondria (14). In addition, artificially tagging GFP to a protein may disrupt its structure and thus affect its endogenous localization (15). Alternatively, these 12 genes may represent false positives from the APEX approach. Supporting evidence for mitochondrial localization of each protein from results from other species and by prediction tools is summarized in Dataset S1. Altogether, when adding the previously unidentified 11 proteins to the positive list, our APEX results show 83% specificity with the positive list (Fig. 4E).

We compared our results with the *Drosophila* mitochondrial proteome obtained in two previous studies (16, 17) that identified proteins from whole mitochondria, including matrix, both mitochondrial membranes, and intermembrane space, following traditional mitochondrial isolation (Fig. 4E and F and Fig. S7). Yin et al. (16) identified 718 proteins corresponding to 698 genes based on the FlyBase release 5.54 gene annotation; 49% of these were represented in our positive control list, and 9% were in our negative control list. In addition, Lotz et al. (17) identified 1,089 proteins, of which 57% were in our positive control list. On the other hand, 210 genes were identified in all three fly studies, and 325 genes were identified both by the APEX labeling method and by at least one of the isolation-based experiments (Fig. 4F). These genes are very likely to encode proteins specifically localized to the matrix or partially exposed to the matrix. In contrast, proteins that are only obtained by the isolation-based approach may represent proteins localized in mitochondrial subcompartments other than the matrix, such as the intermembrane space (Fig. S7). Altogether, our analyses indicate that the APEX-based method is able to facilitate proteomic mapping of finer subcellular compartments (matrix vs. whole mitochondria) and provide high coverage and excellent specificity (83% compared with 49–57% for the isolation-based approach).

**Fig. 5.** Validation of potential mitochondrial proteins in S2R+ cells. S2R+ cells were transfected with constructs carrying different proteins fused with HA at the C terminus. Not all cells are positive for the overexpressed constructs because of variability in transfection efficiency. Localization of proteins of interest was detected by HA staining (red in A–K; gray in A′–K′). Mitochondria are visualized by ATP5x staining (green in A–K; gray in A′–K′). Nuclei are detected by DAPI staining (blue in A–K).
requires the exposure of electron-rich residues such as tyrosine on the surfaces of target proteins (1), proteins that lack tyrosine residues or that are obscured by membranes or macromolecular complexes may not be detected with APEX. Nevertheless, we have identified not only unannotated mitochondrial genes but also nonconserved genes that are unlikely to be identified by orthologous mapping. These fly-specific mitochondrial genes may provide insights into mitochondrial–nuclear coevolution (22). Altogether, our analysis of APEX labeling in live Drosophila tissues indicates that the application of APEX not only provides a means to avoid potential problems during purification of organelles but also provides an opportunity to characterize the proteome of specific cell types under different physiological conditions.

Note that in our experiments, we are unable to determine the labeling radius, because the mitochondrion is a membrane-bound organelle and we are using a signal peptide to target APEX to the mitochondrial matrix rather than fusing APEX to a specific mitochondrial protein. However, in previous studies, the reactive phenoxyl radicals have been considered to have a half-life shorter than 1 ms and a <20-nm labeling radius (23–26), and thus APEX should prove useful to label subcellular domains beyond organelles in vivo.

Furthermore, we have generated a high-quality inventory of Drosophila proteins with submitochondrial compartmental annotation by integrating our results with those of previous studies. The MitoMax database for Drosophila genes encoding mitochondrial localized proteins is publicly available (genepath.med.harvard.edu/~perrimon/MitoMax.html). It provides a resource for systematic functional analysis of mitochondria, and in particular will facilitate investigation of mitochondrial diseases.

Methods

Generation of APEX Drosophila Lines. Plasmids encoding APEX were obtained from Martell et al. (7). APEX is wild-type APX with three engineered mutations (K41D, W41F, E112K). Signal peptide used in this study are nuclear localization signal (3): PKKKKRKK; nuclear export signal (4): LAL-KLALGD; and mitochondrial signal peptide (5): N-terminal 29 aa of human COXIVIII. The USAG/Gal4 system (2) was used for overexpression studies using Dme2-Gal4 (27) and ptc-Gal4 (28, 29) drivers. For APEX labeling, fly tissues were dissected and incubated with biotin-phenol. APEX was activated for protein labeling with H2O2. Samples were fixed for immunostaining or lysed for Western blotting and further proteolytic analysis.

Proteomic Analyses. Enrichment of biotinylated proteins from cell lysates was performed using streptavidin beads. On-bead digestion was subsequently performed to retrieve peptides of biotinylated proteins. The resulting digested peptides were processed for 4-plex iTRAQ labeling. Labeled peptides were separated by StageTips strong cation exchange (SCX) using a protocol adapted from Rappsilber et al. (30). Only proteins identified by >1 unique peptide with quantified ratios were retained for further analysis. All of the genes identified by iTRAQ along with their annotation are listed in Dataset S1. For details on how the cutoff of the iTRAQ ratio was selected, see SI Methods.

Bioinformatics Analyses. COMPLEAT (12) was used to identify complexes enriched among the genes identified by APEX. To build MitoMax, a comprehensive database for Drosophila mitochondrial genes with subcellular annotation, genes identified from isolation-based studies and/or APEX labeling were combined and integrated with genes from annotation as well as Drosophila genes annotated at MitoMiner (18) and MitoDrome (19).

Acknowledgments. We thank the Bloomington Stock Center for stocks, Spyros Artavanis-Tsakonas for reagents, Christians Villalta for technical assistance, Maria Ericson for assistance with electron microscopy, Hyun-Woo Rhee and Peng Zou from the A.Y.T. laboratory for reagents and advice, and members of the N.P. laboratory for discussions. C.-L.C. is a Postdoctoral Fellow of Ellison Medical Foundation/AFAR (American Federation for Aging Research). N.P. is an Investigator of the Howard Hughes Medical Institute (HHMI). This work was supported in part by National Institutes of Health Grants P01-CA120964 and R01-DK088718 (to N.P.) and HHMI's Collaborative Innovative Awards (HCIA).

References