Enzyme-based Reporters for Mapping Proteome and Imaging Proteins in Living Cells

By

Peng Zou

B.S. in Chemistry and in Physics (2007)
Peking University

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY
APR 1 [illegible] 2013
LIBRARIES

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 2013

© 2013 Massachusetts Institute of Technology
All rights reserved

Signature of the Author: ____________________
Department of Chemistry
Dec. 7, 2012

Certified by: ____________________
Alice Y. Ting
Ellen Swallow Richards Associate Professor of Chemistry
Thesis supervisor

Accepted by: ____________________
Robert W. Field
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Alexander M. Klibanov
Firmenich Professor of Chemistry and Bioengineering
Thesis committee chair

Alice Y. Ting
Ellen Swallow Richards Associate Professor of Chemistry
Thesis supervisor

Monty Krieger
Whitehead Professor of Molecular Genetics
Thesis committee member

Enzyme-based Reporters for Mapping Proteome and Imaging Proteins in Living Cells

By

Peng Zou

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

ABSTRACT

Each eukaryotic cell is exquisitely divided into organellar compartments whose functions are uniquely defined by the set of proteins they possess. For each individual protein, precise targeting to a specific sub-cellular location and trafficking between compartments are often key to its proper function. In fact, many human diseases are linked to mutations that cause mis-targeting and/or defective trafficking. This thesis describes the development of enzyme-based reporters for measuring protein localization and trafficking. We employ two complementary approaches: a top-down approach, involving proteomics, to simultaneously acquire the sub-cellular localization information for hundreds of proteins; and a bottom-up approach, involving fluorescence imaging, to record detailed spatial information for proteins on an individual basis. This thesis is therefore divided into the following two parts.

Part A describes a promiscuous protein labeling technique for proteomic mapping of organelles. This method capitalizes on peroxidase as a source of free radical generator. Compared to traditional sub-cellular fractionation methods, this novel approach obviates the need of organelle purification, thereby not only eliminating the potential artifacts associated with purification, but also greatly improving the temporal resolution of the proteomic mapping. Applying this technique to study the proteome of mitochondrial matrix and endoplasmic reticulum lumen has led to the discovery of several mitochondrial proteins whose localizations have previously been unknown or ambiguous.

Part B discusses the development and application of site-specific protein labeling methods for studying receptor trafficking mechanisms. Building upon previous work in our lab, we utilized the *Escherichia coli* biotin ligase BirA and its acceptor peptide to site-specifically label the low-density lipoprotein receptor and studied its internalization and trafficking both at the ensemble imaging and single-molecule level. We discovered that this receptor internalizes as an oligomer into cells.

Thesis Supervisor: Alice Y. Ting
Title: Ellen Swallow Richards Associate Professor of Chemistry

Acknowledgements

First of all, I would like to thank Alice for being a great mentor in the past five years. I came to this lab with a strong interest in chemical biology; however, the focus of my undergraduate research was in a quite different field, physical chemistry. Knowing this, Alice has patiently trained me by designing projects around my strengths in quantitative analysis while at the same time encouraging me to develop new experimental skills of organic synthesis and cutting-edge fluorescence microscopy. She recommended me to enroll in the Physiology summer course at the Marine Biological Laboratory in Woods Hole, Massachusetts. This highly interdisciplinary educational experience ignited my interest in building physical and chemical tools for quantitative analysis of biological phenomena, an interest that has lasted till today. Throughout the years, Alice has taught me to think logically, to act efficiently, to pay attention to details, and above all, to keep a deep passion for science. All of these would be invaluable for my future academic career.

Next, I want to thank members of my thesis committee: Professor Monty Krieger and Professor Alex Klibanov. As I developed this thesis, Monty has kindly shared with me his knowledge in cell biology, especially his expertise on the low-density lipoprotein receptor. He read through my manuscript word-by-word, and provided detailed comments for revision. As my thesis committee chair, Alex has always been warm and encouraging. I truly appreciate his advice on my career development at our thesis meetings. I am also grateful to our collaborators: Professor Vamsi Mootha (Harvard Medical School) is extremely knowledgeable about mitochondrial biology; Dr. Steve Carr and Dr. Namrata Udeshi (Broad Institute) are experts on mass spectrometry-based proteomics. This thesis would not be possible without their contributions.

Then, to the group of brilliant and creative scientists in the Ting lab: I felt extremely fortunate to spend five years with you and to learn from each one of you. Hyun-Woo and Jeff, my teammates for the proteomics project, it has been a fun and intellectually stimulating journey working with you. Dan and Tao, my best buddies, I have always enjoyed our conversations, both in the lab and on our trips to national parks. Thank you for sharing with me your passion for science and many brilliant ideas. I want to thank everyone else in the Ting lab for moral support over the years, for helping me at various stages of my career, for sharing laughter on lab trips and parties, for working hard in summer volleyball games, and for all the good memories.

Last but certainly not least, I am deeply grateful to my family. Dad and mom, it was your constant support that kept me moving forward. Five years and ten thousand miles away from home, your voices over the phone have always been a great comfort to me, and also a reminder that I should live my life to the fullest potential, to make each day count. I want to thank my wife, Xiaochen, for sharing with me my happiness and sorrows. In times of difficulty, when I felt lost, she encouraged me to believe in myself and to fight for my dreams. When I was obsessed with science and seemed to forget anything else, she reminded me to open up my mind and to realize that there are other important things in life. This thesis is dedicated to my family, for their unconditional love and support.

Table of Contents

Title Page 1

Signature Page 2

Abstract 3

Acknowledgements 4

Table of Contents 5

List of Figures 10

List of Tables 13

List of Abbreviations 14

Part A Proteomic analysis of sub-cellular protein localization

Chapter 1 Introduction to sub-cellular proteomic mapping 19

- 1.1 Introduction 21
- 1.2 Existing methods for sub-cellular fractionation 22
  - Sub-cellular fractionation of mitochondria 22
  - Sub-cellular fractionation of endoplasmic reticulum 23
  - Limitations of sub-cellular fractionation 24
- 1.3 Proximity-dependent promiscuous protein labeling 25
  - Promiscuous biotin ligase 26
  - Enzyme-mediated activation of radical source 27
  - Direct crosslinking to endogenous proteins 28
- 1.4 Existing mass spectrometry techniques for proteomics 30
  - Protein identification by mass spectrometry 31
  - Quantitative mass spectrometry 32
- 1.5 Conclusion 34
- References 35

Chapter 2 Promiscuous protein labeling with ascorbate peroxidase 41

- 2.1 Introduction 43

2.2 Characterization of ascorbate peroxidase as a promiscuous labeling enzyme 46
Screening of biotin-phenol substrates 46
In vitro characterization of labeling product 49
Improving ascorbate peroxidase through mutagenesis 50
Comparison between APX and promiscuous BirA labeling 52

2.3 Assessing the spatial specificity of APEX labeling 53
Evaluating APEX labeling by microscopy and Western blot analysis 53
APEX labeling does not cross cellular membranes 56
Preserving spatial specificity by immediate quenching of labeling reaction 58

2.4 Controlling labeling radius 59
2.5 Conclusion 64
2.6 Experimental methods 65
Genetic constructs 65
Reagents 66
In vitro enzyme kinetics 67
Mass spectrometry analysis of *in vitro* labeling 67
Immunostaining and confocal fluorescence imaging 67
STORM super-resolution fluorescence imaging 68
Streptavidin blotting of whole cell lysate 70
Imaging assay to determine the membrane-permeability of the phenoxyl radical 70
Proteomic labeling with promiscuous BirA 71

References 72

Chapter 3 Spatially-resolved organellar proteomic mapping 75

3.1 Introduction 77
3.2 Proteomic mapping of the mitochondrial matrix 77
Promiscuous protein labeling with APEX in mitochondrial matrix 77
Determine mitochondrial matrix proteome 80
Specificity analysis of mitochondrial matrix proteome 83
Depth of coverage analysis of mitochondrial matrix proteome 85

Insights from the proteomic dataset 87
Detection of biotin-labeled peptides 90

3.3 Secretory pathway proteome 92
Promiscuous protein labeling in the endoplasmic reticulum lumen with HRP 92
Specificity analysis of our secretome 93
Analysis of endoplasmic reticulum membrane proteins 94

3.4 Conclusion 96
3.5 Experimental methods 97
Labeling of the mitochondrial matrix proteome 97
Enrichment of biotinylated proteins, in-gel protein digestion and extraction 98
Determination of the cut-off point for our matrix proteome 99
Structural mapping of biotinylated peptides 101

Reference 102

Part B Fluorescence imaging analysis of protein localization and trafficking

Chapter 4 Introduction to fluorescence imaging analysis of protein trafficking 107

4.1 Introduction 109
4.2 Existing methods for fluorescent labeling of specific proteins 109
4.3 Fluorescence microscopy for studying protein localization and trafficking 112
4.4 Conclusion 115
References 116

Chapter 5 Ensemble fluorescence imaging analysis of receptor trafficking 119

5.1 Introduction 121
Clathrin-mediated endocytosis 121
Existing methods for measuring receptor oligomerization state 122

5.2 Fluorescence imaging assay for receptor internalization 123
Cell surface fluorescence labeling of LDL receptor 123

Cell surface fluorescence quenching 125
Internalization assay 127

5.3 Determine receptor oligomerization state during endocytic internalization 129
Oligomerization assay 130
LDL ligand does not affect oligomerization state 136

5.4 Enzymatic incorporation of unnatural probe desthiobiotin 138
Enzyme-mediated DTB ligation to AP peptide 139
Applications of intracellular DTB labeling 144

5.5 Conclusion 145
5.6 Experimental methods 146
Reagents and constructs 146
Fluorescence measurement 147
Testing membrane permeability of QSY21-spermine 148
Mammalian cell culture and transfection 148
Fluorescence microscopy and data analysis 149
Cell surface labeling of AP-receptors with mSA-Alexa568 150
Control expression level of AP-LDLR in T-REx CHO A7 stable cell line 150
Receptor internalization assay 152
Evaluation of LDLR co-expression efficiency using immunostaining 154
Co-immunoprecipitation analysis of LDLR oligomerization 154

References 156

Chapter 6 Single molecule imaging analysis of receptor internalization 159

6.1 Introduction 161
6.2 Engineering small QDs for studying receptor-mediated endocytosis 162
Establishing a model system 162
Engineer small, monovalent QD reagents 164

6.3 Single-molecule imaging with phycoerythrin 167
6.4 Conclusion 171
6.5 Experimental methods 172

Quantum dot ligand synthesis 172
Mammalian cell culture and imaging 177
Cell surface labeling with R-PE 177

References 179

Curriculum Vitae 181

List of Figures

Figure 1-1. Scheme of proximity-dependent promiscuous protein labeling 25
Figure 2-1. Structure of soybean cytosolic ascorbate peroxidase in complex with an aromatic substrate. 44
Figure 2-2. Catalytic cycle of heme peroxidase 45
Figure 2-3. Scheme of phenoxyl radical crosslinking. 46
Figure 2-4. Reactivity of eight different biotin substrates towards APX and HRP. 48
Figure 2-5. Amino acid specificity of labeling. 49
Figure 2-6. Gel filtration chromatography was used to analyze the dimerization tendency of mutant and wild-type soybean APX. 50
Figure 2-7 Engineering a monomeric peroxidase with high activity towards biotin-phenol. 51
Figure 2-8. Biotinylation of mitochondrial proteins by mito-pBirA is undetectable after 10 min or 1 hour of labeling. 52
Figure 2-9. Comparison of promiscuous biotinylation by APEX and pBirA with Western blot. 53
Figure 2-10. Confocal microscopy analysis of APEX and HRP-catalyzed promiscuous biotinylation in many cellular compartments. 54
Figure 2-11. STORM imaging analysis of APEX promiscuous labeling in mitochondria. 55
Figure 2-12. Western blot analysis of APEX and HRP-catalyzed promiscuous biotinylation in various cellular compartments. 56
Figure 2-13. Peroxidase-generated phenoxyl radicals do not cross the plasma membrane 57
Figure 2-14. Quencher cocktail effectively reduces biotin-phenol labeling intensity. 58
Figure 2-15. Titration of H_2O_2 does not affect labeling radius. 59
Figure 2-16. Biotin-phenol probe titration does not affect labeling radius 60
Figure 2-17. Addition of free radical quencher reduces the labeling radius 63
Figure 2-18. Determine the resolution of STORM setup. 69
Figure 3-1. Scheme of promiscuous labeling in mitochondrial matrix. 78
Figure 3-2. Electron microscopy (EM) showing the localization of mito-APEX in the mitochondrial matrix of HEK293T cells. 78
Figure 3-3. Labeling the mitochondrial matrix proteome with biotin-phenol. 79
Figure 3-4. SILAC labeling scheme. 81

Figure 3-5. Determine the cutoff point for enrichment 82
Figure 3-6. Defining our mitochondrial matrix proteome. 82
Figure 3-7. Our proteome is specific to mitochondrial matrix proteins. 83
Figure 3-8. Analysis of labeling specificity for inner mitochondrial membrane protein complexes. 84
Figure 3-9. Analysis of labeling specificity for mitochondrial translocase complexes. 85
Figure 3-10. Depth of coverage of our mitochondrial matrix proteome 86
Figure 3-11. Correlation between SILAC ratio and protein expression level in U2OS cells. 86
Figure 3-12. Immunofluorescence imaging analysis of a subset of "mito orphan" proteins. 87
Figure 3-13. Literature model of CPOX and PPOX localization within the mitochondria. 88
Figure 3-14. EM study of sub-mitochondrial localization of the heme biosynthesis enzymes CPOX and PPOX. 89
Figure 3-15. Example of MS/MS spectrum of directly labeled peptide. 90
Figure 3-16. Mapping biotinylated peptides to soluble matrix proteins. 91
Figure 3-17. Mapping biotinylated peptides to transmembrane proteins. 91
Figure 3-18. Promiscuous protein labeling in the endoplasmic reticulum lumen. 92
Figure 3-19. Distribution of protein enrichment ratio. 93
Figure 3-20. Distribution of protein localizations in our secretome dataset. 94
Figure 3-21. Analysis of 20 abundant endoplasmic reticulum transmembrane proteins in U2OS cells. 95
Figure 5-1. BirA-mediated site-specific biotinylation of AP peptide. 123
Figure 5-2. Scheme for selective labeling of cell surface AP-LDLR. 124
Figure 5-3. Specific labeling of AP-LDLR 125
Figure 5-4. Characterization of QSY21-spermine. 126
Figure 5-5. Testing membrane permeability of QSY21-spermine. 127
Figure 5-6. Internalization analysis time course of wild-type LDL receptor. 128
Figure 5-7. Internalization analysis of mutant LDLR. 129
Figure 5-8. Scheme for oligomerization assay. 130
Figure 5-9. Co-internalization analysis of Δ792 LDLR. 131
Figure 5-10. Internalization analysis of Y807 LDLR mutants. 132

Figure 5-11. Internalization analysis of Y807C LDLR co-expressed with transferrin receptor (TfnR) as a negative control. 133
Figure 5-12. Co-immunoprecipitation analysis of LDLR oligomerization. 135
Figure 5-13. Internalization analysis of wild-type and mutant LDLR in the presence of LDL. 137
Figure 5-14. Summary of LDLR internalization analysis. 138
Figure 5-15. Reversible binding of streptavidin to DTB. 139
Figure 5-16. In vitro desthiobiotin ligation using purified BirA. 140
Figure 5-17. Cell surface desthiobiotinylation of AP-CFP-TM protein with exogenous BirA. 141
Figure 5-18. Wild-type BirA efficiently ligates DTB-AMP to AP peptide on the cell surface. 142
Figure 5-19. Titration of DTB adenylate probe and quantitation. 142
Figure 5-20. Intracellular DTB ligation using AM ester. 143
Figure 5-21. Blocking endogenous biotin signal using enzyme-mediated destiobiotinylation. 144
Figure 5-22. Titration of AP-LDLR expression level in T-REx CHO A7 cells. 151
Figure 5-23. The mSA label on AP-LDLR does not affect LDL binding and internalization. 152
Figure 5-24. Confirmation of LDLR co-expression by epi-fluorescence imaging. 153
Figure 6-1. Two proposed models for receptor entrapment into the CCPs. 161
Figure 6-2. Single-molecule fluorescence imaging experiment with commercial QDs. 163
Figure 6-3. The synthetic scheme of novel QD coating ligand. 165
Figure 6-4. Specific targeting to biotinylated LDL receptor on the cell surface. 166
Figure 6-5. Separating mSA-conjugated QDs with new coating. 166
Figure 6-6. Scheme for using lipoic acid ligase (LplA) to conjugate a hydrazine functional group to LplA acceptor peptide (LAP) fusion proteins. 168
Figure 6-7. Protein labeling with phycoerythrin on the surface of living HEK293T cells. 169
Figure 6-8. Single-molecule imaging with phycoerythrin on the surface of living COS7 cells. 170

List of Tables

Table 2-1. Michaelis-Menten kinetics of APX and HRP toward guaiacol 46
Table 2-2. Genetic constructs 65
Table 3-1. Summary of MS analysis.. 100
Table 5-1. Filter set information 149

List of Abbreviations

ADP	adenosine diphosphate
AM	acetoxymethyl
AMP	adenosine monophosphate
APEX	enhanced ascorbate peroxidase
APX	ascorbate peroxidase
ATP	adenosine triphosphate
BioID	proximity-dependent biotin identification
BirA	biotin ligase
BSA	bovine serum albumin
CCP	clathrin-coated pit
CHO	Chinese hamster ovary cell
co-IP	co-immunoprecipitation
DHFR	dihydrofolate reductase
DHLA	dihydrolipoic acid
DIC	differential interference contrast
DNA	deoxyribonucleic acid
DTB	desthiobiotin
EGF	epidermal growth factor
EGFP	enhanced green fluorescent protein
EMARS	enzyme-mediated activation of radical source
ER	endoplasmic recitulum
ESI	electrospray ionization
FACS	fluorescence activated cell sorting
FPR	false-positive rate
FLIP	fluorescence loss in photobleaching
FRAP	fluorescence recovery after photobleaching
HEK	human embryonic kidney cell
HOMO	highest occupied molecular orbital
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
ICAT	isotope-coded affinity tag
ICPL	isotope-coded protein label
IMM	inner mitochondrial membrane
IMS	inter-membrane space
iTRAQ	isotope tags for relative and absolute quantitation
kDa	kilo Dalton
LAP	LplA acceptor peptide
LC	liquid chromatography
LC-MS/MS	liquid chromatography-tandem mass spectrometry
LDL	low-density lipoprotein
LDLR	low-density lipoprotein receptor
LplA	lipoic acid ligase
LUMO	lowest unoccupied molecular orbital
MALDI	matrix-assisted laser desorption/ionization

MS	mass spectrometry
mSA	monovalent streptavidin
msec	millisecond
NHS	N-hydroxysuccinimide
OMM	outer mitochondrial membrane
PALM	photo-activatable localization microscopy
pBirA	promiscuous biotin ligase BirA
PE-Ald	R-PE modified with 4-formylbenzamide
PEG	polyethyleneglycol
RASSL	receptor activated solely by a synthetic ligand
R-PE	R-phycoerythrin
SA	streptavidin
SILAC	stable isotope labeling with amino acids in cell culture
STED	stimulated emission depletion
STORM	stochastic optical reconstruction microscopy
TfnR	transferrin receptor
THF	tetrahydrofuran
TIRF	total internal reflection fluorescence microscopy
TMT	tandem mass tags
UV	ultraviolet

Part A

Proteomic analysis of sub-cellular protein localization

Chapter 1 Introduction to sub-cellular proteomic mapping

1.1 Introduction

An important feature of eukaryotic cells is the exquisite division of cellular space into organellar compartments that vary both in shape and in molecular composition. Because the lipid membrane of these compartments presents a barrier to the diffusion of hydrophilic molecules, functionally distinct aqueous environments can be created within the cell. The biochemical properties of these environments differ considerably. While the cytosolic space is abound with reducing agents, the secretory pathway comprising of the endoplasmic reticulum (ER) and the Golgi apparatus has highly oxidizing redox potential (*1*). Whereas most compartments have nearly neutral pH, the late endosome and lysosome have highly acidic luminal space (*2,3*).

Compartmentalization allows precise control of biochemical reactions within the cell. This is exemplified by the creation of concentration gradients across the membranes that are useful for many physiological processes. The proton gradient across the mitochondrial inner membrane is established by the action of electron transport chain and ultimately drives the synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate (*4,5*). Calcium ion is largely sequestered within the endoplasmic reticulum, and in response to signaling, it can be released into the cytosol to trigger a variety of physiological processes, ranging from muscle contraction to apoptosis (*6,7*). Another advantage of cellular compartmentalization is to dramatically increase the surface area of lipid membrane, which is home to many essential membrane-bound metabolic enzymes. For example, many enzymes involved in lipid metabolism are located on the mitochondrial inner and outer membranes (*8*), and both oxidative phosphorylation and photosynthesis require membrane protein complexes to produce the proton gradient necessary for ATP synthesis (*9,10*).

The function of each organelle is uniquely defined by the set of proteins that it contains. In order to understand how an organelle is engaged in biological pathways, it is often desirable to know its complete molecular inventory. Since the discovery of signal peptides (*11*), it has also become possible to predict the sub-cellular localization of proteins based on their amino acid sequences (*12*). However, predicted localization ultimately needs to be experimentally validated. Currently, two complementary approaches have been employed to verify protein localizations. On the one hand, the top-down approach, involving purification of an organelle of interest and subsequent proteomic analysis, aims to identify the sub-cellular localization information for hundreds of proteins in parallel; on the other hand, the bottom-up approach, involving specific protein labeling and fluorescence imaging, reports detailed spatial information for proteins on an individual basis. In this chapter, we review the currently available techniques for purification and identification of the sub-cellular proteome. Existing techniques for protein labeling and fluorescence microscopy will be reviewed in Chapter 4.

1.2 Existing methods for sub-cellular fractionation

Purification of organelles was traditionally done with sub-cellular fractionation and differential centrifugation. In the first step, cells are homogenized in isotonic buffer solution to disrupt the plasma membrane. Homogenization could be achieved with osmotic shock, freeze-thaw cycle, shearing, grinding, or ultra-sonication. The homogenate then undergoes repeated centrifugation steps to separate organelles based on their sizes and densities. Whereas intact mitochondria can be purified via sub-cellular fractionation approach, the endoplasmic reticulum is too delicate to survive the purification process. Rather, the endoplasmic reticulum is always fragmented into vesicles known as the microsomes, which can be isolated from other cellular components in the fractionation process. Here we summarize our current knowledge about the purification and proteomic analysis of these two compartments.

Sub-cellular fractionation of mitochondria

In one of the earliest studies on mitochondrial proteome, classical cell fractionation of placental tissue was employed to isolate mitochondria (*13*). The mitochondrial suspension was subsequently lysed in denaturing condition, and its protein contents separated by 2-dimensional gel electrophoresis. Gel pieces containing protein spots were excised and analyzed with either protein sequencing or in-gel trypsin digestion followed by mass spectrometry (MS) analysis. Due to limited sensitivity of the method, merely 46 proteins were identified. Another issue with this method is contamination by proteins from other cell components, such as abundant secreted protein (choriomammotropin) in secretion granules whose density matches that of mitochondria, and cytoskeletal proteins (Rho protein and actin) at the mitochondrion-cytosol interface. In addition, bovine serum albumin (BSA) introduced at the initial purification stages were also retained as a contaminant (*13*).

The same strategy of cell fractionation has been employed in all subsequent proteomics studies of mitochondria. As liquid chromatography-tandem mass spectrometry (LC-MS/MS) became a standard tool for protein identification, the coverage has greatly improved due to increased sensitivity. In 2003, Taylor *et al.* identified 615 mitochondrial proteins of human heart using sucrose density gradient centrifugation to purify intact protein complexes (*14*). Mootha *et al.* identified 591 mitochondrial proteins from four mouse tissues based on combined information from proteomics survey and existing genomics annotation. Notably, 163 proteins in this list had not been associated with mitochondria in previous research (*15*).

While it is relatively easy to purify the intact mitochondria, it is challenging to obtain sub-mitochondrial fragments. For instance, the mitochondrial matrix proteome has previously been studied using hypotonic shock to remove the outer membrane from purified rat liver mitochondria (*16*), but only 37 matrix proteins were identified, perhaps due to the harshness of the procedure. Da Cruz *et al.* combined sub-cellular fractionation and membrane extraction with organic acid to selectively purify inner mitochondrial membrane (IMM) proteins from mouse liver (*17*). A total of 182 proteins were identified, including 20 unknown mitochondrial proteins. Zahedi *et al.* isolated outer membrane vesicles from highly purified yeast mitochondria (*18*). Subsequent MS analysis revealed 112 proteins, achieving coverage of about 85% of known outer mitochondrial membrane (OMM) proteins.

Even more challenging is to selectively identify proteins located in the mitochondrial inter-membrane space (IMS). Recently, Vögtle *et al.* presented the first study on yeast IMS proteome (*19*). This experimental design is based on the conserved mechanism of Bax-induced mitochondrial outer membrane permeabilization, which releases IMS proteins into the cytosol. When combined with stable isotope labeling, this approach allows for the specific identification of released IMS proteins. A total of 51 proteins were identified, among which 20 novel IMS protein candidates were discovered. Remarkably, 10 protein candidates have never been associated with mitochondria in previous research.

Sub-cellular fractionation of endoplasmic reticulum

Endoplasmic reticulum forms an extended network of tubules and sheets (*20*). The lumenal space is continuous throughout the endoplasmic reticulum and is connected with the nuclear envelope. The endoplasmic reticulum is morphologically and functionally divided into rough and smooth endoplasmic reticulum subdomains. The rough endoplasmic reticulum is decorated with membrane-bound ribosomes and is the entry point of the secretory pathway. The smooth endoplasmic reticulum, on the other hand, lacks ribosome on the surface and is involved in lipid synthesis and calcium storage.

Due to structural complexity, the endoplasmic reticulum is too delicate to be purified directly. Instead, previous purification schemes disrupted the endoplasmic reticulum into small membrane-bound vesicles called microsomes. Rough and smooth endoplasmic reticulum generates rough and smooth microsomes, respectively. In general, microsomes remain in suspension at low speed centrifugation (10,000 × g for 20 min) but are pelleted at high speed (100,000 × g for 1 hour) (*21*). Sucrose gradient density centrifugation is capable of differentiating between these two types of microsomes (*22*). Rough microsomes can be further purified by removing surface-attached ribosomes with high salt concentration and addition of puromycin (*23-25*). In addition, in order to enrich for integral membrane proteins, peripheral membrane proteins can be removed via treatment with sodium carbonate at basic pH (*6*).

Using this approach, Gilchrist *et al.* compiled a list of over 1,400 proteins in the rat liver secretory pathway including 345 proteins that have previously unknown functions (*22*). Among these, 832 proteins are unique to the endoplasmic reticulum, 193 proteins are unique to the Golgi, and 405 are in both the endoplasmic reticulum and the Golgi. More recently, Song *et al.* performed quantitative proteomic survey of mouse liver and identified 903 proteins in the rough endoplasmic reticulum and 1042 proteins in the smooth endoplasmic reticulum (*26*). However, comparison between these two proteomics studies showed that merely 662 proteins were found in common. Even considering that these endoplasmic reticulum proteomes are derived from different animals, this poor overlap still suggests lack of reliability in extracting the endoplasmic reticulum proteome.

Limitations of sub-cellular fractionation

Density centrifugation of organelles is often plagued by artifacts. On the one hand, false positive hits arise because of incomplete purification or contaminants introduced into the sample; on the other hand, if the purification condition is too harsh, then loosely interacting proteins could be lost, resulting in false negatives. Protein correlation profiling has been employed to battle against contaminating proteins in sub-cellular fractionation (*27*). This strategy utilizes known protein markers for the protein complexes or organelles that are being purified, and reject proteins whose abundance distribution profiles are different from that of the markers.

Another drawback of organelle purification processes is the requirement of intensive labor and extensive period of time. During hours of purification, the physiology of organelles could change significantly. Furthermore, many regions of interest in the cell – such as the synaptic cleft or the contact site between mitochondria and the endoplasmic reticulum – simply cannot be purified using conventional methods and therefore their proteomes are largely unknown.

1.3 Proximity-dependent promiscuous protein labeling

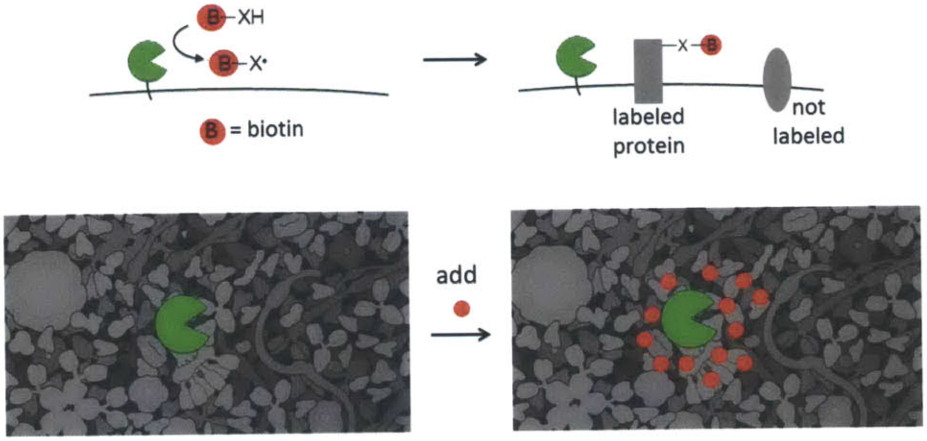


Figure 1-1. Scheme of proximity-dependent promiscuous protein labeling.
The green packman represents the labeling source, which in our case is an engineered enzyme. It catalyzes the conversion of a stable compound into a highly reactive intermediate which then diffuses away from the active site and forms a covalent bond with nearby proteins. Ideally, such intermediate is short lived, so that it does not diffuse far from the source before getting inactivated. Therefore, inside a living cell, chemical labeling occurs in a proximity-dependent manner.

We sought to develop new methods for proteomic mapping that preserves spatial information yet circumvents the many drawbacks of traditional cellular lysis and purification. In this alternative approach, the proteome of interest is tagged with a chemical handle such as biotin while the cell is still alive, with all cellular membranes, macromolecular complexes, and spatial arrangements preserved. Ideally, the yield of such chemical reactions should be strictly dependent on the proximity between the labeling source and endogenous protein targets, and it should be able to label any proteins, regardless of their amino acid sequence, three-dimensional structure and physicochemical properties such as size, charge, and hydrophobicity. The scheme of such proximity-dependent promiscuous labeling reaction is outlined in Figure 1-1. This strategy has been employed in previous research, using various labeling sources to generate reactive intermediates. These are reviewed below.

Promiscuous biotin ligase

One candidate promiscuous protein labeling source is engineered biotin ligase. The Cronan group discovered that the R118G mutant form of *Escherichia coli* biotin ligase BirA is capable of catalyzing the formation of biotin-adenylate ester (biotin-AMP) and subsequently releasing it from the active site (*28*). Roux *et al.* further developed this method by expressing BirA R118G mutant protein in living mammalian cells. Due to its high reactivity toward nucleophilic groups such as the side chain of lysine residue, biotin-AMP reacts with proteins proximal to the enzyme, resulting in covalent modification of these proteins with biotin. These biotinylated proteins could then be selectively recovered with streptavidin-based affinity purification, and subsequently analyzed by Western blot or identified by mass spectrometry. This labeling technique was termed BioID (proximity-dependent biotin identification) (*29*).

BioID has been demonstrated to be useful for analyzing the proteome of nuclear lamina. In this study, promiscuous BirA was fused to lamin A, a component of the nuclear lamina, and expressed in a human cell line (*29*). MS identified about 100 protein candidates, including both known protein components of the nuclear lamina and several uncharacterized proteins. SLAP75 belongs to the latter group, and follow-up immunofluorescence study confirmed its localization to the nuclear envelope.

However, BioID suffers from slow labeling kinetics and poor spatial resolution. It requires at least 6 hours of labeling (*29*). This prolonged incubation could cause biological perturbation, because artificially modifying proteins with biotin could interfere with protein functions such as blocking protein-protein interactions. Promiscuous covalent modification of thousands of proteins in a living cell could therefore alter the physiology of the organelle under investigation or even be toxic to cells. In addition, the poor temporal resolution of this method makes it impossible to analyze any rapid changes in the organellar proteome, such as acute response to drug treatment. Furthermore, the proposed mechanism of promiscuous labeling with biotin-AMP implies a long half-life and hence a large labeling radius (*28*). The authors claimed that the labeling radius is within 20 nm, based on the examination of proteomics data. However, it contradicts with our knowledge of the stability of acyl-adenylate. For example, the hydrolysis rate constant of glutaminyl-adenylate in buffered aqueous solution was measured to be merely 0.7×10^{-3} s^{-1}, corresponding to a half-life of about 10^3 sec (*30*). Such long life time would allow the intermediate to diffuse across the entire cell. It might be still useful for mapping organelle proteomes, because the adenylate is membrane-impermeant due to its negative charges. However, it would be implausible to map non-membrane-bound sub-cellular regions or protein complexes. Finally, the selectivity toward primary amines poses another limitation of BioID, because this creates a bias toward proteins with more surface-accessible lysine residues. The labeling reaction yield would also be lower at acidic environments as fewer lysine amino groups are deprotonated.

Enzyme-mediated activation of radical source

Horseradish peroxidase (HRP)-catalyzed generation of nitrene from aryl azide is another candidate. It is well-known that aryl azide can be photo-activated into a highly reactive and short-lived nitrene radical, which can form covalent bond with nearby biomolecules via insertion reactions (*31*). Recently, Honke *et al.* reported that in addition to ultraviolet (UV) light, HRP is also capable of activating arylazide into nitrene (*32*). By tethering aryl azide to biotin and targeting HRP to a specific cellular location via antibody, it is now possible to label the surrounding proteins with an affinity tag. Thereafter, these labeled proteins can be enriched and identified using streptavidin-based affinity purification. This technique was termed EMARS (enzyme-mediated activation of radical source).

To demonstrate the power of EMARS labeling, HRP was targeted via antibody to various cell surface proteins including growth factor receptors and integrin. EMARS labeling and subsequent analysis by antibody array and MS identified proteins co-clustering with these markers on the cell surface (*32-34*). HRP has also been targeted to lipid raft markers such as gangliosides GM1 and GD3, and fluorescein-conjugated aryl azide was used to fluorescently labeling cell surface molecules within the lipid raft. Enrichment with anti-fluorescein antibody-coated resins and MS analysis identified membrane proteins and secreted proteins as candidates in the lipid raft microdomain (*35,36*).

Electron microscopy showed that the labeled proteins are positioned within 200 – 300 nm of the HRP enzyme (*32*). This small radius is consistent with the short lifetime of the nitrene radical. It is noteworthy that the EMARS approach bears resemblance to the commercial "tyramide signal amplification" method (*37*), which uses antibody-targeted HRP to generate fluorophore-tyramide deposits on fixed cells for imaging. This method is also known to preserve the spatial information of labeling, with minimal diffusion of fluorophore-tyramide signal beyond the location of HRP enzyme. One limitation of this method is that HRP is inactive when expressed in the mammalian cytosol (*38*). Therefore, all the applications so far have been on cell surface protein clusters, and it is not possible to extend this technique to labeling cytosolic proteins.

Direct crosslinking to endogenous proteins

A special case of proximity-dependent labeling is protein crosslinking, in which a protein of interest is introduced into cells as bait. In order to discover its endogenous interaction partners (preys), a crosslinking reagent was added to covalently conjugate the bait to nearby proteins. Affinity purification of the bait also pulls down the crosslinked preys, which are subsequently identified. In this scheme, the bait protein itself serves as the label.

A variety of reagents have been used for protein crosslinking experiments. Formaldehyde, *N*-hydroxysuccinimidyl esters and carbodiimides are reactive toward primary amines, and are therefore useful for crosslinking proteins via the side chain of lysine residues or the *N*-termini. Photocrosslinkers such as benzophenone (*39,40*), aryl azide (*41*), and diazirines (*42,43*) have been incorporated into proteins in living cells through unnatural amino acid mutagenesis. Subsequent UV irradiation generates highly reactive intermediates (triplet-state benzophenone, nitrene, or carbene, respectively) that enable crosslinking of proteins of interest to other endogenous targets. The main concern regarding photocrosslinking is the low yield, which necessitates long exposure time.

A variation of the photocrosslinker strategy is the development of targeted releasable affinity probe, in which a benzophenone probe is targeted to the bait via a tetra-cysteine motif (*44*). After photo-induced crosslinking to endogenous proteins, this group can be dissociated from the bait with treatment of excess dithiols. The preys are therefore released from the bait and left with a covalent modification that can be used for affinity purification. This technique has been applied to cultured mammalian cells and it revealed the interaction between the membrane protein phospholamban and the extracellular matrix protein fibronectin (*16*). The ability to release the prey from the bait reduces sample complexity, which facilitates MS analysis. However, the caveat of using the tetra-cysteine motif is the oxidation and palmitoylation of cysteine residues that could block probe binding (*45*).

Transition metal-mediated oxidative protein crosslinking is another possibility (*46*). Pioneered by Kodadek *et al.*, this technique capitalizes on the oxidizing power of high-valence transition metal complexes and primarily targets electron-rich residues such as tyrosine, tryptophan and cysteine (*47,48*). In the case of tyrosine, one-electron oxidation generates a tyrosyl radical which rapidly couples to an adjacent protein, often with another tyrosine residue (*49,50*).

Various transition metal complexes have been employed for oxidative crosslinking reactions. Highly reactive high-valence metal complexes are often prepared from stable, low-valence complexes through either chemical oxidation or photo-oxidation. One prominent example of chemical oxidation is the nickel-peptide complex. In the presence of magnesium monoperoxyphthalate or potassium peroxomonosulfate as the stoichiometric oxidant, Ni(II) complexed with a tripeptide Gly-Gly-His (*51*) or a hexapeptide His_6 (*52,53*) is oxidized into Ni(III) oxo, which is highly reactive and mediates crosslinking. In addition, by introducing these Ni(II)-binding peptides to proteins through genetic fusion, it is possible to target redox active nickel complexes in recombinant proteins (*23,26,27,54*). This method has been applied to study the interaction between actin and myosin (*55*).

In the case of photo-oxidation, Ru(II)-tris(bipyridine) (*21*) and Pd(II)-tetrakis-(4-methylpyridyl) porphyrin (*56*) have been utilized. In the presence of electron acceptors such as ammonium persulfate or $Co(III)(NH_3)_5Cl_2^+$, light irradiation excites an HOMO electron in the metal complex to a LUMO state which is then captured by the electron acceptor. The resulting high-valence metal complex then mediates the crosslinking process by taking an electron from tyrosine residues, in the same way as oxidized nickel-peptide complexes. Photo-induced crosslinking has been used in the proteomic study of the 20S proteasome (*57*), and the covalent labeling of G-protein coupled receptors with their ligands on the cell surface (*58*).

The crosslinking yield needs to be carefully optimized. While too little crosslinking fails to capture weak and transient interactions, over-crosslinking could generate insoluble protein precipitate and/or block epitope tag recognition.

1.4 Existing mass spectrometry techniques for proteomics

Mass spectrometry (MS) is an analytical tool for measuring the mass-to-charge ratio of ions. In history, the most primitive model was built by physicist Wien in 1898, when he applied electric and magnetic fields to deflect a stream of ionized gas that later came to be known as the proton (*59*). This technique was further developed in the early 1900s by Thomson and his student Aston, who designed the first mass spectrograph to discover different isotopes of neon (*60,61*), and Dempster, who introduced both the magnetic sector field mass spectrometer and electron impact ionization source (*62*). In the following decades, various mass analyzers have been developed, and MS has become a powerful tool for both physicists and chemists. However, the application of MS to analyzing proteins and nucleic acids had been limited, mainly because traditional ionization sources were too harsh for these biomolecules.

The introduction of soft ionization techniques such as electrospray ionization (ESI) (*63*) and matrix-assisted laser desorption/ionization (MALDI) (*64,65*) has revolutionized the field of bioanalytical chemistry. In ESI, highly charged droplets containing the analyte are generated by passing the liquid phase through the tip of a capillary tube that is held at high voltage. The droplets are subsequently evaporated, leading to the formation of ionized analyte in the gas phase. In MALDI, the analyte is first dispersed in a solid matrix containing UV light-absorbing molecules. Subsequent UV laser irradiation causes ablation of the matrix surface and generates a plume of ionized analyte. Both ESI and MALDI are compatible with large biomolecules such as proteins and nucleic acids, to the extent that they could even preserve non-covalent interactions during ionization (*66,67*).

Over the past decades, MS has gained increasing popularity as a major tool for proteomics studies. Recent developments in mass analyzers such as the Orbitrap technology have achieved unprecedented sensitivity, accuracy (1 – 2 ppm), and mass resolution (200,000) (*68,69*). At the same time, various liquid chromatography techniques have been improved to efficiently couple with MS analysis. Tandem MS allows simultaneous detection of the parent ion and deduction of its sequence. This section reviews the basics of mass spectrometry in proteomic experiments, with a focus on quantitative MS approaches.

Protein identification by mass spectrometry

Modern liquid chromatography-tandem MS (LC-MS/MS) is highly efficient in detecting peptide fragments within a complex sample. Proteins are generally too large for this analysis, and therefore need to be broken down to peptides first. Enzymatic digestion is a favored approach to break down proteins, because the fragment length is optimal for LC-MS/MS and the cleavage site is often predictable. For example, trypsin has been routinely used to cleave proteins at the *C*-termini of basic amino acid residues lysine and arginine. In this way, peptides are identified first, and protein identifications are then deduced based on peptide sequences. Availability of genome information of the organism under investigation could help map peptide fragments onto known open-reading frames, thereby greatly facilitating the protein identification process. This approach is termed "bottom-up proteomics" (*70,71*).

Protease digestion can be performed directly in the solution of protein sample. Alternatively, proteins can be first separated by gel electrophoresis, in either one dimension or two dimensions. Thereafter, gel spots containing proteins can be sliced out and treated individually with protease. After this in-gel digestion, the resulting peptides can be recovered through extraction. Although in-gel digestion is more time-consuming than the in-solution form, its main advantage is the reduction of sample complexity, which helps improve the coverage of proteomics study. A hybrid method utilizing micro-reactors has been employed to combine the resolving power of in-gel approach and the speed of in-solution approach. In this method, trypsin is immobilized within a solid matrix so that enzymatic digestion can be performed on-line to the liquid chromatography (*72-74*). The main limitation of this approach is the requirement of sophisticated instrumentation.

Digested peptides are separated by liquid chromatography (LC), usually on a reverse phase column. The LC is directly coupled to tandem mass spectrometer. The first stage mass spectrometer measures the mass-to-charge ratio of ionized peptides, and then transports peptides of interest to the second stage mass spectrometer. These selected peptide ions are mixed in the gas phase with inert gas molecules, and collision-induced dissociation causes peptide ions to fragment, most frequently at the peptide bond. Because dissociation can occur at different sites, a series of fragment ions are generated and analyzed. Sequence information of each parent peptide ion can thus be deduced from this measurement.

Quantitative mass spectrometry

While extremely powerful for identifying proteins, traditional MS is not inherently good at quantifying proteins in a complex sample. The MS peak intensity is affected by many parameters other than protein or peptide abundance. These parameters include size, charge state, and hydrophobicity of proteins or peptides. Internal standards can be added to the sample to calibrate the MS instrument, but this has to be done on an individual peptide basis, and it is challenging to extend this on a proteomic level. Recently, using stable isotope labels, various MS methods have been developed to allow quantitation of relative protein abundance between multiple samples. Proteins labeled with different isotope tags are pre-mixed before any sample processing occurs. Because these proteins differ only in the isotope that they carry, their physical and chemical properties are identical, so that they are affected in exactly the same manner in subsequent biochemical (protein enrichment, protease digestion, peptide extraction, etc.) and instrumental (chromatographic separation, ionization, MS response) analyses. This effectively minimizes quantitation errors. MS analysis resolves multiplet peaks, with each corresponding to a unique monoisotopic species.

In cultured cell lines, it is often convenient to metabolically incorporate isotope labels into proteins. Stable isotope labeling with amino acids in cell culture, or SILAC, has been widely used for this purpose (*75*). Among 20 standard amino acids, arginine and lysine are favored candidates for isotope labeling for two reasons: 1) trypsin digestion is specific to the arginine and lysine residues; and 2) lysine (an essential amino acid) and arginine do not easily convert to other standard amino acids during metabolism (excess arginine can be converted to proline via the arginase catabolism pathway (*76*), but this can be prevented by careful titration of arginine). Together, these ensure that each tryptic peptide (except for the most *C*-terminal peptide) carries one and only one isotope-labeled residue at the *C*-terminus, which greatly simplifies data analysis. ^{13}C and ^{15}N are commonly used to label lysine ($^{13}C_6$-Lys, or $^{13}C_6$,$^{15}N_2$-Lys) and arginine ($^{13}C_6$-Arg, or $^{13}C_6$,$^{15}N_4$-Arg) residues. After culturing cells in SILAC media for 6 – 8 doublings, more than 90% of proteins can be labeled (*75*). Despite its success with immortalized cell cultures, SILAC is generally not considered as cost-effective when labeling cells in live animals, such as worms, fruit flies and rats. Typically, SILAC labeling can be applied up to three conditions (light state – no labeling, medium state – ^{13}C label only, and heavy state – both ^{13}C and ^{15}N labels).

Alternative labeling methods are required if SILAC is not a viable solution. For example, ^{18}O can be efficiently incorporated to the carboxyl termini of peptides if enzymatic digestion is performed in $H_2{}^{18}O$ (*77,78*). Isotope-coded affinity tag (ICAT) method has been developed to selectively modify cysteine residues with deuterated reagents, such as biotin derivatives with a thiol-reactive group, for both affinity purification and isotopic labeling (*79*). Similarly, isotope-coded protein label (ICPL) employs *N*-hydroxysuccinimide chemistry to modify the *N*-terminal and lysine residue side chain amino groups with an isotope tag (*80*). Variations of this method include isotope tags for relative and absolute quantitation (iTRAQ) and tandem mass tags (TMT), which employs the same derivatization chemistry but utilizes isobaric tags that provide fragmentation signatures during the tandem MS analysis (*81,82*). Finally, the *C*-terminal and acidic amino acid (Asp and Glu) side chain carboxyl groups can be labeled via esterification with deuterated alcohols (*83,84*). These techniques are collectively referred to as post-biosynthetic labeling (*85*).

Quantitation of peptide abundance is achieved by taking the ratio of integrated peak intensity of differentially labeled peptides. Protein level abundance information is then inferred from peptide ratios, usually by taking the mean or median value of peptides belonging to the same protein. This step can now be automated using commercially available software, such as MaxQuant (*86,87*).

1.5 Conclusion

Here we have reviewed the current technology involved in proteomic studies of cells and organelles. All of these experiments consist of two steps: purification of proteins of interest, and subsequent identification of these proteins. Thanks to advances in MS technology, the second step can be routinely done, and is limited only by the sensitivity and resolving power of the instrument. The challenge remains in the first step, where the purification scheme varies from case to case. Proximity-dependent promiscuous protein labeling offers a novel way of addressing the problem as the physical purification step can be bypassed and replaced with a chemical reaction. However, the currently available labeling reactions do not fulfill all the requirements.

We envision that an enzyme-catalyzed free radical reaction could be beneficial, due to the short lifetime and reactive nature of free radicals in the cellular environment. In the following two chapters, we discuss in detail the discovery, improvement, characterization and applications of a class of heme peroxidase for promiscuous protein labeling.

References

1 Merksamer, P. I., Trusina, A. & Papa, F. R. Real-time redox measurements during endoplasmic reticulum stress reveal interlinked protein folding functions. *Cell* **135**, 933-947 (2008).

2 Nilsson, C., Kagedal, K., Johansson, U. & Ollinger, K. Analysis of cytosolic and lysosomal pH in apoptotic cells by flow cytometry. *Methods Cell Sci* **25**, 185-194 (2003).

3 Overly, C. C., Lee, K. D., Berthiaume, E. & Hollenbeck, P. J. Quantitative measurement of intraorganelle pH in the endosomal-lysosomal pathway in neurons by using ratiometric imaging with pyranine. *Proc Natl Acad Sci U S A* **92**, 3156-3160 (1995).

4 Perry, S. W., Norman, J. P., Barbieri, J., Brown, E. B. & Gelbard, H. A. Mitochondrial membrane potential probes and the proton gradient: a practical usage guide. *Biotechniques* **50**, 98-115 (2011).

5 Nakamoto, R. K., Ketchum, C. J. & al-Shawi, M. K. Rotational coupling in the F0F1 ATP synthase. *Annu Rev Biophys Biomol Struct* **28**, 205-234 (1999).

6 Somlyo, A. V., Bond, M., Somlyo, A. P. & Scarpa, A. Inositol trisphosphate-induced calcium release and contraction in vascular smooth muscle. *Proc Natl Acad Sci U S A* **82**, 5231-5235 (1985).

7 Scorrano, L. *et al.* BAX and BAK regulation of endoplasmic reticulum Ca2+: a control point for apoptosis. *Science* **300**, 135-139 (2003).

8 McGarry, J. D. & Brown, N. F. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. *Eur J Biochem* **244**, 1-14 (1997).

9 Hatefi, Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu Rev Biochem* **54**, 1015-1069 (1985).

10 Allen, J. Photosynthesis of ATP-electrons, proton pumps, rotors, and poise. *Cell* **110**, 273-276 (2002).

11 Blobel, G. Protein targeting (Nobel lecture). *Chembiochem* **1**, 87-102 (2000).

12 Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953-971 (2007).

13 Rabilloud, T. *et al.* Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. *Electrophoresis* **19**, 1006-1014 (1998).

14 Taylor, S. W. *et al.* Characterization of the human heart mitochondrial proteome. *Nat Biotechnol* **21**, 281-286 (2003).

15 Mootha, V. K. *et al.* Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629-640 (2003).

16 Forner, F., Foster, L. J., Campanaro, S., Valle, G. & Mann, M. Quantitative proteomic comparison of rat mitochondria from muscle, heart, and liver. *Mol Cell Proteomics* **5**, 608-619 (2006).

17 Da Cruz, S. *et al.* Proteomic analysis of the mouse liver mitochondrial inner membrane. *The Journal of biological chemistry* **278**, 41566-41571 (2003).

18 Zahedi, R. P. *et al.* Proteomic analysis of the yeast mitochondrial outer membrane reveals accumulation of a subclass of preproteins. *Mol Biol Cell* **17**, 1436-1450 (2006).

19 Voegtle, F. N. *et al.* Intermembrane space proteome of yeast mitochondria. *Mol Cell Proteomics* (2012).

20 Voeltz, G. K., Prinz, W. A., Shibata, Y., Rist, J. M. & Rapoport, T. A. A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* **124**, 573-586 (2006).

21 Chen, X., Karnovsky, A., Sans, M. D., Andrews, P. C. & Williams, J. A. Molecular characterization of the endoplasmic reticulum: insights from proteomic studies. *Proteomics* **10**, 4040-4052 (2010).

22 Gilchrist, A. *et al.* Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265-1281 (2006).

23 Chen, X. *et al.* Quantitative organellar proteomics analysis of rough endoplasmic reticulum from normal and acute pancreatitis rat pancreas. *J Proteome Res* **9**, 885-896 (2010).

24 Adelman, M. R., Sabatini, D. D. & Blobel, G. Ribosome-membrane interaction. Nondestructive disassembly of rat liver rough microsomes into ribosomal and membranous components. *The Journal of cell biology* **56**, 206-229 (1973).

25 Zahedi, R. P. *et al.* Analysis of the membrane proteome of canine pancreatic rough microsomes identifies a novel Hsp40, termed ERj7. *Proteomics* **9**, 3463-3473 (2009).

26 Song, Y. *et al.* Quantitative proteomic survey of endoplasmic reticulum in mouse liver. *J Proteome Res* **9**, 1195-1202 (2010).

27 Foster, L. J. *et al.* A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187-199 (2006).

28 Choi-Rhee, E., Schulman, H. & Cronan, J. E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein science : a publication of the Protein Society* **13**, 3043-3050 (2004).

29 Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology* **196**, 801-810 (2012).

30 Gruic-Sovulj, I., Uter, N., Bullock, T. & Perona, J. J. tRNA-dependent aminoacyl-adenylate hydrolysis by a nonediting class I aminoacyl-tRNA synthetase. *The Journal of biological chemistry* **280**, 23978-23986 (2005).

31 Torres, M. J., Zayas, J. & Platz, M. S. A Formal CH Insertion Reaction of an Aryl Nitrene into an Alkyl CH Bond - Implications for Photoaffinity-Labeling. *Tetrahedron Lett* **27**, 791-794 (1986).

32 Kotani, N. *et al.* Biochemical visualization of cell surface molecular clustering in living cells. *Proc Natl Acad Sci U S A* **105**, 7405-7409 (2008).

33 Kotani, N., Ishiura, Y., Yamashita, R., Ohnishi, T. & Honke, K. Fibroblast Growth Factor Receptor 3 (FGFR3) Associated with the CD20 Antigen Regulates the Rituximab-induced Proliferation Inhibition in B-cell Lymphoma Cells. *The Journal of biological chemistry* **287**, 37109-37118 (2012).

34 Yamashita, R., Kotani, N., Ishiura, Y., Higashiyama, S. & Honke, K. Spatiotemporally-regulated interaction between beta1 integrin and ErbB4 that is involved in fibronectin-dependent cell migration. *J Biochem* **149**, 347-355 (2011).

35 Jiang, S. *et al.* A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. *Proteomics* **12**, 54-62 (2012).

36 Hashimoto, N. *et al.* Proteomic analysis of ganglioside-associated membrane molecules: Substantial basis for molecular clustering. *Proteomics* **12**, 3154-3163 (2012).

37 van Gijlswijk, R. P. M. *et al.* Fluorochrome-labeled tyramides: Use in immunocytochemistry and fluorescence in situ hybridization. *Journal of Histochemistry & Cytochemistry* **45**, 375-382 (1997).

38 Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically-encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143-1148 (2012).

39 Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **99**, 11020-11024 (2002).

40 Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2**, 201-206 (2005).

41 Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *Journal of the American Chemical Society* **124**, 9026-9027 (2002).

42 Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2**, 261-267 (2005).

43 Tippmann, E. M., Liu, W., Summerer, D., Mack, A. V. & Schultz, P. G. A genetically encoded diazirine photocrosslinker in Escherichia coli. *Chembiochem* **8**, 2210-2214 (2007).

44 Yan, P. *et al.* A targeted releasable affinity probe (TRAP) for in vivo photocrosslinking. *Chembiochem* **10**, 1507-1518 (2009).

45 Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224 (2006).

46 Antos, J. M. & Francis, M. B. Transition metal catalyzed methods for site-selective protein modification. *Curr Opin Chem Biol* **10**, 253-262 (2006).

47 Fancy, D. A. & Kodadek, T. Chemistry for the analysis of protein-protein interactions: rapid and efficient cross-linking triggered by long wavelength light. *Proc Natl Acad Sci U S A* **96**, 6020-6024 (1999).

48 Kodadek, T., Duroux-Richard, I. & Bonnafous, J. C. Techniques: Oxidative cross-linking as an emergent tool for the analysis of receptor-mediated signalling events. *Trends Pharmacol Sci* **26**, 210-217 (2005).

49 Brown, K. C., Yu, Z., Burlingame, A. L. & Craik, C. S. Determining protein-protein interactions by oxidative cross-linking of a glycine-glycine-histidine fusion protein. *Biochemistry* **37**, 4397-4406 (1998).

50 Fancy, D. A. & Kodadek, T. A critical role for tyrosine residues in His6Ni-mediated protein cross-linking. *Biochem Biophys Res Commun* **247**, 420-426 (1998).

51 Brown, K. C., Yang, S. H. & Kodadek, T. Highly specific oxidative cross-linking of proteins mediated by a nickel-peptide complex. *Biochemistry* **34**, 4733-4739 (1995).

52 Fancy, D. A., Melcher, K., Johnston, S. A. & Kodadek, T. New chemistry for the study of multiprotein complexes: the six-histidine tag as a receptor for a protein crosslinking reagent. *Chem Biol* **3**, 551-559 (1996).

53 Fancy, D. A. & Kodadek, T. Site-directed oxidative protein crosslinking. *Tetrahedron* **53**, 11953-11960 (1997).

54 Amini, F., Kodadek, T. & Brown, K. C. Protein affinity labeling mediated by genetically encoded peptide tags. *Angew Chem Int Ed Engl* **41**, 356-359 (2002).

55 Van Dijk, J. *et al.* Conformational changes in actin-myosin isoforms probed by Ni(II).Gly-Gly-His reactivity. *J Muscle Res Cell Motil* **25**, 527-537 (2004).

56 Kim, K., Fancy, D. A., Carney, D. & Kodadek, T. Photoinduced protein cross-linking mediated by palladium porphyrins. *Journal of the American Chemical Society* **121**, 11896-11897 (1999).

57 Denison, C. & Kodadek, T. Toward a general chemical method for rapidly mapping multi-protein complexes. *J Proteome Res* **3**, 417-425 (2004).

58 Duroux-Richard, I. *et al.* Crosslinking photosensitized by a ruthenium chelate as a tool for labeling and topographical studies of G-protein-coupled receptors. *Chemistry & Biology* **12**, 15-24 (2005).

59 Wien, W. *Verl. Phys. Ges. Berlin* **17**, 10 (1898).

60 Thomson, J. J. Rays of positive electricity. *Philos Mag* **20**, 752-767 (1910).

61 Aston, F. W. The constitution of atmospheric neon. *Philos Mag* **39**, 449-455 (1920).

62 Dempster, A. J. A new method of positive ray analysis. *Phys Rev* **11**, 316-325 (1918).

63 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).

64 Tanaka, K. W., H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of flight Mass Spectrometry. *Rapid Commun Mass Spectrom* **2**, 151-153 (1988).

65 Karas, M. & Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical Chemistry* **60**, 2299-2301 (1988).

66 Katta, V. & Chait, B. T. Observation of the Heme Globin Complex in Native Myoglobin by Electrospray-Ionization Mass-Spectrometry. *Journal of the American Chemical Society* **113**, 8534-8535 (1991).

67 Ganem, B., Li, Y. T. & Henion, J. D. Detection of Noncovalent Receptor Ligand Complexes by Mass-Spectrometry. *Journal of the American Chemical Society* **113**, 6294-6296 (1991).

68 Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J Am Soc Mass Spectrom* **17**, 977-982 (2006).

69 Makarov, A. *et al.* Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal Chem* **78**, 2113-2120 (2006).

70 Chait, B. T. Chemistry. Mass spectrometry: bottom-up or top-down? *Science* **314**, 65-66 (2006).

71 Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).

72 Liu, J. Y. *et al.* On-chip enzymatic microreactor using trypsin-immobilized superparamagnetic nanoparticles for highly efficient proteolysis. *J Chromatogr A* **1176**, 169-177 (2007).

73 Bonneil, E., Mercier, M. & Waldron, K. C. Reproducibility of a solid-phase trypsin microreactor for peptide mapping by capillary electrophoresis. *Anal Chim Acta* **404**, 29-45 (2000).

74 Slysz, G. W. & Schriemer, D. C. Blending protein separation and peptide analysis through real-time proteolytic digestion. *Analytical Chemistry* **77**, 1572-1579 (2005).

75 Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

76 Ong, S. E., Kratchmarova, I. & Mann, M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* **2**, 173-181 (2003).

77 Yao, X., Freas, A., Ramirez, J., Demirev, P. A. & Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* **73**, 2836-2842 (2001).

78 Reynolds, K. J., Yao, X. & Fenselau, C. Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J Proteome Res* **1**, 27-33 (2002).

79 Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999 (1999).

80 Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4-15 (2005).

81 Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**, 1154-1169 (2004).

82 Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).

83 Goodlett, D. R. *et al.* Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun Mass Sp* **15**, 1214-1221 (2001).

84 Salomon, A. R. *et al.* Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc Natl Acad Sci U S A* **100**, 443-448 (2003).

85 Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031 (2007).

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

87 Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).

Chapter 2 Promiscuous protein labeling with ascorbate peroxidase

Work described in this chapter has not yet been published. Most of the experiments described were performed in collaboration with Dr. Hyun-Woo Rhee, who synthesized various biotin-phenol compounds and characterized the activity of ascorbate peroxidase with biochemical assays. The focus of my work was on characterizing the spatial specificity of ascorbate peroxidase labeling.

2.1 Introduction

As outlined in Chapter 1, we envision that a promiscuous labeling enzyme would be useful for acquiring proteomic information. For this method to work, the following criteria have to be met:

1. The enzyme should be genetically targetable to and be active at different parts of the cell. The enzyme should withstand both oxidizing (such as the secretory pathway) and reducing (such as the cytosol) environments as well as pH levels ranging from around 4.5 (lysosome (*1*)) to 8.0 (mitochondrial matrix (*2*)).
2. The substrate of the enzyme should be a membrane-permeant small molecule that can be added exogenously and triggers the reaction. Ideally, the enzyme does not recognize endogenous metabolites present in mammalian cells, or its activity can be temporally modulated by small molecules.
3. The substrate contains a functional handle for subsequent enrichment. Biotin is commonly used as an affinity tag, which is recognized by streptavidin with femtomolar dissociation constant. Alternatively, functional groups enabling bio-orthogonal reactions can be used as a handle for enrichment. For example, if the substrate contains an azido (or alkynyl) group, then the labeled proteins can be selectively immobilized onto an alkyne-derivatized (or azide-derivatized) resin surface via click-chemistry (*3*).
4. The enzyme converts the substrate into a highly reactive intermediate with short half-life, such that it has the potential to label any endogenous proteins and has small diffusion radius.

Although to our knowledge no naturally occurring enzyme-substrate pair fulfills these conditions, a class of free radical generating enzymes appears to be promising candidates. In biological systems, the free radical is a two-edged sword. On the one hand, they have the potential to cause deleterious effects to cells, such as DNA cleavages, and therefore their formation as metabolic byproducts is under tight control (*4*). As a result, cells have evolved a sophisticated system to rapidly clear away free radicals. On the other hand, the power of free radicals can also be harnessed in enzymatic reactions to achieve otherwise difficult chemical transformations. This has been demonstrated in the case of deoxyribonucleotide biosynthesis (*5*). Because free radicals have extremely short life-time and high reactivity, we envision that through protein engineering, it would be possible to identify enzyme mutants and synthetic substrates that are useful for promiscuous protein labeling. Among various free radical-generating enzymes, we were particularly interested in a class of enzyme called heme peroxidases. In these enzymes, because the small molecule substrate is non-covalently bound to the active site, the resulting radical could be released to attack nearby biomolecules.

Heme peroxidases are categorized into three classes based on sequence homology, cellular localization and biological origin (*6*). Class I is cytosolic peroxidases. Examples in this group are cytochrome c peroxidase and ascorbate peroxidase. Class II is fungal secretory peroxidases, which contains lignin peroxidase and manganese peroxidase. Class III is secreted plant peroxidases, which contains horseradish peroxidase. Both class II and class III peroxidases contain four disulfide bonds and two conserved calcium ions binding sites. Among these enzymes, horseradish peroxidase (HRP) is known for recognizing a variety of substrates and has found broad applications in biotechnology, ranging from chemiluminescence, enzyme-linked immunosorbent assay (with 3,3',5,5'-tetramethylbenzidine), to electron microscopy contrast generation (with 3,3'-diaminobenzidine). Unfortunately, HRP is inactive when expressed in the cytosol.

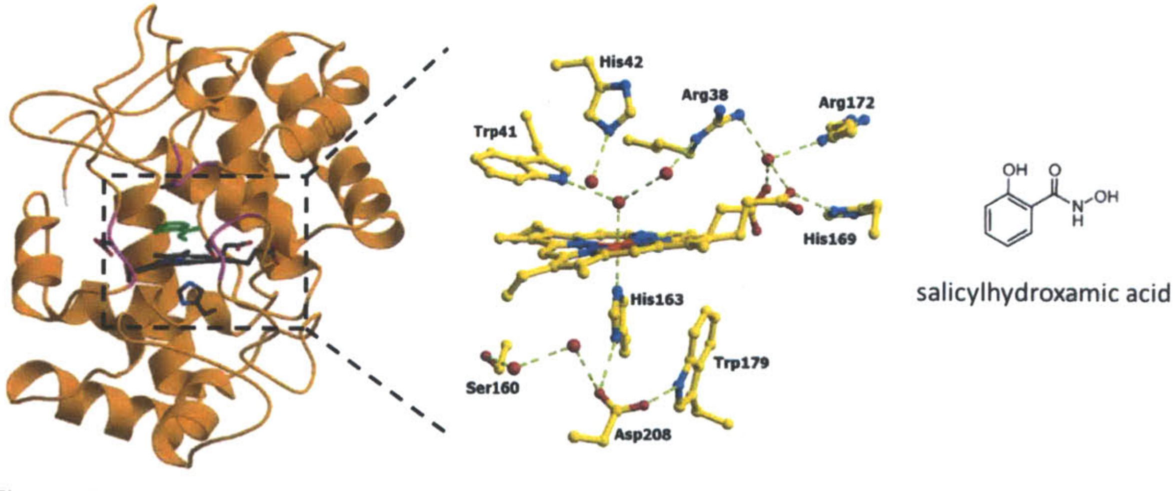


Figure 2-1. Structure of soybean cytosolic ascorbate peroxidase in complex with an aromatic substrate.
Left: overall structure of ascorbate peroxidase monomer, with heme co-factor colored in black, distal histidine in blue and bound aromatic substrate (salicylhydroxamic acid, structure shown on the right) in green. Middle: close up view of residues around the heme active site. This figure is adapted from references (*7*) and (*8*). PDB ID: 1V0H

In an effort to search for peroxidases as electron microscopy reporters, Jeffrey Martell in our lab found that pea ascorbate peroxidase (APX), a class I heme peroxidase, is active toward a panel of phenol-containing substrates, such as guaiacol and Amplex Red (*9*). Unlike HRP, APX lacks disulfide bonds and is active in the mammalian cytosol and other organelles. The active site of ascorbate peroxidase features a non-covalently bound heme molecule (Figure 2-1). The catalytic cycle is depicted in Figure 2-2.

In the case of phenol-containing substrates, a phenoxyl radical is formed. Previous work has shown that phenoxyl radicals are short lived (<1 msec (*10*)) and have limited diffusion radius prior to labeling (<14 nm (*11,12*)). Quenching of phenoxyl radical occurs either through hydrogen atom abstraction from a reducing agent or via covalent bond formation upon collision with another free radical. In aqueous solution, these reactions are often diffusion-limited (*13*). Together, these two reactions provide one mechanism of protein labeling with phenoxyl radicals, as depicted in Figure 2-3. Previous work has shown that phenoxyl radicals can covalently react with electron-rich amino acids such as tyrosine, phenylalanine, tryptophan, and cysteine. In the case of tyrosine residue, dityramine formation was suggested as the labeling product (*14-18*).

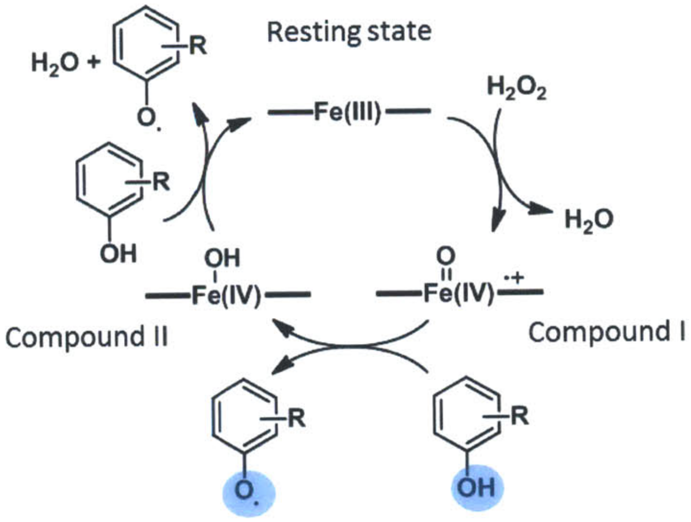


Figure 2-2. Catalytic cycle of heme peroxidase.
Heme Fe (III) at its resting state gets oxidized by hydrogen peroxide into high-valence Fe(V) state, forming Compound I. Compound I then undergoes one-electron transfer to abstract a hydrogen atom from the aromatic substrate, with itself being reduced to Compound II. Compound II is capable of abstracting another hydrogen atom from the substrate in a one-electron transfer reaction, returning to the resting state. In each catalytic cycle, two free radicals are generated.

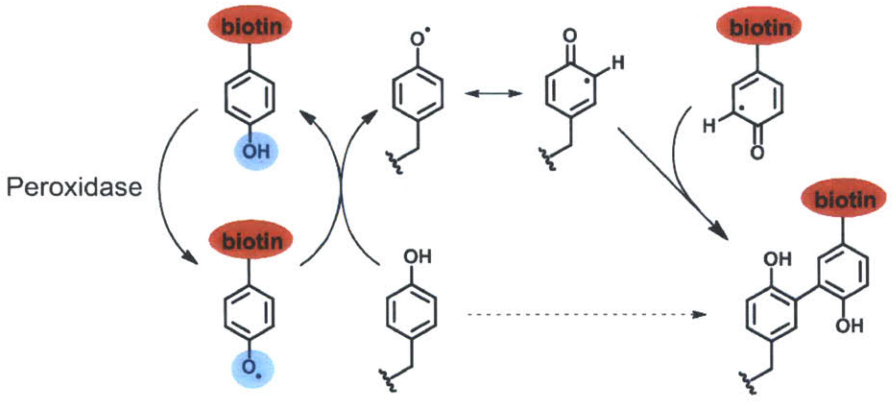


Figure 2-3. Scheme of phenoxyl radical crosslinking.
In the first step, the radical is transferred from biotin-phenoxyl group to an electron-rich amino acid on the endogenous protein target, such as the tyrosine residue. Hydrogen atom transfer results in the formation of a tyrosyl radical on the protein surface. Because of electron delocalization, the radical may stay on carbon center at the ortho- or para-position. Then, another nearby biotin-phenoxyl radical reacts with the tyrosyl radical on the protein surface to form a covalent carbon-carbon bond. Subsequent tautomerization yields the dityramine product.

2.2 Characterization of ascorbate peroxidase as a promiscuous labeling enzyme

Screening of biotin-phenol substrates

We measured the *in vitro* enzyme kinetics of ascorbate peroxidase toward aromatic substrates. Guaiacol was used as a model aromatic substrate of peroxidase for determining Michaelis-Menten kinetics. In this colorimetric assay, APX catalyzes the one-electron oxidation of guaiacol into its phenoxyl radical, which subsequently polymerizes into tetraguaiacol, leading to an increase in absorbance at 470 nm ($\varepsilon_{470} = 22 \times 10^3$ M^{-1} cm^{-1}) (*19*). For comparison, we also measured the kinetic parameters for HRP. Dr. Hyun-Woo Rhee performed these experiments and summarized the results in Table 2-1.

Table 2-1. Michaelis-Menten kinetics of APX and HRP toward guaiacol

Enzyme	k_{cat} (s^{-1})	K_m (mM)	k_{cat} / K_m (M^{-1} s^{-1})
wild-type APX	76.3 ± 8.0	6.9 ± 0.4	1.1×10^4
wild-type HRP	416 ± 14	1.7 ± 0.2	2.4×10^5

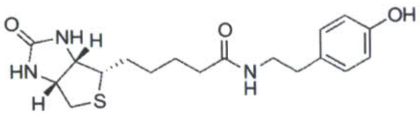
We therefore decided to test whether APX, in combination with a biotinylated phenol substrate, could covalently tag cellular proteins in a spatially-resolved manner. The strong interaction between biotin and streptavidin would allow for affinity capture of biotinylated proteins under extremely harsh conditions. For example, stringent wash conditions (2M urea and 200 mM NaCl) can be applied to remove contaminants. Different substitution groups could fine-tune the redox potential of the phenoxyl radical by changing the charge density of the aromatic system. While electron-donating groups, such as hydroxy group and methoxy group, stabilize the phenoxyl radical, electron-withdrawing groups, such as nitro group, destabilize it. Linker structures between the biotin moiety and phenol affects both cell membrane permeability and enzyme recognition.

We expressed APX in the cytosol of human embryonic kidney (HEK293T) cells, and tested labeling with a panel of phenol derivatives conjugated to biotin (Figure 2-4). For comparison, we also tested this panel of substrates with HRP targeted to the cell surface. The simplest biotin conjugate (**1**) gave the strongest labeling of intracellular proteins following 1 min reaction in the presence of 1 mM H_2O_2, as seen by streptavidin staining (Figure 2-4). Interestingly, ortho-substituted phenols, whether it is nitro, hydroxy, or methoxy groups (**2**, **3**, and **4**) all gave lower labeling yield. Putting a long linker between biotin and phenol (**5**) did not affect reaction by HRP on the cell surface, but greatly diminished labeling with APX inside cells. This is likely due to poor cellular membrane permeability of this substrate. Substitution groups outside but adjacent to the aromatic ring also reduces labeling yield (**6** and **7**), presumably because of steric clashes with residues in the active site. Aryl azide **8** was used in a previous study with HRP and was believed to form a nitrene radical (*20*), but we were unable to reproduce this labeling reaction. Based on its superior performance in labeling yield, we utilized biotin-phenol probe **1** for all subsequent labeling experiments.

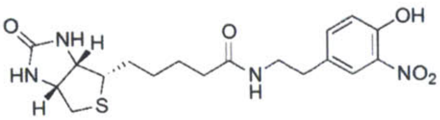
Figure 2-4. Reactivity of eight different biotin substrates towards APX and HRP.
This figure was generated by Dr. Hyun-Woo Rhee. Left: structures of biotin conjugates tested. These differ in electronic properties, linker length, sterics, and charge. Right: immunostain of HEK293T cells labeled with corresponding biotin-phenol conjugate. Cells expressed either cytosolic Flag-^{W41F}APX-NES (left column) or cell surface HRP-myc-TM (right column). Confocal fluorescence images were overlaid on top of DIC images. NES = nuclear export signal. TM = transmembrane domain of the PDGF receptor. For APX, labeling was performed by incubating cells with 500 μM of the indicated biotin substrate for 30 min, then adding 1 mM H_2O_2 for 1 min to initiate biotinylation. For cell surface labeling with HRP, 100 μM of the indicated substrate was added for 10 min, then 1 mM H_2O_2 was added for 1 min to initiate labeling. Cells were then fixed, permeabilized, and stained. Scale bars, 10 μm.

Substrate:

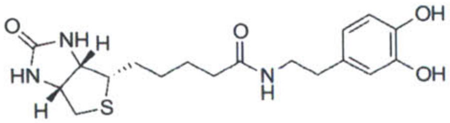
1



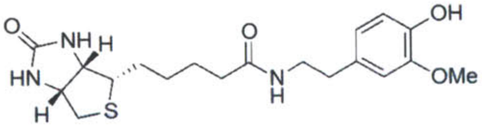
2



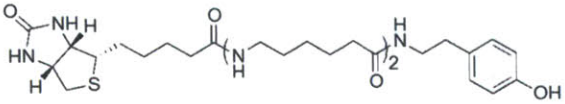
3



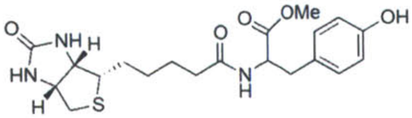
4



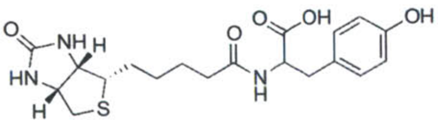
5



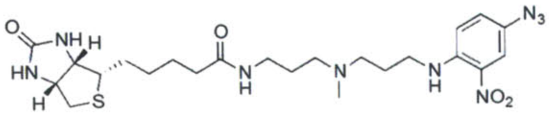
6

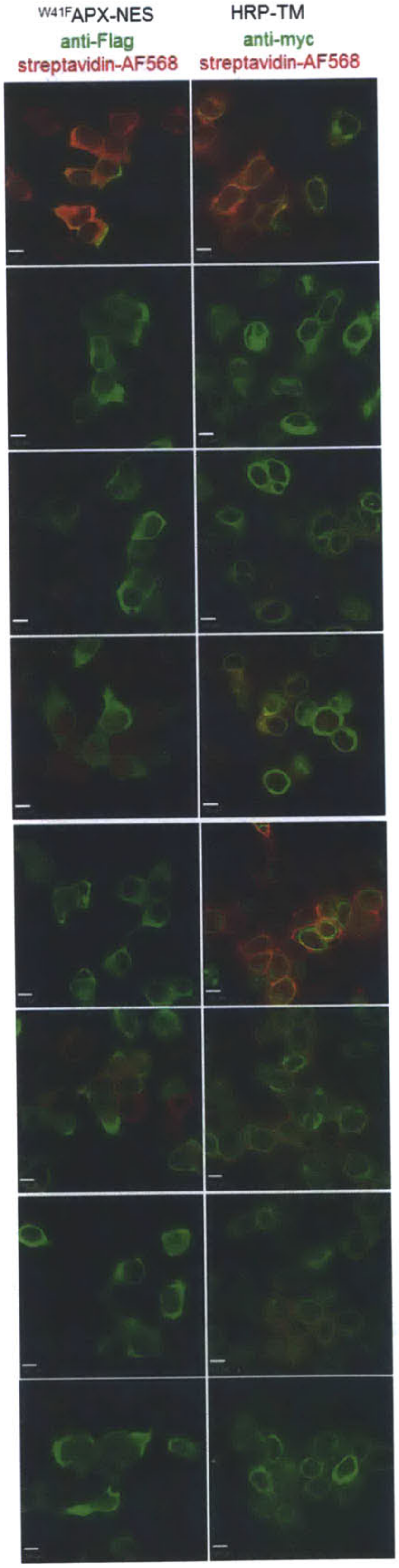


7



8




W41FAPX-NES
anti-Flag
streptavidin-AF568
HRP-TM
anti-myc
streptavidin-AF568

In vitro characterization of labeling product

We confirmed with mass spectrometry (MS) analysis that peroxidase-generated biotin-phenoxyl radical forms a covalent adduct with tyrosine, but we were not able to detect reaction products with any of the other nineteen amino acids (Figure 2-5). Statistics from protein structure analysis shows that 80% of tyrosine residues are solvent-exposed *(21)*. In the human proteome, merely 2.1% proteins have no tyrosine residue. 3.3% proteins have 1 tyrosine residue, which has 80% chance of being surface-exposed. The remaining 94.6% proteins have 2 or more tyrosine residues, which almost surely have at least one being surface-exposed. Based on this information, we estimate that >90% of proteins have one or more surface-exposed tyrosine residues, so the amino acid preference shown in phenoxyl radical labeling should not greatly limit our potential depth of coverage.

In addition to testing with amino acids, we also labeled bovine serum albumin with biotin-phenoxyl radical *in vitro*, and analyzed the reaction product by MS after tryptic digestion. Four tryptic peptides with biotin-phenol labeling on tyrosine have been identified by MS.

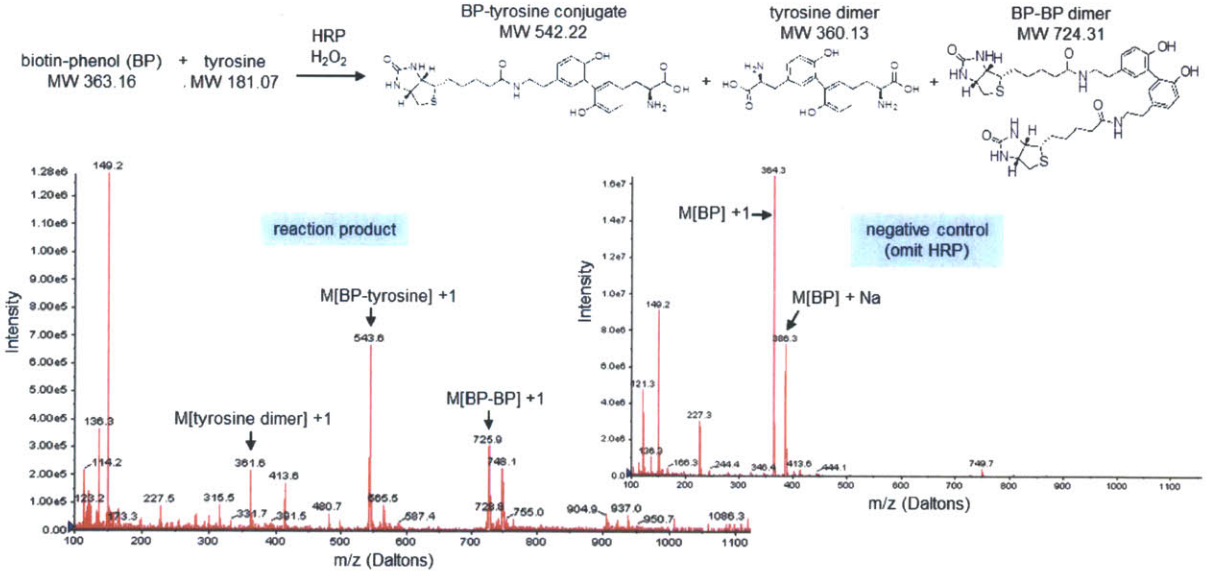


Figure 2-5. Amino acid specificity of labeling.
This figure was generated by Dr. Hyun-Woo Rhee. All 20 amino acids (each at 500 μM) were reacted *in vitro* with 500 μM biotin-phenol, 10 nM HRP and 1 mM H_2O_2 for 10 min at room temperature. HRP was used because it has the fastest *in vitro* kinetics. Apart from biotin-phenol dimer which was detected in all reactions, only the tyrosine reaction gave product, with the masses and predicted structures shown above. Product was not detected when HRP or H_2O_2 were omitted from the tyrosine reaction.

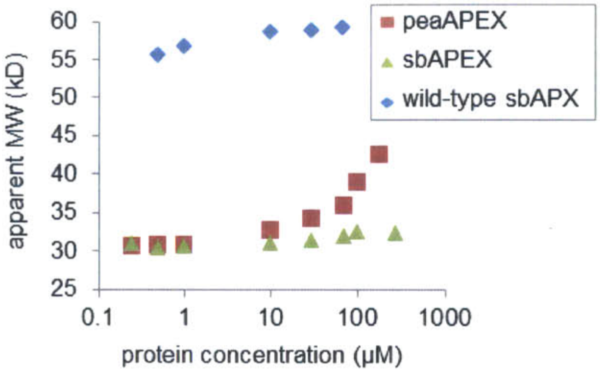


Figure 2-6. Gel filtration chromatography was used to analyze the dimerization tendency of mutant and wild-type soybean APX.
This figure was generated by Dr. Hyun-Woo Rhee and Jeffrey Martell. sbAPEX is triple mutant K14D, W41F, E112K of soybean APX, and peaAPEX is from a previous study (*9*) The expected molecular weight (MW) for monomer is 29 kDa. As expected (*7*), wild-type APX runs as a dimer at all protein concentrations tested (250 nM–68 µM; apparent MW ~58 kDa). sbAPEX runs as a monomer (apparent MW ~30 kD), even at a concentration of 270 µM, while peaAPEX runs as a monomer at lower concentrations but shows signs of dimerization at higher concentrations (as previously observed (*9*)).

Improving ascorbate peroxidase through mutagenesis

One problem with APX is that it is a homo-dimer (*7*). This limits its application as a fusion tag to cellular proteins. In Jeffrey Martell's earlier work on using APX as an EM reporter (*9*), ascorbate peroxidase from pea was engineered to be monomeric by introducing two mutations (K14D and E112K) at the dimer interface. Examination of crystal structures of recombinant ascorbate peroxidase from soybean and pea revealed that the dimer interfaces are conserved (*7*). We therefore tested homologous mutations in both soybean APX and pea APX. Gel filtration chromatography showed that the soybean APX mutant had even lower dimeric association affinity (Figure 2-6).

However, the monomeric form suffers from low activity towards the biotin-phenol substrate (data not shown). In the crystal structure, the active site contains a bulky tryptophan residue (Trp41, in Figure 2-1). We reasoned that by mutating this tryptophan into another residue with smaller side chain could help accommodate binding of biotin-phenol substrate. For this, we mutated W41 into all other 19 amino acid residues and tested labeling by expressing the enzyme in the cytosol. As shown in Figure 2-7, mutant W41F has the highest activity overall. An aromatic residue at this position is favored probably because it facilitates aromatic substrate binding through π-π stacking interaction. For the same reason, W41H also shows improved activity, although not as good as W41F. Interestingly, mutation to another aromatic residue, W41Y, fails to increase activity. Mutation to residues with small, hydrophobic side chains, such as alanine, valine and cysteine also improves activity, which supports the hypothesis of opening up the binding pocket. Among these, mutant W41V has especially high activity in some cells, but we also observed a number of cells displaying decreased activity. This is probably due to protein misfolding which compromises its activity.

Based on its reduced dimerization affinity and enhanced activity, we chose soybean APX triple mutant (K14D, W41F, E112K) for subsequent characterization and proteomic studies. For simplicity, we name this triple mutant "APEX", for enhanced ascorbate peroxidase. We used both biochemical and imaging methods to characterize the spatial specificity of APEX labeling.

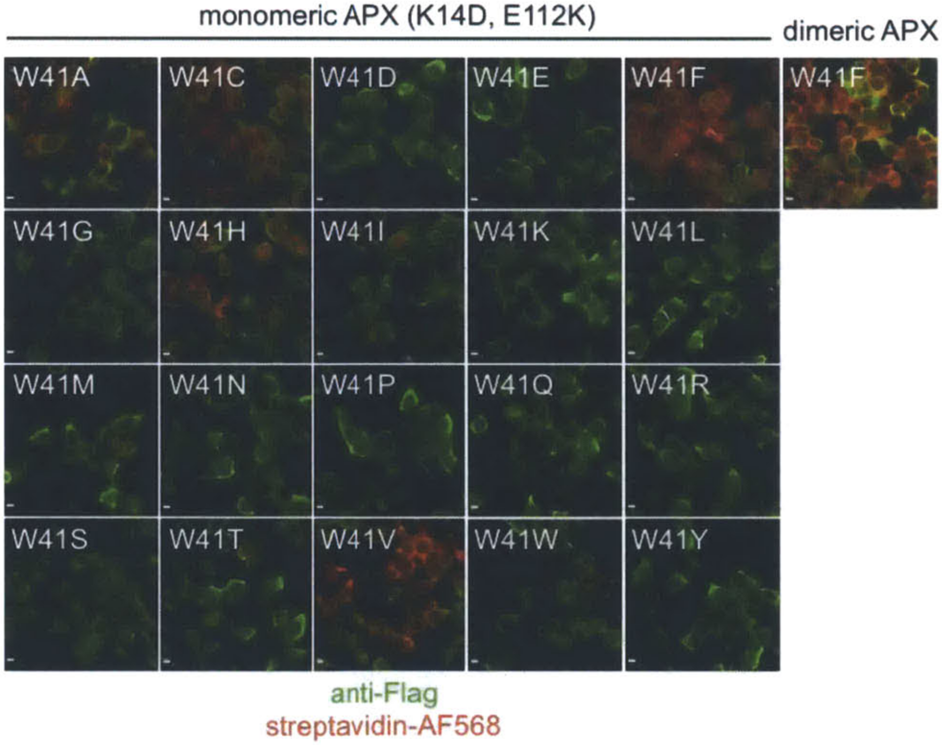


Figure 2-7 Engineering a monomeric peroxidase with high activity towards biotin-phenol. This figure was generated by Dr. Hyun-Woo Rhee. Because previous studies have shown that mutations at the W41 position boosted activity towards aromatic substrates (*22*), we tested all possible mutations at W41 on a monomeric APX background. NES (nuclear export sequence)-targeted constructs were expressed in HEK293T and labeling was performed live as in Figure 2-4. Peroxidase expression was detected with anti-Flag staining, and biotinylated proteins were visualized with streptavidin-AlexaFluor568 (AF568). Dimeric ^{W41F}APX-NES (used in Figure 2-4) is included for comparison. Confocal images of anti-Flag, streptavidin, and DIC images are overlaid. Scale bars, 10 μm.

Comparison between APX and promiscuous BirA labeling

As described in Chapter 1, promiscuous biotin ligase (pBirA) has also been used for proteomic labeling (*23*), but it suffers from poor temporal and spatial resolutions. We verified that the kinetics of promiscuous biotinylation with pBirA is much slower than APEX labeling, requiring as much as 24 hours. Whereas 1 minute reaction with APEX gives strongly labeled proteome, there is almost no product formation with pBirA in the same time window, as measured by both fluorescence imaging (Figure 2-8) and Western blot analysis (Figure 2-9) of biotinylated proteins. Even after 1-hour labeling with pBirA, the biotinylation level barely stands out from endogenous biotinylated proteins (Figure 2-8). Labeling with pBirA became detectable after 24 hours, and even so, we observed extremely low activity of pBirA targeted to the endoplasmic reticulum lumen, suggesting that it is not a universally applicable enzyme (Figure 2-9).

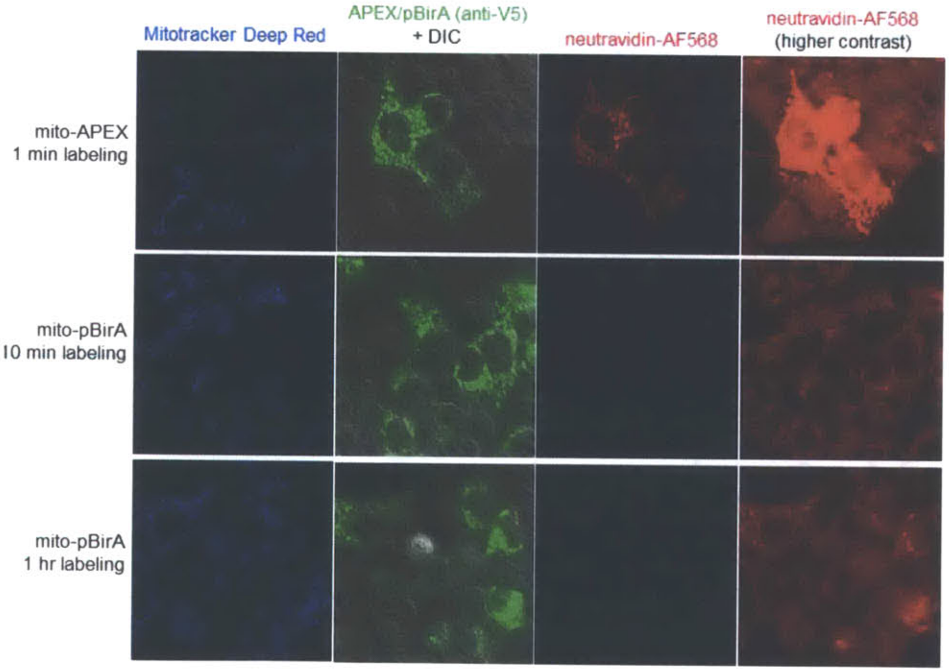

Mitotracker Deep Red
APEX/pBirA (anti-V5) + DIC
neutravidin-AF568
neutravidin-AF568 (higher contrast)
mito-APEX 1 min labeling
mito-pBirA 10 min labeling
mito-pBirA 1 hr labeling

Figure 2-8. Biotinylation of mitochondrial proteins by mito-pBirA is undetectable after 10 min or 1 hour of labeling.
Confocal fluorescence images of immunostained HEK293T cells. For comparison, the top row shows 1 min biotinylation by mito-APEX. For the bottom two rows, cells were transfected with mitochondria-localized promiscuous BirA (mito-pBirA). 24 hours after transfection, 50 μM biotin was added to the culture medium to permit promiscuous biotinylation. Mitotracker deep red staining was also used to visualize mitochondria. The neutravidin channel is also shown at higher contrast. The staining seen in rows 2 and 3 reflects endogenous biotinylated proteins (located in mitochondria) because the staining is equivalent between transfected (V5-positive) and untransfected cells.

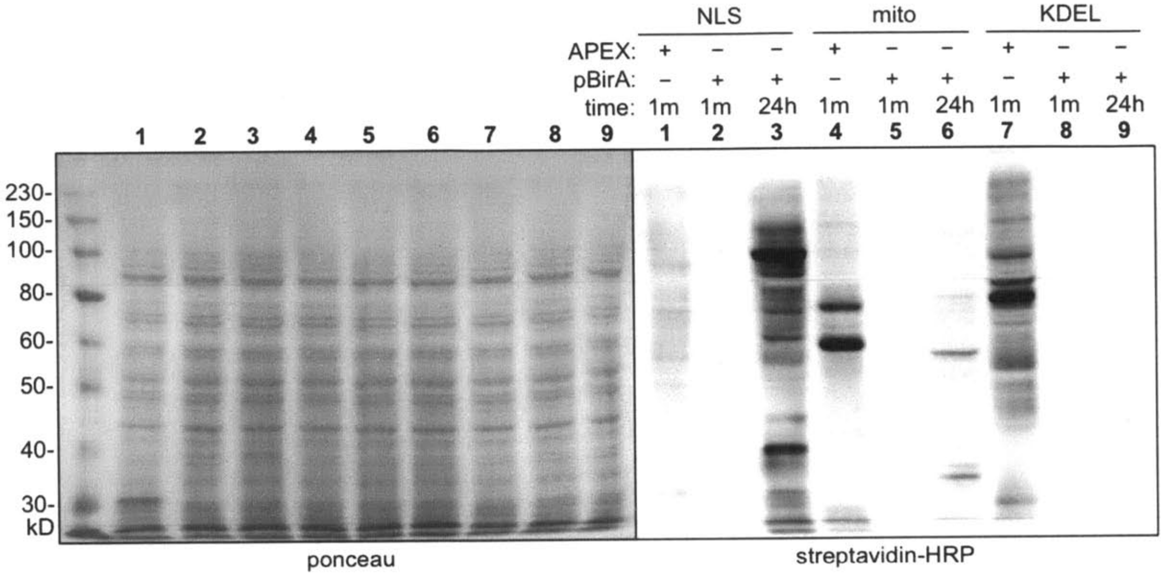


Figure 2-9. Comparison of promiscuous biotinylation by APEX and pBirA with Western blot. This figure was generated by Dr. Hyun-Woo Rhee. Labeling with APEX and promiscuous biotin ligase (pBirA = R118G BirA) (*23*) was performed in live HEK293T cells in three different compartments (NLS = nucleus-targeted with nuclear localization sequence; mito = mitochondrial matrix; KDEL = endoplasmic reticulum-targeted via KDEL endoplasmic reticulum-retention sequence). APEX labeling is performed for 1 min (1m) with 500 μM biotin-phenol and 1 mM H_2O_2. pBirA labeling is performed for 1 min (1m) or 24 hours (24h) with 50 μM biotin, as previously described (*23*). Streptavidin blot analysis of whole cell lysate is shown at right. The same samples are visualized by Ponceau stain at left.

2.3 Assessing the spatial specificity of APEX labeling

Evaluating APEX labeling by microscopy and Western blot analysis

To assess the spatial specificity of APEX-mediated biotinylation, we targeted APEX to various organelles and compartments (nucleus, cytosol, mitochondrial matrix, endoplasmic reticulum lumen, and cell surface) in human embryonic kidney (HEK293T) cells via genetic fusion to targeting peptides, and imaged the resulting pattern of biotinylated targets with streptavidin stain (Figure 2-10, panel A). In all of these compartments, the streptavidin stain overlaps tightly with peroxidase localization. In the case of mitochondrial matrix and endoplasmic reticulum lumen, we examined the localizations of enzyme and biotinylated targets in COS-7 cells, because these organelles are more clearly discernible in these cells (Figure 2-10, panel B). Immunostain shows complete overlap between the enzyme, biotinylated proteins and the organelle markers. The resolution of this analysis is limited by light diffraction, which is about 230 nm in confocal microscopy.

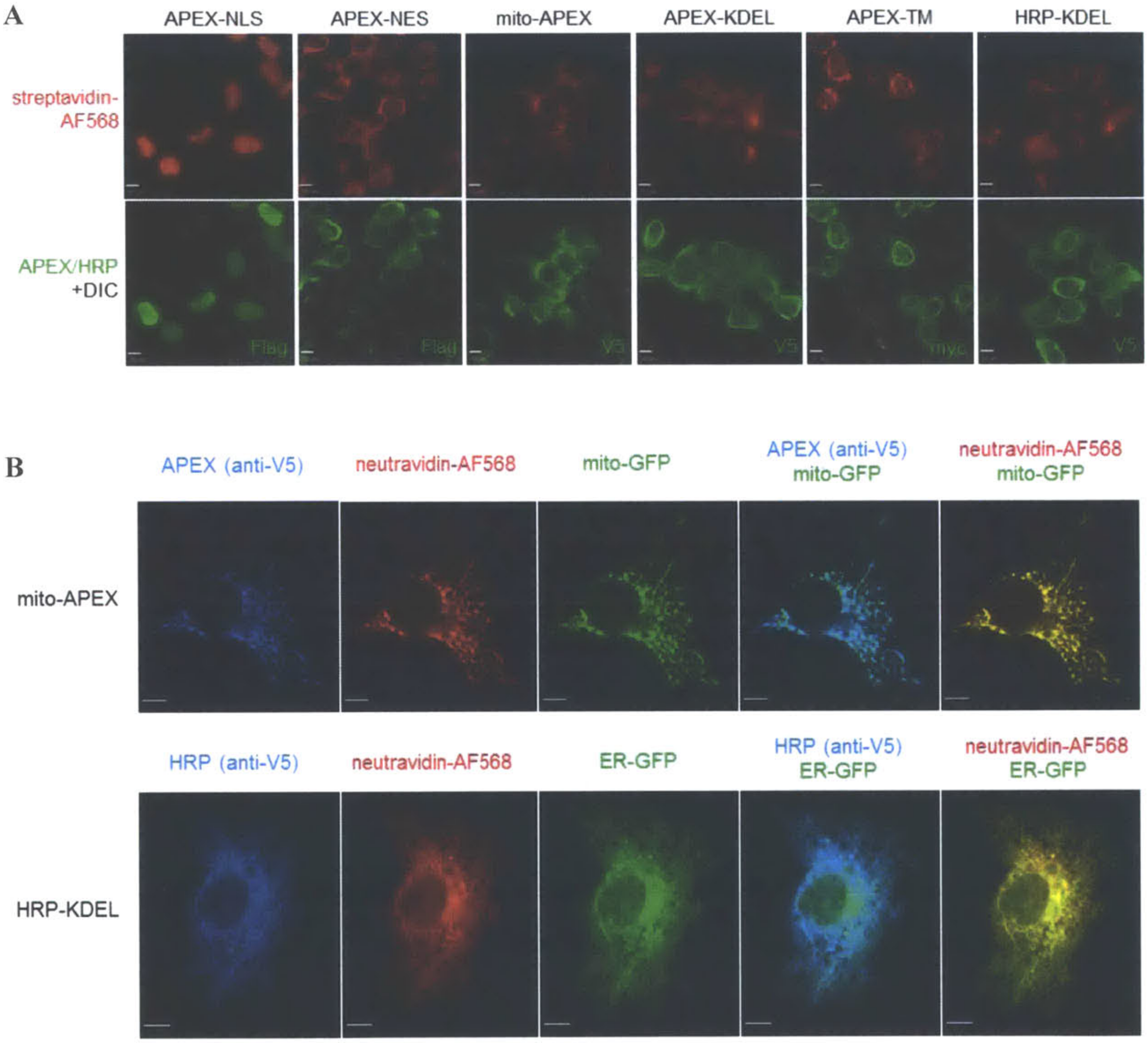

A
APEX-NLS
APEX-NES
mito-APEX
APEX-KDEL
APEX-TM
HRP-KDEL
streptavidin-AF568
APEX/HRP +DIC
Flag
Flag
V5
V5
myc
V5
B
APEX (anti-V5)
neutravidin-AF568
mito-GFP
APEX (anti-V5) mito-GFP
neutravidin-AF568 mito-GFP
mito-APEX
HRP (anti-V5)
neutravidin-AF568
ER-GFP
HRP (anti-V5) ER-GFP
neutravidin-AF568 ER-GFP
HRP-KDEL

Figure 2-10. Confocal microscopy analysis of APEX and HRP-catalyzed promiscuous biotinylation in many cellular compartments.

A) HEK293T cells were transfected with APEX or HRP constructs targeted to various compartments. B) COS-7 cells were transfected with APEX or HRP constructs and organelle markers (ER-GFP utilized the N-terminal 29 amino acid residues of cytochrome P450 (*24*)). Labeling was performed with biotin-phenol and H_2O_2 for 1 min. Cells were then fixed and stained with streptavidin-AlexaFluor568 (AF568) to visualize biotinylated proteins, and antibody to visualize Flag, V5, or myc tags on the APEX/HRP enzymes. NLS = nuclear localization signal; NES = nuclear export signal; mito = mitochondrial matrix signal; KDEL = endoplasmic reticulum retention motif; TM = transmembrane helix of PDGF receptor. Scale bars, 10 μm.

We assessed the spatial specificity of biotin-phenol labeling at even higher resolution. Electron microscopy is potentially useful and has been used in previous studies of similar problems (*11,12*). However, sample preparation requires using gold nanoparticles, which tend to generate non-specific background. We decided to use two-color super-resolution light microscopy for comparing the localizations of enzyme and biotinylated targets at high resolution. In recent years, many fluorescence-based imaging techniques have been developed to break the diffraction limit barrier of ~200 nm. These super-resolution microscopy techniques have been widely used for visualization of detailed cellular structures, such as the cytoskeleton. Among these, stochastic optical reconstruction microscopy, or STORM (*25*), has produced the best spatial resolution. STORM relies on the sequential imaging and localization of individual small organic fluorophores that can be photo-switched between bright and dark states.

In this analysis, we target APEX into the mitochondrial matrix of COS-7 cells and performed promiscuous labeling with biotin-phenol as previously described. To compare the localizations of biotinylated targets and the enzyme under super-resolution, we immunostained the cells with streptavidin-Cy3/Cy5 and antibody conjugated with AlexaFluor405/AlexaFluor647 against APEX, respectively. Two-color STORM image in Figure 2-11 showed that the patterns of biotinylated proteins and overlapped tightly with that of APEX in the mitochondrial matrix. The lateral resolution of STORM image on our setup was measured to be 22 nm.

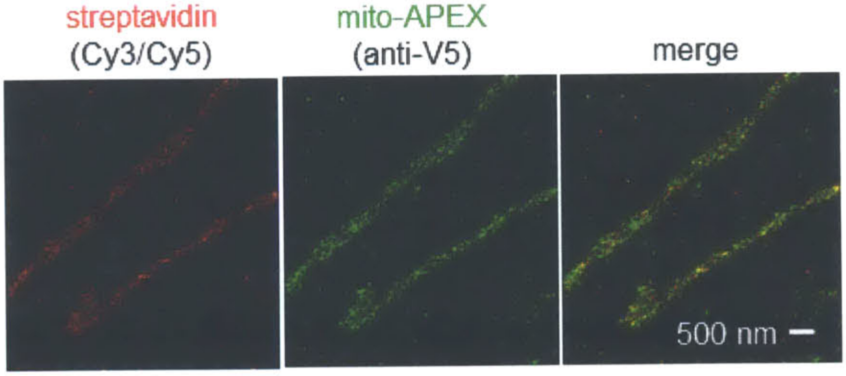


Figure 2-11. STORM imaging analysis of APEX promiscuous labeling in mitochondria. Biotinylated targets were labeled with streptavidin-dye conjugate, and mito-APEX, with V5 epitope tag, was labeled with anti-V5 antibody. We estimated the degree of cross-talk between channels by two-color imaging of fixed cell samples prepared as described above. We estimate that the maximum crosstalk of streptavidin into the anti-V5 channel is 5% (AlexaFluor405/AlexaFluor647 is activated by the 561 nm laser by 5% relative to activation by the 405 nm laser). The maximum crosstalk for anti-V5 into the streptavidin channel is 14% (Cy3/Cy5 is activated by the 405 nm laser by 14% relative to activation by the 561 nm laser).

In addition to imaging analysis, we collected these labeled cells and analyzed the whole cell lysate using Western blot, in which streptavidin-HRP specifically detects biotinylated proteins. Figure 2-12 shows that APEX yields different patterns, or "fingerprints", of biotinylated endogenous proteins in these compartments.

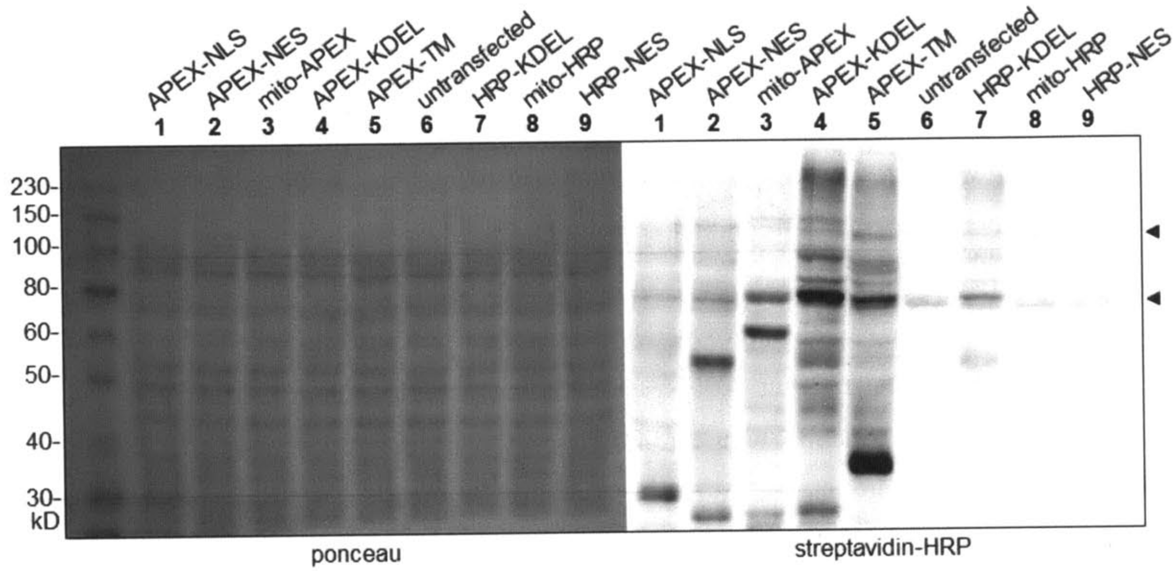


Figure 2-12. Western blot analysis of APEX and HRP-catalyzed promiscuous biotinylation in various cellular compartments.
This figure was generated by Dr. Hyun-Woo Rhee. HEK293T cells were transfected and labeled with biotin-phenol and H_2O_2, and then lysed. Whole cell lysates were analyzed by 10% SDS-PAGE, with Ponceau staining (left) or streptavidin-HRP visualization of biotinylated proteins (right). Arrowheads point to endogenously biotinylated proteins (*26*).

APEX labeling does not cross cellular membranes

So far, our imaging results suggest that the reactive species generated by APEX does not cross a variety of cellular membranes, but the resolution at best is still limited to ~22 nm. We therefore developed a cell surface labeling assay to more rigorously demonstrate that APEX-catalyzed tagging does not cross the plasma membrane (Figure 2-13). In this assay, HEK293T cells were transfected with ^{W41A}APX targeted to the cytosol (via the NES targeting sequence) or the inner leaflet of the plasma membrane (via the CAAX targeting sequence). Labeling was performed with alkyne-phenol and H_2O_2 for 1 min. Thereafter click chemistry was performed at the cell surface with a membrane-impermeant AlexaFluor647-picolyl azide (*27*) conjugate. Only if the phenoxyl radical can cross the plasma membrane will alkyne be present at the cell surface and detectable by the AlexaFluor647 reagent. Dimeric ^{W41A}APX was used in this experiment because it has higher activity in the mammalian cytosol than monomeric APEX.

As a control, cells were fixed and permeabilized prior to click chemistry to detect the total pool of alkyne-phenol labeling. Quantitation shows that the signal at the surface represents less than 2% of the total signal. The same results were obtained when we used biotin-phenol instead of alkyne-phenol and detected labeling at the cell surface with a membrane-impermeant streptavidin-AlexaFluor568 conjugate.

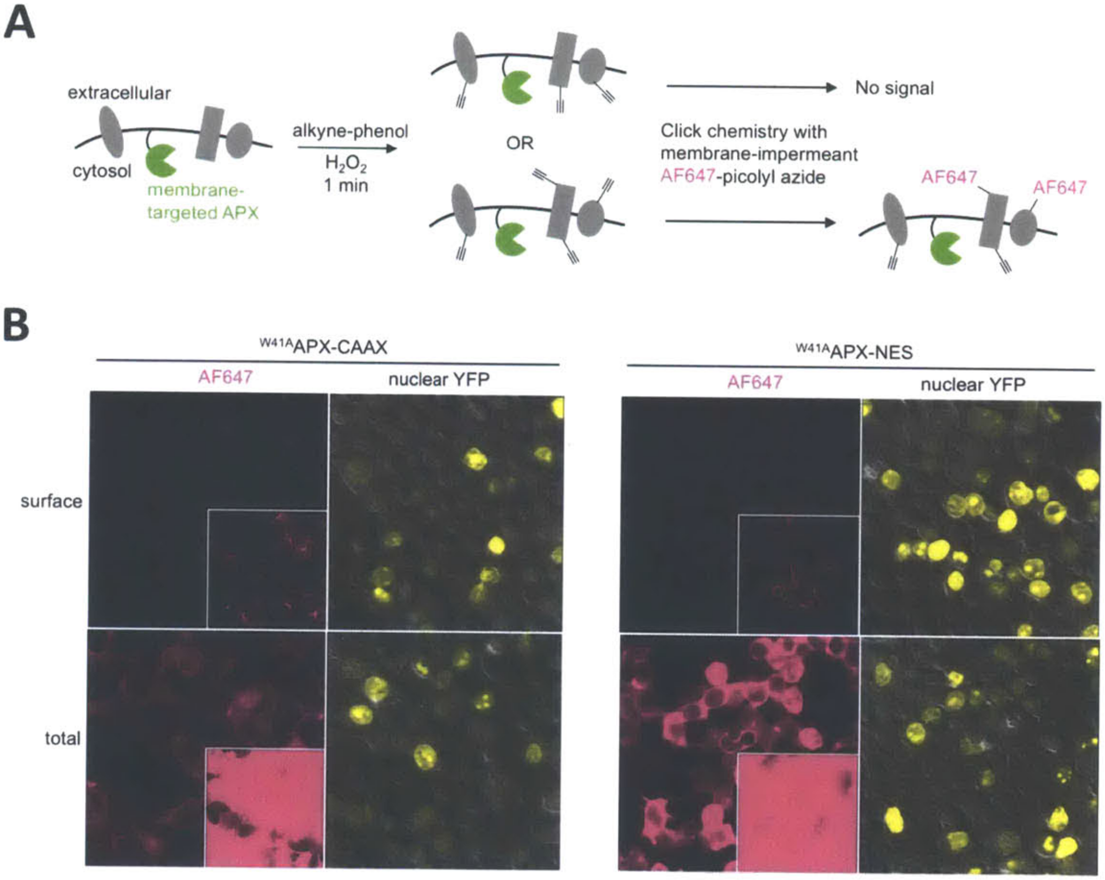

A
extracellular
cytosol
membrane-targeted APX
alkyne-phenol
H_2O_2
1 min
OR
No signal
Click chemistry with membrane-impermeant AF647-picolyl azide
AF647
AF647
B
W41AAPX-CAAX
W41AAPX-NES
AF647
nuclear YFP
AF647
nuclear YFP
surface
total

Figure 2-13. Peroxidase-generated phenoxyl radicals do not cross the plasma membrane.
(A) Assay scheme. APX is targeted to the inner leaflet of the plasma membrane. Alkyne-phenol is used here and the labeling is subsequently detected with AlexaFluor647 (AF647)-picolyl azide using click chemistry. (B) Images with ^{W41A}APX-CAAX (left) and ^{W41A}APX-NES (right). Nuclear-localized YFP was a transfection marker. Bottom rows (labeled "total") show detection of the entire alkyne-phenol labeling in fixed and permeabilized cells. Insets show the same fields of view with 50-fold greater contrast.

Preserving spatial specificity by immediate quenching of labeling reaction

When developing the labeling protocol, we found that APEX is a particularly stable enzyme, and is not easily inactivated in the presence of 4% formaldehyde during cell fixation. This poses the risk of decreasing the spatial resolution from residual APEX activity after cell fixation or cell lysis. To solve this problem, we added a quencher cocktail of Trolox (5 mM, a free radical quencher), ascorbate (10 mM, a free radical quencher) and sodium azide (10 mM, a peroxidase inhibitor) to eliminate residual activity. We showed by imaging that brief treatment of HEK293T cells with this quencher cocktail could reduce subsequent biotin-phenol labeling by over 90% (Figure 2-14). By Western blot analysis, we also showed that this treatment indeed eliminated labeling of OMM proteins with a matrix-targeted enzyme.

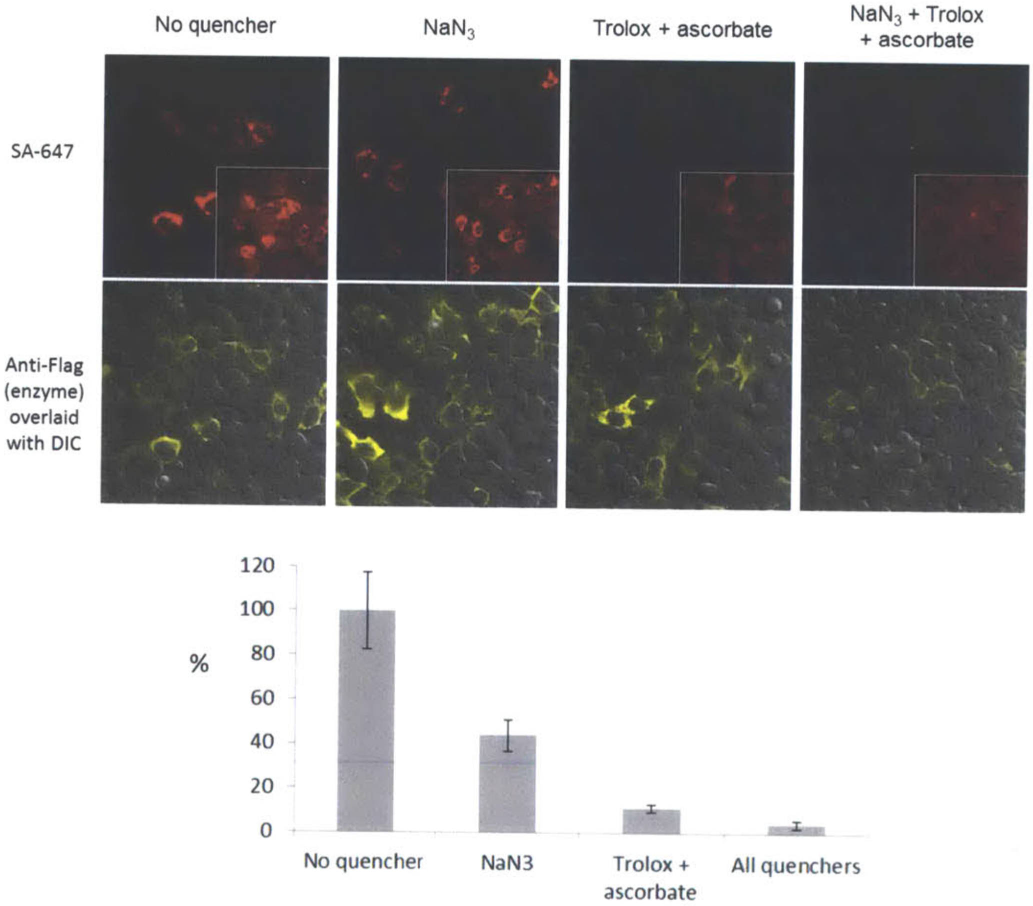


Figure 2-14. Quencher cocktail effectively reduces biotin-phenol labeling intensity.
Top: HEK293T cells express mitochondrial matrix-targeted ^{W41V}APX were treated with various quenchers for 1 min followed by biotin-phenol labeling. Cells were then fixed and stained with streptavidin-AlexaFluor647 (SA-647) to visualize biotinylated proteins, and antibody to visualize Flag tag on the enzyme. Bottom: quantitation of streptavidin-AlexaFluor647 fluorescence signal.

2.4 Controlling labeling radius

While the high spatial specificity of APEX labeling in membrane-bounded compartments are useful for studying organellar proteomics, a further challenge lies in carefully defining, and possibly controlling, the radius of peroxidase-catalyzed labeling when it is not membrane-enclosed. A powerful application of labeling radius control would be the proteomic mapping of macromolecular complexes. The labeling radius of APEX is determined by the diffusion of the product radical and its life time, which in turn is determined by the dynamic equilibrium between radical generation and radical quenching.

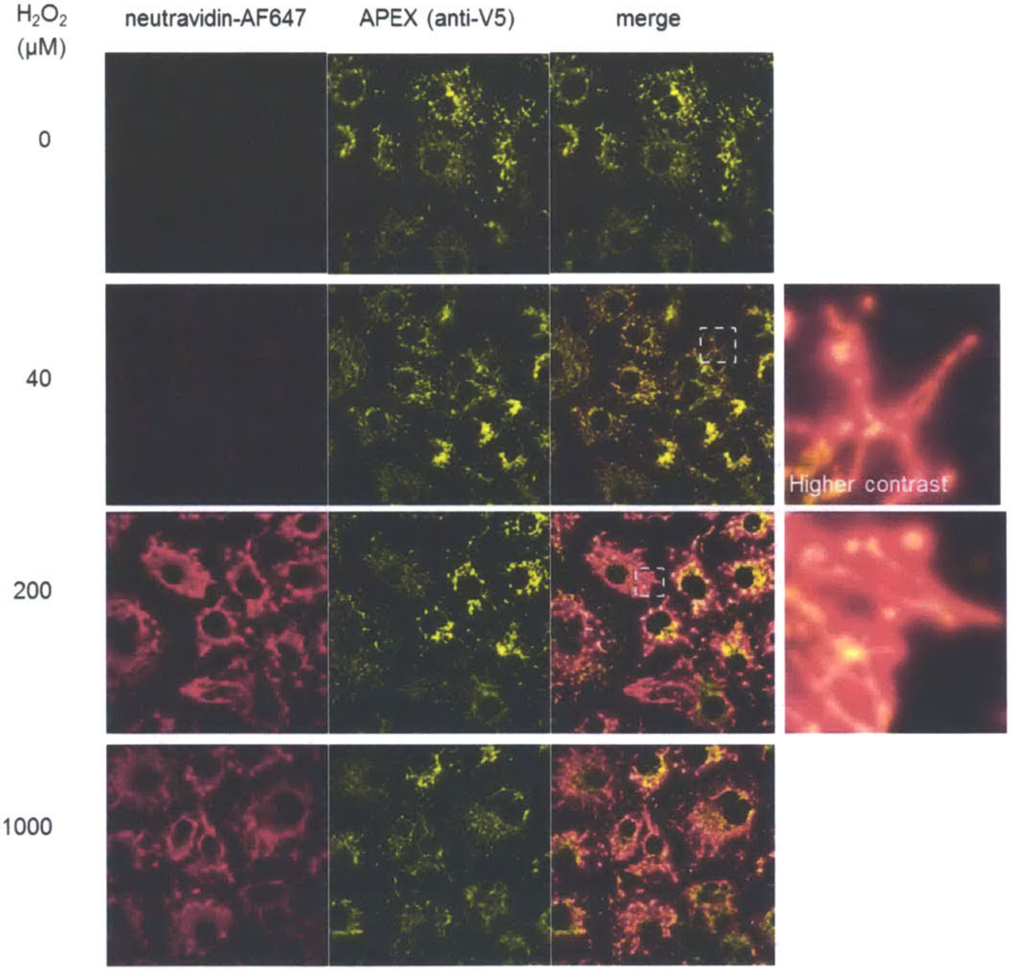


Figure 2-15. Titration of H_2O_2 does not affect labeling radius.
HRP was targeted to mitochondria in fixed HeLa cells via anti-Tom20 antibody. 100 μM biotin-phenol and various concentration of H_2O_2 were added to cells for 1 min. Thereafter, cells were washed, stained, and imaged. Fluorescence images are shown with enzyme marked in yellow and biotin signal in pink. Zoom-in views at concentrations 40 μM and 200 μM H_2O_2 show comparison of the extent of signal diffusion from the enzyme source at different H_2O_2 concentrations (biotin signal is not normalized).

An imaging readout following live cell labeling can be misleading because biotinylated proteins can themselves diffuse during the 1 min labeling window. Moreover, it is difficult to control various factors affecting the labeling radius in an *in vivo* experiment. Instead, we chose to perform promiscuous labeling on fixed cells and then use fluorescence imaging to measure the radius of labeling, where a larger radius corresponds to wider distribution of biotin signal relative to the enzyme localization. Note that, in this experiment, we chose HRP as the promiscuous labeling enzyme instead of APEX for the following reasons: 1) Both enzymes generate the same biotin-phenoxyl radical, so conclusions of radical diffusion and lifetime drawn from studying HRP can be extended to APEX labeling; 2) HRP is more active than APEX, thereby giving stronger labeling signal that is more readily detectable; and 3) various HRP-antibody conjugates are readily available, allowing highly specific targeting of HRP enzyme to subcellular structures on fixed cells.

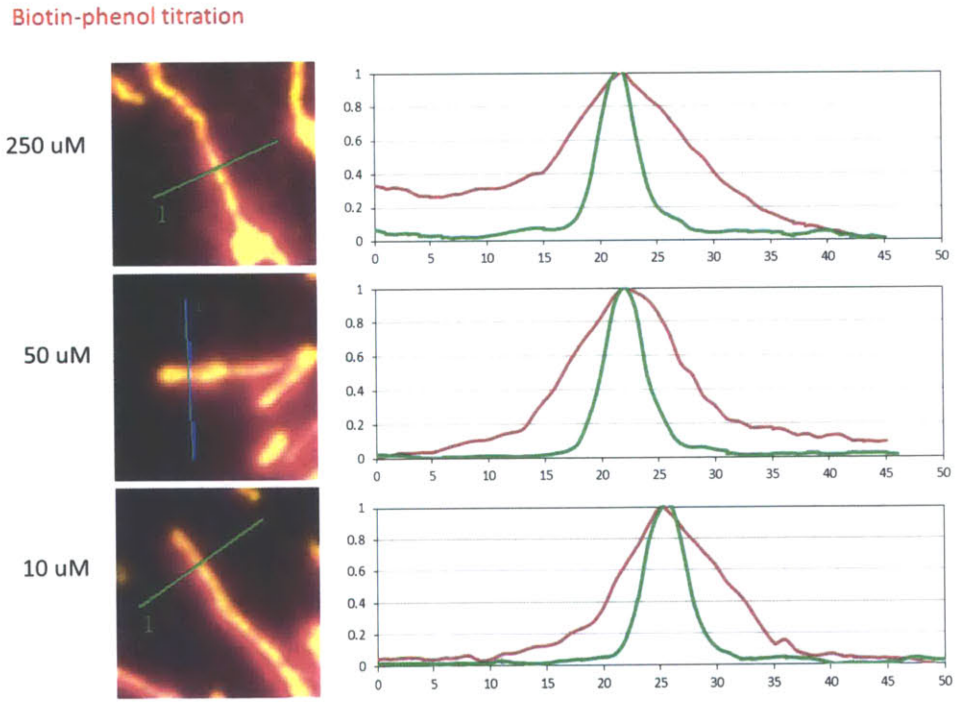


Figure 2-16. Biotin-phenol probe titration does not affect labeling radius.
HRP was targeted to mitochondria in fixed HeLa cells via anti-Tom20 antibody. 1 mM H_2O_2 and various concentrations of biotin-phenol were added to cells for 1 min. Thereafter, cells were washed, stained, and imaged. Left: fluorescence images, with enzyme shown in yellow and biotin signal shown in pink. Right: cross-section analysis of intensity distribution of two channels. Horizontal axis is the coordinate in unit of pixels along the line drawn in images on the left, and vertical axis is the normalized fluorescence intensity. Red line, showing wider spread, represents fluorescence cross-section distribution of the biotin signal, and the green line represents the enzyme signal. Each pixel is 160 nm in size.

Using fluorescence microscopy, we investigated the effect of the following factors on the radius of biotin-phenol labeling by HRP enzyme: H_2O_2 and biotin-phenol concentrations, and radical quenchers. We started by testing whether lower substrate concentration could result in smaller radius of labeling. HRP enzyme was targeted to fixed and permeabilized HeLa cells via using an anti-Tom20 antibody. Labeling was performed with varying concentrations of biotin-phenol and H_2O_2 for 1 min. At lower concentrations of H_2O_2, the labeling intensity significantly decreased. However, we found that the labeling radius did not change significantly by titrating H_2O_2 (Figure 2-15). Titration of biotin-phenol also showed decreased signal intensity at lower probe concentration, but the labeling radius remained the same (Figure 2-16).

Phenoxyl radicals are known to be quenched by a variety of small molecule reducing agents, such as Trolox (vitamin E analog), ascorbate (vitamin C), thiols and gallate. These quenchers readily donate a hydrogen atom to the phenoxyl radical, reducing it to phenol. The resulting quencher radical is lower in energy and thus incapable of regenerating phenoxyl radicals. Varying the amount of quencher would allow us to tune the radical lifetime, thereby controlling its labeling radius. We tested various substituted phenols and thiols for their ability to reduce the APEX labeling radius (Figure 2-17). Gallate emerged as a potent quencher of biotin-phenol labeling reaction and it significantly reduced the labeling radius at sub-stoichiometric concentration (5 μM gallate with 100 μM biotin-phenol). Mercaptoethanol and cysteamine (mercaptoethylamine) also reduced the labeling intensity and radius when they are present at high concentrations (100 μM). Trolox (10 μM) and ascorbate (20 μM) were potentially useful for reducing the labeling radius, but at this concentration they almost completely abolished the labeling signal. Titration is required in future studies to find the optimal concentration for these two quenchers. The other two phenol structures (4-iodophenol and vanillin) failed to reduce the labeling radius.

In the future, our next goal is to apply these quenchers to fine-tune the labeling radius of biotin-phenol in live cells. This is potentially useful for mapping the component and structure of macromolecular protein complexes. We also noted that the intracellular concentration of thiols typically far exceeds the thiol quencher concentration (100 μM) tested in the above analysis. For example, glutathione is present in millimolar concentrations. Therefore, even without additional quenchers, the biotin-phenol labeling with APEX in live cells may have already had limited radius due to the quenching effect by those endogenous thiol molecules.

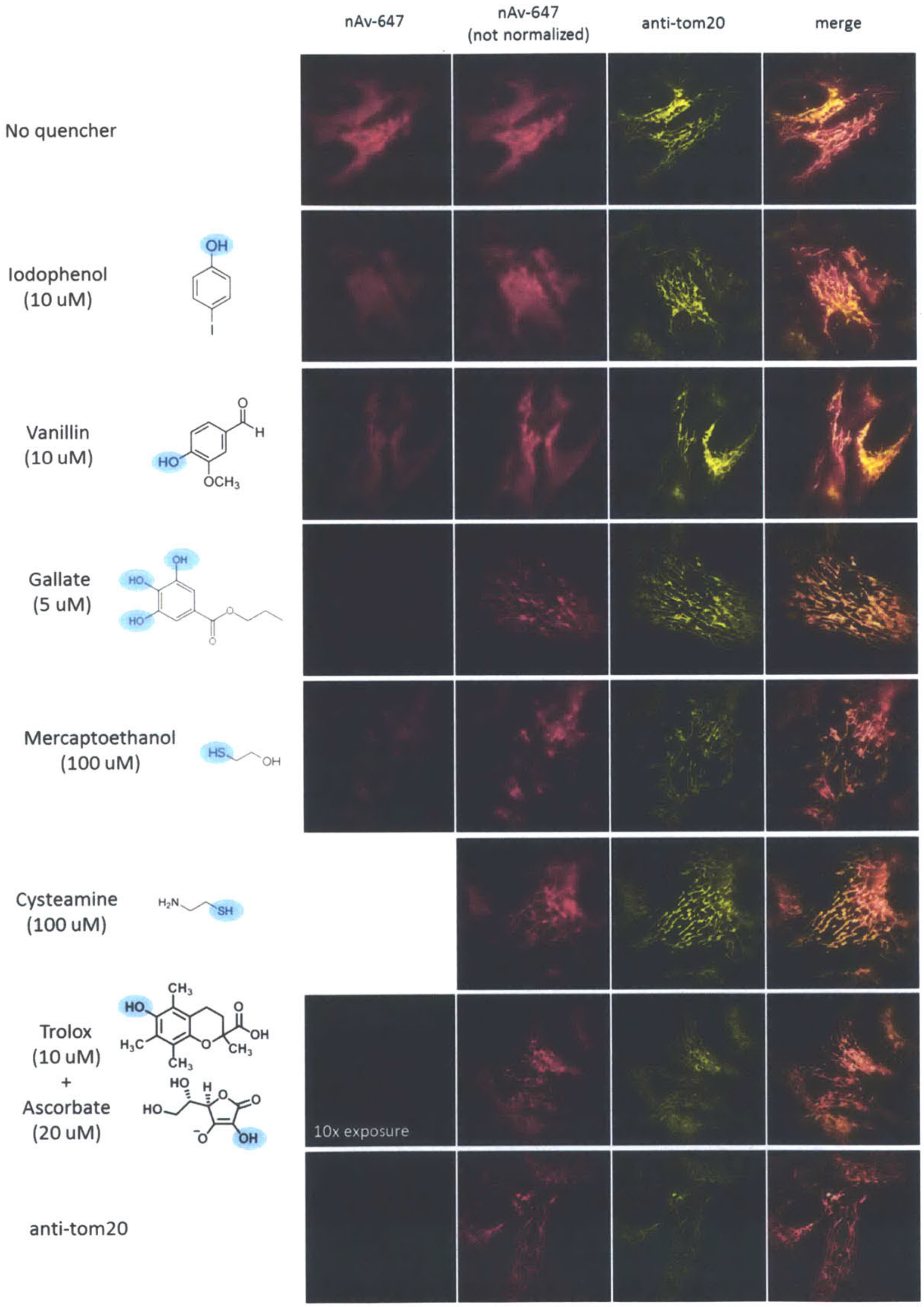

nAv-647
nAv-647 (not normalized)
anti-tom20
merge
No quencher
Iodophenol (10 uM)
OH
I
Vanillin (10 uM)
O
H
HO
OCH₃
Gallate (5 uM)
OH
HO
HO
O
Mercaptoethanol (100 uM)
HS
OH
Cysteamine (100 uM)
H₂N
SH
Trolox (10 uM)
+
Ascorbate (20 uM)
HO
CH₃
H₃C
O
OH
HO
H
⁻O
OH
10x exposure
anti-tom20

Figure 2-17. Addition of free radical quencher reduces the labeling radius.
HRP was targeted to mitochondria in fixed HeLa cells via anti-Tom20 antibody. 100 μM biotin-phenol, 1 mM H_2O_2 and various quenchers (structures and concentrations shown on the left, with the reactive hydrogen atom shaded in light blue) were added together to cells for 1 min. Thereafter, cells were washed, stained, and imaged. Biotin signal was detected with neutravidin-AlexaFluor647 (nAv-647), and enzyme signal was detected with anti-Tom20 immunofluorescence. Fluorescence images were shown next to the quencher structures. In the case of Trolox + ascorbate, the biotin signal was too low, so 10-times longer exposure was used in the nAv-647 channel. The bottom row, "anti-tom20", is the immunostain of enzyme with antibody-AlexaFluor647 conjugate, which serves as a reference for comparison with the same imaging wavelength.

2.5 Conclusion

In summary, we have implemented proximity-dependent promiscuous protein labeling using ascorbate peroxidase and a biotin-phenol substrate. In this approach, the enzyme acts as a point source that constantly generates biotin-phenoxyl radicals which diffuses away to attack proteins nearby. Through rational design and screening, we have identified an enzyme-substrate pair that gives the highest activity of labeling. Using confocal and super-resolution microscopy, as well as a membrane-permeability assay, we demonstrated that the promiscuous labeling is restricted to membrane-bounded compartments. We further tested the effect of radical quenchers on biotin-phenol labeling and indeed found that they helped reduce the labeling radius on fixed cells. In addition to co-applying radical scavengers such as gallate and thiols to cells, it should also be possible to modulate the labeling radius using electron withdrawing substituents on the aromatic ring. A future direction would be to screen biotin-phenoxyl radicals with small labeling radius and high reactivity, while at the same time identify enzymes with improved activity toward these substrates through protein engineering.

The next chapter describes the application of this promiscuous protein labeling technique to the study of mitochondrial matrix and endoplasmic lumen proteomes.

2.6 Experimental methods

Genetic constructs

Table 2-2 summarizes the plasmids used for peroxidase expression in mammalian cells.

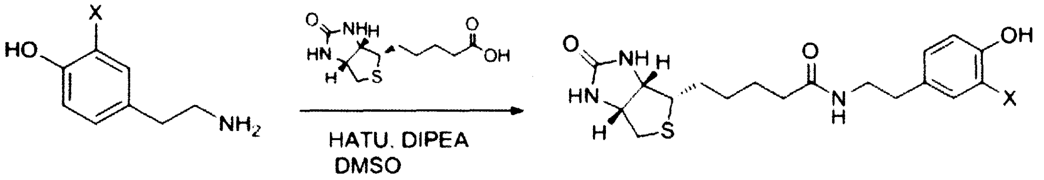
Table 2-2. Genetic constructs

Construct name	Features
Flag-APEX-NES	Flag: DYKDDDDK NES: LQLPPLERLTLD
Flag-^{W41F}APX-NES	
Flag-APEX-NLS	NLS: PKKKRKVDPKKKRKVDPKKKRKV
Flag-^{W41A}APX-CAAX	CAAX: RSKLNPPDESGPGCMSCKCVLS
ss-HA-APEX- myc-TM	ss: Ig-κ chain signal sequence in pDisplay HA: YPYDVPDYA myc: EQKLISEEDL TM: transmembrane domain of PDGF receptor
ss-HA- HRP- myc-TM	
ss-APEX-V5-KDEL	KDEL: endoplasmic reticulum retention motif V5: GKPIPNPLLGLDST
ss-HRP-V5-KDEL	
mito-V5-APEX	mito matrix targeting sequence: MLATRVFSLVGKRAISTSVCVRAH
LACTB$_{(1-68)}$-V5-APEX	LACTB$_{(1-68)}$ leader sequence: MYRLLSSVTARAAATAGPAWDGGRRGAHRRPGLPV LGLGWAGGLGLGLGLALGAKLVVGLRGAVPIQS (reference (*28*))
PPOX- V5-APEX	PPOX GenBank ID NM_000309.3
CPOX- APEX-Flag	CPOX GenBank ID NM_000097.5
CPOX$_{(1-120)}$-V5-APEX-CPOX$_{(121-453)}$	CPOX$_{(1-120)}$ is the mitochondrial targeting signal sequence of CPOX (*29*)

For bacterial expression, His_6-APEX and His_6-wtAPX were cloned into pTRC99 vector, under the control of *lac*O operator. In general, genetic constructs were prepared by directional ligation into restriction enzyme-cut vectors, overlap-extension PCR, and QuikChange mutagenesis. wtAPX = wild-type soybean ascorbate peroxidase (GenBank ID NM_001250856), a gift from Emma Raven (University of Leicester). APEX = wtAPX with three mutations (K14D, W41F, E112K). The HRP (horse radish peroxidase) gene was a gift from Frances Arnold (Caltech). Human CPOX and PPOX genes were obtained from the Broad Institute. The LACTB gene was from Timothy A. Brown and David A. Clayton (HHMI Janelia Farm Research Campus). Sub-cellular targeting sequences for the mitochondrial matrix (*9*), nucleus (*30*), cytosol (*31*), and plasma membrane (*32*) are based on previous literature. Note that HA tags are not used for immunofluorescence or Western blot detection in this work, because they are tyrosine-rich and hence damaged by the peroxidase/biotin-phenol reaction.

Reagents

Biotin conjugate **5** (biotin-XX-tyramide) was purchased from Invitrogen. Biotin conjugate **8** (EZ-link photo-activatable biotin) was purchased from Pierce. The other biotin conjugates were synthesized by Dr. Hyun-Woo Rhee according to this general scheme:



All products were purified by HPLC and characterized by Q-TRAP MS instrument (Applied Biosystems). Biotin-phenol conjugate **1** was characterized by ^{1}H-NMR.

Horseradish peroxidase (HRP) was purchased from Sigma. APEX and wtAPX were expressed in BL21-DE3 *E. coli* and purified as previously described (*9*).

In vitro enzyme kinetics

This assay (*9*) measures the formation of tetraguaiacol by absorbance at 470 nm, after one-electron oxidation of guaiacol into its phenoxyl radical. 20 nM of each enzyme was used, and guaiacol concentration ranged from 0.25-30 mM. For APEX and APX, 90 μM H_2O_2 was used to exceed its K_m of 20 μM (*33*). For HRP, 10 mM H_2O_2 was used to exceed its K_m of 1.53 mM (*34*). To achieve high reproducibility, we found it was important to pre-mix H_2O_2 and guaiacol, establish a baseline, and then add enzyme to initiate the reaction. Absorbance at 470 nm was measured for 1 min at 4 second intervals.

Mass spectrometry analysis of in vitro labeling

Reactions were assembled as follows: 500 μM amino acid (all 20 were purchased from Sigma), 500 μM biotin-phenol, 1 mM H_2O_2, and 10 nM HRP in PBS pH 7.4. HRP was added last to initiate the reaction. After 10 minutes at room temperature, reactions were quenched by addition of sodium azide (final concentration = 1 mM). Product mixtures were analyzed by LC-MS. The LC was a Shimadzu UFLC-XR equipped with a 4.6 x 50 mm C18 column running a gradient of 5% to 95% acetonitrile in 0.1% trifluoroacetic acid/water. The MS was an Applied Biosystems Q-TRAP.

Immunostaining and confocal fluorescence imaging

HEK293T cells were transfected and labeled with biotin-phenol as described below for proteomics. After labeling, cells were washed with a quencher cocktail (10 mM sodium ascorbate, 10 mM sodium azide, 5 mM Trolox) in DPBS (Dulbecco's Phosphate Buffered Saline) three times before fixation in 4% formaldehyde in PBS at room temperature for 15 min. Thereafter, cells were washed three times with PBS, and permeabilized with pre-chilled methanol at -20 °C for 5 min. Cells were blocked overnight with 1% casein in PBS ("blocking buffer") at 4 °C. For immunostaining, blocked cells were incubated with primary antibody (mouse anti-Flag (Sigma) at 1:300 dilution, mouse anti-V5 (Invitrogen) at 1:2000 dilution, or mouse anti-myc (Millipore) at 1:1000 dilution) in blocking buffer for 1 hr at room temperature. Samples were washed 4× 5 min with PBS at room temperature, then incubated with secondary antibody (goat anti-mouse AlexaFluor488 conjugate (Invitrogen) at 1:1000 dilution) and neutravidin- AlexaFluor647 or streptavidin- AlexaFluor568 (50 nM) in blocking buffer for 1 hr at room temperature. The neutravidin/streptavidin conjugates were prepared by coupling neutravidin (Invitrogen) or homemade streptavidin protein with AlexaFluor647-NHS or AlexaFluor568-NHS esters (Invitrogen) following the manufacturer's instructions. Labeled cells were washed 4× 5 min with PBS before imaging.

Fixed cells were imaged in DPBS at room temperature. Confocal fluorescence images were acquired on a Zeiss AxioObserver.Z1 microscope with a 63× oil-immersion objective, equipped with a Yokogawa spinning disk confocal head containing a Quad-band notch dichroic mirror (405/488/568/647 nm). YFP/ AlexaFluor488 (491 nm laser excitation, 528/38 emission filter), AlexaFluor568 (561 nm laser excitation, 617/73 emission filter), AlexaFluor647 (640 nm laser excitation, 700/75 emission filter), and DIC images were collected using Slidebook software (Intelligent Imaging Innovations). Images were acquired for 20 to 500 milliseconds and recorded on a Cascade II:512 camera. Fluorescence images in each experiment were normalized to the same intensity range.

STORM super-resolution fluorescence imaging

U2OS cells were plated on Lab-Tek™ II Chambered Coverglass (Thermo Scientific) pre-coated with human fibronectin. At 60-70% confluence, cells were transfected with mito-APEX (0.2 μg plasmid DNA per well) using Lipofectamine 2000 (0.5 μL per well). 4 hours after transfection, the media was replaced. 24 hrs later, cells were labeled with biotin-phenol as described below for proteomics. After 1 minute of labeling, samples were immediately quenched by aspirating off the probe/H_2O_2 solution and replacing with a quencher solution (10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in DPBS) for 1 min at room temperature. Thereafter, cells were washed three times with PBS, fixed with 3.7% formaldehyde at room temperature for 20 min, and permeabilized with pre-chilled methanol at -20 °C for 5 min. Finally, cells were blocked overnight with blocking buffer (1% casein in PBS) at 4 °C.

Samples were then stained with mouse anti-V5 antibody (Invitrogen) at 1:2000 dilution in blocking buffer for 1 hour at room temperature, then washed 4× 5 min with PBS, before incubating with a mixture of donkey anti-mouse AlexaFluor405/AlexaFluor647 conjugate (1:2000 dilution) and streptavidin-Cy3/Cy5 conjugate in blocking buffer for 5 min at room temperature. The antibody conjugate was prepared by amide-coupling donkey anti-mouse antibody (Jackson Immunoresearch) with AlexaFluor405-NHS and AlexaFluor647-NHS esters (Invitrogen). The streptavidin conjugate was prepared by coupling homemade streptavidin protein with Cy3-NHS and Cy5-NHS esters (Invitrogen). Labeled cells were washed 4× 5 min with PBS.

For STORM imaging, fixed cells were transferred to STORM imaging buffer (10% glucose (w/v), 50 mM Tris pH 8.2, 10 mM NaCl, 0.56 mg/mL glucose oxidase, 0.8 mg/mL catalase, and either 1% β-mercaptoethanol (v/v) or 7.7 mg/mL cysteamine). STORM images were acquired on a home-built setup constructed on an Olympus IX-81 stand. A polychroic mirror (Di01-R405/488/561/635, Semrock) was used to reflect lasers onto the sample. 405 nm and 561 nm laser lines (0.1 – 10 mW) were used alternately to activate fluorophore pairs, and a 642 nm laser line (140 mW) was used to image AlexaFluor647 and Cy5. Fluorescence emissions from AlexaFluor647 filtered with a dual-pass filter (FF01-577/690, Semrock). Samples were illuminated by lasers in the pseudo-TIRF mode, and images were collected on an EMCCD camera (Evolve, Photometrics). Super-resolution images were typically reconstructed from more than 10,000 frames of localizations for each channel. Data were analyzed using Insight3 software from Prof. Xiaowei Zhuang (Harvard University). The lateral resolution is determined by measuring straying dye single molecules that non-specifically adsorbs on the glass substrate (Figure 2-18).

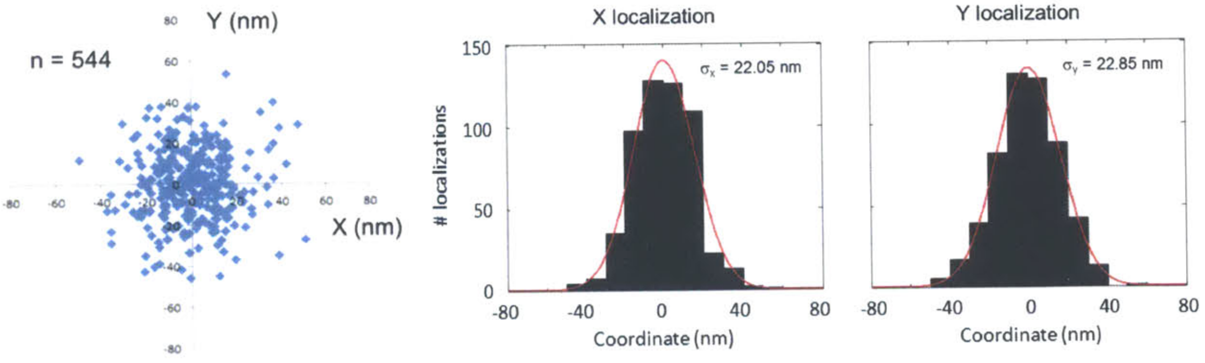


Figure 2-18. Determine the resolution of STORM setup.
Left: overlay of single-molecule clusters by their center of mass. Right: modeling with Gaussian in x and y dimensions.

Streptavidin blotting of whole cell lysate

HEK293T cells were transfected, labeled, and lysed as described below for proteomics. Lysates were separated on 10% SDS-PAGE gels, and then transferred to nitrocellulose membranes. Total protein was visualized by Ponceau S staining (10 minutes in 0.1% (w/v) Ponceau S in 5% acetic acid/water). Thereafter membranes were washed extensively to remove the Ponceau stain, and blocking was performed with 3% (w/v) BSA in PBS with 0.1% (v/v) Tween-20 ("blocking buffer") at 4 °C overnight. To detect biotinylated proteins, membranes were immersed in a solution of 0.3 mg/mL streptavidin-HRP (Thermo Scientific) at room temperature for 60 min before rinsing with blocking buffer 4× 5 min. Blots were developed with the SuperSignal West Pico reagent (Thermo Scientific), and imaged using an Alpha Innotech digital camera.

Imaging assay to determine the membrane-permeability of the phenoxyl radical

HEK293T cells were transfected with ^{W41A}APX targeted to either the cytosol (NES) or inner leaflet of the plasma membrane (CAAX). 24 hrs after transfection, cells were incubated with alkyne-phenol and 1 mM H_2O_2 for 1 min at room temperature. After washing cells three times with DPBS, click chemistry was performed on the cell surface by incubating with membrane-impermeant AlexaFluor647-picolyl azide conjugate (10 μM) and a pre-formed mixture of $CuSO_4$ (200 μM), THPTA (1 mM) and sodium ascorbate (2.5 mM) for 10 min at room temperature (*27*). Cells were quickly washed with DPBS three times and immediately imaged live by confocal microscopy as described above.

For comparison, click chemistry was performed also on fixed and permeabilized cells, in order to visualize the total population of alkyne-phenol-labeled molecules. For this labeling, cells were first blocked after fixation using 1% bovine serum albumin + 1% casein in PBS for 30 min at room temperature. Then cells were treated with AlexaFluor647-picolyl azide conjugate (2.5 μM) and a pre-formed mixture of $CuSO_4$ (1 mM), TBTA (100 μM) and sodium ascorbate (2.5 mM) in the same blocking buffer for 1 hour at room temperature.

Proteomic labeling with promiscuous BirA

The promiscuous BirA (pBirA) gene was obtained through site-directed mutagenesis of wild-type *E. coli* BirA to introduce the R118G mutation (*23,35*). HEK293T cells were transfected with pBirA or APEX plasmids using Lipofectamine 2000 (Invitrogen). After 4 hours transfection, labeling was initiated by addition of 50 μM biotin to the culture medium for 24 hours. Alternatively, 28 hours after transfection, 50 μM biotin was added for 1 min. Labeling was terminated by cell lysis in SDS-PAGE loading buffer.

References

1 Diwu, Z., Chen, C. S., Zhang, C., Klaubert, D. H. & Haugland, R. P. A novel acidotropic pH indicator and its potential application in labeling acidic organelles of live cells. *Chem Biol* **6**, 411-418 (1999).

2 Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. & Tsien, R. Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc Natl Acad Sci U S A* **95**, 6803-6808 (1998).

3 Eichelbaum, K., Winter, M., Diaz, M. B., Herzig, S. & Krijgsveld, J. Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat Biotechnol* **30**, 984-990 (2012).

4 Dai, J., Sloat, A. L., Wright, M. W. & Manderville, R. A. Role of phenoxyl radicals in DNA adduction by chlorophenol xenobiotics following peroxidase activation. *Chem Res Toxicol* **18**, 771-779 (2005).

5 Stubbe, J., Ge, J. & Yee, C. S. The evolution of ribonucleotide reduction revisited. *Trends Biochem Sci* **26**, 93-99 (2001).

6 Welinder, K. G. Superfamily of plant, fungal and bacterial peroxidases. *Current Opinion in Structural Biology* **2**, 388-393 (1992).

7 Sharp, K. H., Mewies, M., Moody, P. C. & Raven, E. L. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat Struct Biol* **10**, 303-307 (2003).

8 Sharp, K. H., Moody, P. C., Brown, K. A. & Raven, E. L. Crystal structure of the ascorbate peroxidase-salicylhydroxamic acid complex. *Biochemistry* **43**, 8644-8651 (2004).

9 Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically-encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143-1148 (2012).

10 Wishart, J. F. & Madhava Rao, B. S. *Recent trends in radiation chemistry.* (World Scientific, 2010).

11 Bendayan, M. Tech.Sight. Worth its weight in gold. *Science* **291**, 1363-1365 (2001).

12 Lee, S. W. *et al.* Introduction of tyramide signal amplification (TSA) to pre-embedding nanogold-silver staining at the electron microscopic level. *J Histochem Cytochem* **53**, 249-252 (2005).

13 Hunter, E. P., Desrosiers, M. F. & Simic, M. G. The effect of oxygen, antioxidants, and superoxide radical on tyrosine phenoxyl radical dimerization. *Free Radic Biol Med* **6**, 581-585 (1989).

14 Minamihata, K., Goto, M. & Kamiya, N. Protein heteroconjugation by the peroxidase-catalyzed tyrosine coupling reaction. *Bioconjug Chem* **22**, 2332-2338 (2011).

15 Rogers, M. S. *et al.* Cross-link formation of the cysteine 228-tyrosine 272 catalytic cofactor of galactose oxidase does not require dioxygen. *Biochemistry* **47**, 10428-10439 (2008).

16 Bhaskar, B. *et al.* A novel heme and peroxide-dependent tryptophan-tyrosine cross-link in a mutant of cytochrome c peroxidase. *J Mol Biol* **328**, 157-166 (2003).

17 Buse, G., Soulimane, T., Dewor, M., Meyer, H. E. & Bluggel, M. Evidence for a copper-coordinated histidine-tyrosine cross-link in the active site of cytochrome oxidase. *Protein science : a publication of the Protein Society* **8**, 985-990 (1999).

18 Meunier, S., Strable, E. & Finn, M. G. Crosslinking of and coupling to viral capsid proteins by tyrosine oxidation. *Chem Biol* **11**, 319-326 (2004).

19 Lad, L., Mewies, M. & Raven, E. L. Substrate binding and catalytic mechanism in ascorbate peroxidase: evidence for two ascorbate binding sites. *Biochemistry* **41**, 13774-13781 (2002).

20 Kotani, N. *et al.* Biochemical visualization of cell surface molecular clustering in living cells. *Proc Natl Acad Sci U S A* **105**, 7405-7409 (2008).

21 Bordo, D. & Argos, P. Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J Mol Biol* **217**, 721-729 (1991).

22 Badyal, S. K. *et al.* Conformational mobility in the active site of a heme peroxidase. *The Journal of biological chemistry* **281**, 24512-24520 (2006).

23 Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of cell biology* **196**, 801-810 (2012).

24 Snapp, E. L. *et al.* Formation of stacked ER cisternae by low affinity protein interactions. *The Journal of cell biology* **163**, 257-269 (2003).

25 Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810-813 (2008).

26 Niers, J. M., Chen, J. W., Weissleder, R. & Tannous, B. A. Enhanced in vivo imaging of metabolically biotinylated cell surface reporters. *Anal Chem* **83**, 994-999 (2011).

27 Uttamapinant, C. *et al.* Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew Chem Int Ed Engl* **51**, 5852-5856 (2012).

28 Brown, T. A. *et al.* Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction. *Mol Cell Biol* **31**, 4994-5010 (2011).

29 Dailey, T. A., Woodruff, J. H. & Dailey, H. A. Examination of mitochondrial protein targeting of haem synthetic enzymes: in vivo identification of three functional haem-responsive motifs in 5-aminolaevulinate synthase. *Biochem J* **386**, 381-386 (2005).

30 Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. A Short Amino-Acid Sequence Able to Specify Nuclear Location. *Cell* **39**, 499-509 (1984).

31 Wen, W., Meinkoth, J. L., Tsien, R. Y. & Taylor, S. S. Identification of a Signal for Rapid Export of Proteins from the Nucleus. *Cell* **82**, 463-473 (1995).

32 Keppler, A., Pick, H., Arrivoli, C., Vogel, H. & Johnsson, K. Labeling of fusion proteins with synthetic fluorophores in live cells. *P Natl Acad Sci USA* **101**, 9955-9959 (2004).

33 Dalton, D. A., Diaz del Castillo, L., Kahn, M. L., Joyner, S. L. & Chatfield, J. M. Heterologous expression and characterization of soybean cytosolic ascorbate peroxidase. *Arch Biochem Biophys* **328**, 1-8 (1996).

34 Sock Ying Tham and C.H. Foo, Malaysian Journal of Biochemistry and Molecular Biology (2005) 12, 8-13

35 Choi-Rhee, E., Schulman, H. & Cronan, J. E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein science : a publication of the Protein Society* **13**, 3043-3050 (2004).

Chapter 3 Spatially-resolved organellar proteomic mapping

Work described in this chapter has not yet been published. Most experiments described were performed in collaboration with Dr. Hyun-Woo Rhee. Dr. Namrata Udeshi and Dr. Steven Carr from Broad Institute performed mass spectrometry proteomics experiments. Prof. Vamsi Mootha provided valuable advice on proteomic data analysis. Jeffrey Martell and Dr. Eliza Vasile performed electron microscopy analysis.

3.1 Introduction

Having established that APEX catalyzes covalent biotinylation of endogenous proteins in living cells using biotin-phenol as a substrate, we used it to map the proteome of the mitochondrial matrix and endoplasmic reticulum lumen.

The mitochondrion is the power house of the cell, producing more than 90% of cellular ATP in the heart (*1*). It is also well established that cell apoptosis is initiated by protein release from the inter-membrane space of the mitochondria into the cytosol (*2*). As discussed in Chapter 1, all previous proteomic studies of mitochondria have used density centrifugation as a first step to purify intact mitochondria, which is a procedure known to disrupt mitochondrial structure and function (*3*). For this reason, the specificity of MitoCarta, the most comprehensive inventory of mitochondrial proteins (*4*), was improved by integrating proteomic data with literature curation, fluorescence imaging data, mitochondrial targeting sequence prediction, transcriptional analysis, and phylogenetic profiling. In contrast, high-quality inventories of sub-mitochondrial compartments are not currently available.

The endoplasmic reticulum is another important organelle in the cell. It serves as the entry point of secretory pathway, both for secreted proteins (such as hormones and extracellular matrix proteins) and for plasma membrane proteins (such as receptors, ion channels and transporters). It is also involved in the synthesis of lipids and cholesterol (*5*). The endoplasmic reticulum quality control and endoplasmic reticulum-associated degradation are all important biological processes that have gained increasing attention over the past decades (*6*).

3.2 Proteomic mapping of the mitochondrial matrix

Promiscuous protein labeling with APEX in mitochondrial matrix

We wished to test whether our peroxidase-mediated live cell labeling method could yield a more accurate and complete catalog of mitochondrial matrix proteins. As outlined in Figure 3-1, APEX is transiently expressed in HEK293T cells and is genetically targeted to the mitochondrial matrix by fusing to the targeting sequence of cytochrome c oxidase subunit 4 (*7*).

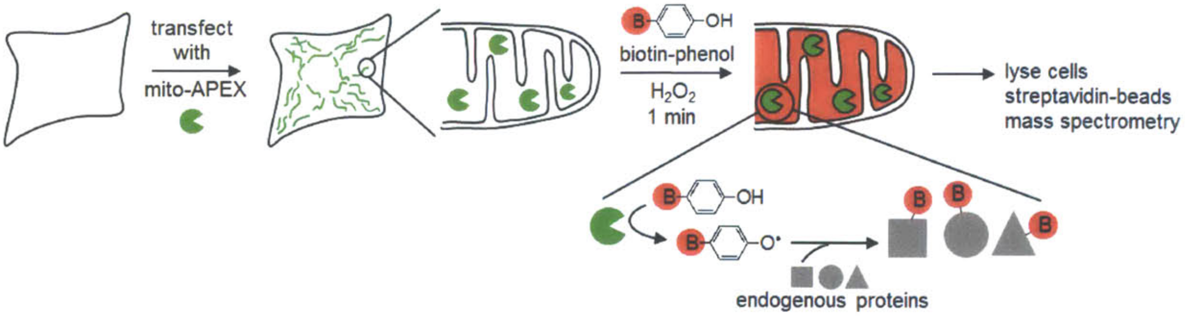
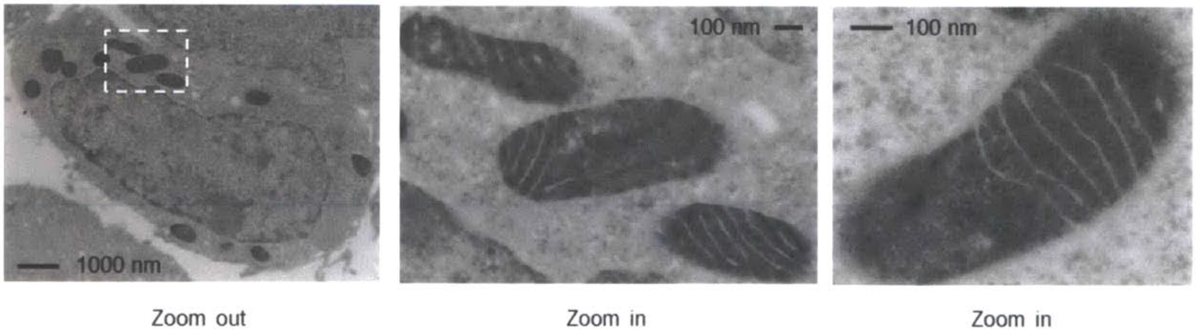


Figure 3-1. Scheme of promiscuous labeling in mitochondrial matrix.
The APEX peroxidase is genetically targeted to the mitochondrial matrix (or other cellular compartment of interest), via fusion to a 24-amino acid mitochondrial targeting peptide. Labeling is initiated by addition of biotin-phenol and H_2O_2 to live cells for 1 min. Cells are then lysed, and biotinylated proteins are recovered with streptavidin-coated beads, eluted, separated on a gel, and identified by mass spectrometry. The peroxidase-generated phenoxyl radical is short-lived and membrane-impermeant and hence covalently tags only neighboring and not distant endogenous proteins.

We confirmed the mitochondrial localization of APEX by electron microscopy because conventional light microscopy cannot resolve sub-mitochondrial spaces. We took advantage that APEX can also act as an EM reporter and performed DAB staining. We confirmed with EM-resolution that APEX is correctly localized to the mitochondrial matrix (Figure 3-2).



Zoom out Zoom in Zoom in

Figure 3-2. Electron microscopy (EM) showing the localization of mito-APEX in the mitochondrial matrix of HEK293T cells.
Contrast (dark regions) was generated by APEX-catalyzed diaminobenzidine polymerization, followed by OsO_4 staining. Zoom out view shows specific staining of mitochondria in transfected cells. Zoom in views highlight staining in the mitochondrial matrix.

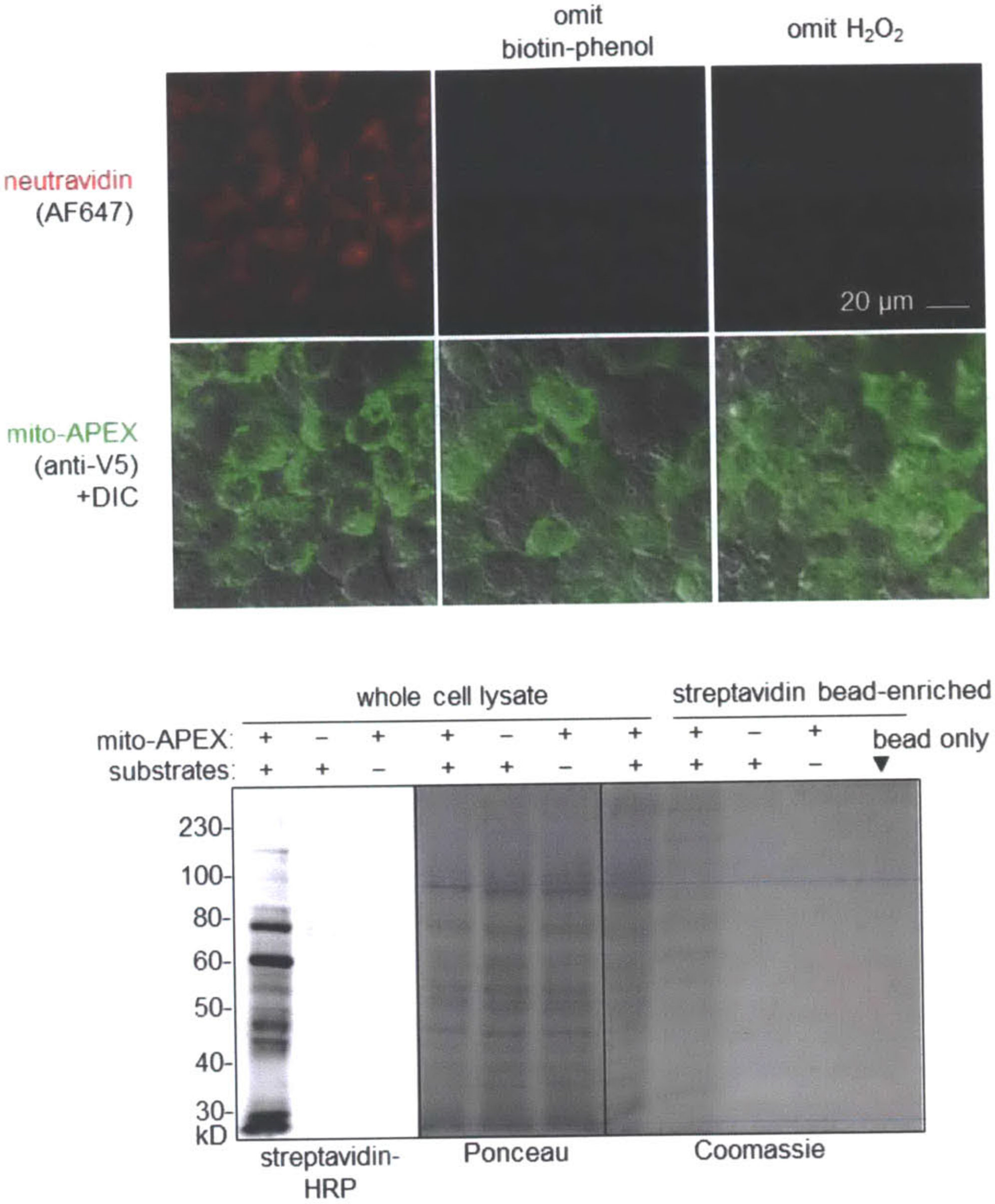

omit biotin-phenol
omit H2O2
neutravidin (AF647)
20 µm
mito-APEX (anti-V5) +DIC
whole cell lysate
streptavidin bead-enriched
mito-APEX: + − + + − + + + − +
substrates: + + − + + − + + + −
bead only
230-
100-
80-
60-
50-
40-
30-
kD
streptavidin-HRP
Ponceau
Coomassie

Figure 3-3. Labeling the mitochondrial matrix proteome with biotin-phenol.
Top: confocal fluorescence imaging of biotinylated proteins. Mitochondrial matrix-targeted APEX in HEK293T is visualized by anti-V5 (AlexaFluor488) staining. Proteins biotinylated by APEX (500 µM biotin-phenol and 1 mM H_2O_2, 1 minute) are stained with neutravidin (AlexaFluor647). Only transfected cells (green) displayed biotin signal (red) in the mitochondria. Negative control omitting either reagent (biotin-phenol or H_2O_2) abolished labeling. Bottom: gel analysis of biotinylated mitochondrial matrix proteins, before and after streptavidin bead-enrichment. Substrates are biotin-phenol and H_2O_2. Consistent with imaging results shown above, labeling occurred only when both enzyme and substrates were present. Ponceau stain showed equal amount of loading of cell lysate into each lane. Coomassie stain showed that proteins enriched from streptavidin-coated beads had a different pattern on the gel, as compared to total cell lysate. Mammalian cells have four endogenously biotinylated proteins (*8*), which are seen in the negative control lanes of the streptavidin blot.

Next we characterized mitochondrial matrix-targeted APEX labeling with biotin-phenol in living HEK293T cells by fluorescence imaging. Promiscuous labeling was achieved by addition of biotin-phenol and H_2O_2. Negative controls in Figure 3-3 show that labeling was not detected when either biotin-phenol or H_2O_2 was omitted. Untransfected cells in the same field of view also did not show streptavidin staining, indicating that endogenous mammalian peroxidases do not activate biotin-phenol to a significant degree. After labeling HEK293T cells for 1 min with APEX, biotin-phenol, and H_2O_2, streptavidin blotting detects numerous biotinylated proteins in the cell lysate (Figure 3-3). Negative controls omitting either the enzyme or the substrate gave no enriched proteins (2nd and 3rd lanes to the right in Figure 3-3, bottom). Comparison of Coomassie stain of streptavidin bead-enriched samples and whole cell lysate (4th and 5th to the right in Figure 3-3, bottom) shows different protein bands, indicating that a subset of proteins are selectively enriched.

Determine mitochondrial matrix proteome

We proceeded to purify labeled proteins using streptavidin-coated magnetic beads. Although these beads bind fairly specifically to biotinylated molecules, they nevertheless have the capacity to adsorb unlabeled proteins through hydrophobic interactions. This creates a protein background which can easily flood sensitive MS readouts. In order to facilitate subtraction of this background due to non-specific binding to streptavidin beads, we established "heavy" and "light" isotope-labeled HEK293T cell cultures using the SILAC approach (stable isotope labeling by amino acids in cell culture (*9*)), as described in Chapter 1.

The SILAC labeling scheme is outlined in Figure 3-4. Two replicate heavy SILAC cultures were labeled by pre-loading biotin-phenol for 30 min, then adding 1 mM H_2O_2 to initiate biotinylation. After 1 minute, cells were lysed using a mixture containing radical quenchers (Trolox and ascorbate) and a peroxidase inhibitor (sodium azide) to halt biotinylation. Two light SILAC cultures were processed in parallel as negative controls. These light cultures were treated identically to the heavy cultures, except that in the first light sample, APEX was omitted, and in the second light sample, biotin-phenol and H_2O_2 were omitted. In this way, the first replicate subtracts background from any endogenous peroxidase activity, while the second replicate normalizes the effect of APEX expression on the endogenous protein expression levels.

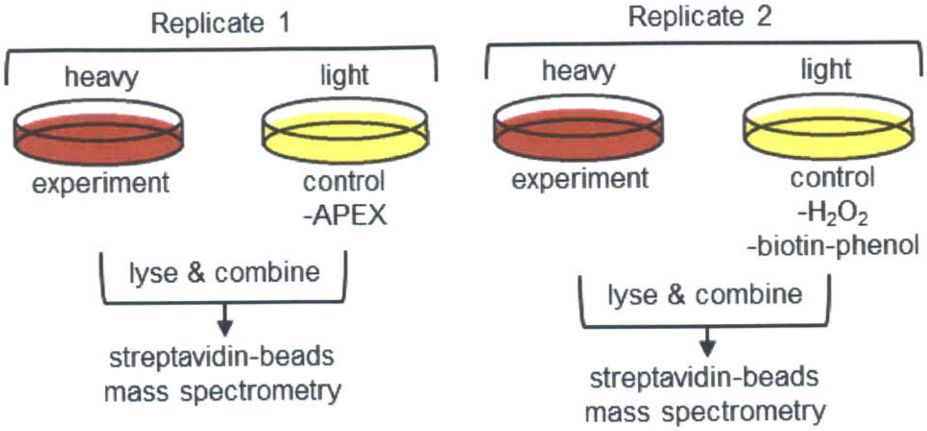

Replicate 1
heavy
light
experiment
control
-APEX
lyse & combine
streptavidin-beads
mass spectrometry
Replicate 2
heavy
light
experiment
control
-H_2O_2
-biotin-phenol
lyse & combine
streptavidin-beads
mass spectrometry

Figure 3-4. SILAC labeling scheme.
In replicate 1, the negative control state is no APEX transfection. This could subtract background labeling caused by endogenous peroxidase activity. In replicate 2, the negative control is no biotin-phenol/H_2O_2 labeling. This could eliminate perturbation of protein expression level due to expression of mitochondrial matrix-targeted enzyme. Heavy medium contains $^{13}C_6,^{15}N_2$-Lys and $^{13}C_6,^{15}N_4$-Arg.

Immediately after labeling and quenching, cells were pelleted and stored at -80 °C. After cell lysis, heavy and light lysates were first combined, and then enriched with streptavidin beads. Proteins eluted from the beads were separated on a polyacrylamide gel, and subsequently cut into gel bands. Each gel bands were treated with trypsin to digested proteins into peptides, which were extracted from the gel and analyzed by LC-MS/MS (liquid chromatography, tandem MS) using a Q Exactive instrument.

To determine the mitochondrial matrix proteome, we calculated heavy/light (H/L) SILAC ratios for each MS-identified protein, and chose cut-offs with <10% false-positive rates (FPR) (Figure 3-5). Replicate 1 contained 527 proteins above the H/L cut-off of 2.07, while replicate 2 contained 579 proteins with H/L ratio above 2.37. We took the intersection of these two groups to be our mitochondrial matrix proteome (495 proteins). Figure 3-6 shows that the correlation between replicate 1 and replicate 2 H/L ratios is very high (r = 0.91) for proteins in our matrix proteome.

Because these 495 proteins in our mitochondrial matrix proteome were obtained in a fundamentally different approach than traditional method, we were very interested in comparing this list with the literature in terms of both specificity and sensitivity. This comparison is described in the next two sections.

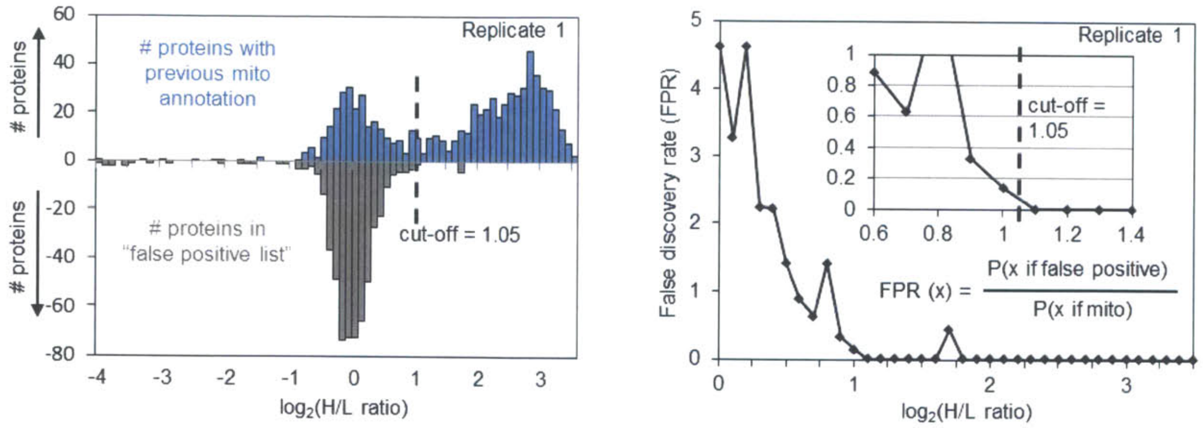


Figure 3-5. Determine the cutoff point for enrichment.
Left: distribution of SILAC ratios (H/L) from the replicate 1 experiment. The top panel in blue shows "true positive" analysis: the number of proteins within each SILAC ratio range (bin size 0.1) with prior mitochondrial annotation. The bottom panel in grey shows "false positive" analysis: the number of proteins within each SILAC ratio range appearing in a hand-curated list of 2410 non-mitochondrial proteins ("false positive list") (*4*). Right: false-positive rate (FPR) as a function of SILAC enrichment factor (H/L). At a given SILAC enrichment level, FPR is defined as the ratio of the (conditional) probability of finding a false-positive protein and the (conditional) probability of finding a mitochondrial protein. Therefore, FPR measures the likelihood of finding a false positive versus a true mitochondrial protein. We selected an FPR of 0.1 (10%) as our cut-off point, i.e. 10 times more likely for a protein to be mitochondrial than a false positive non-mitochondrial protein.

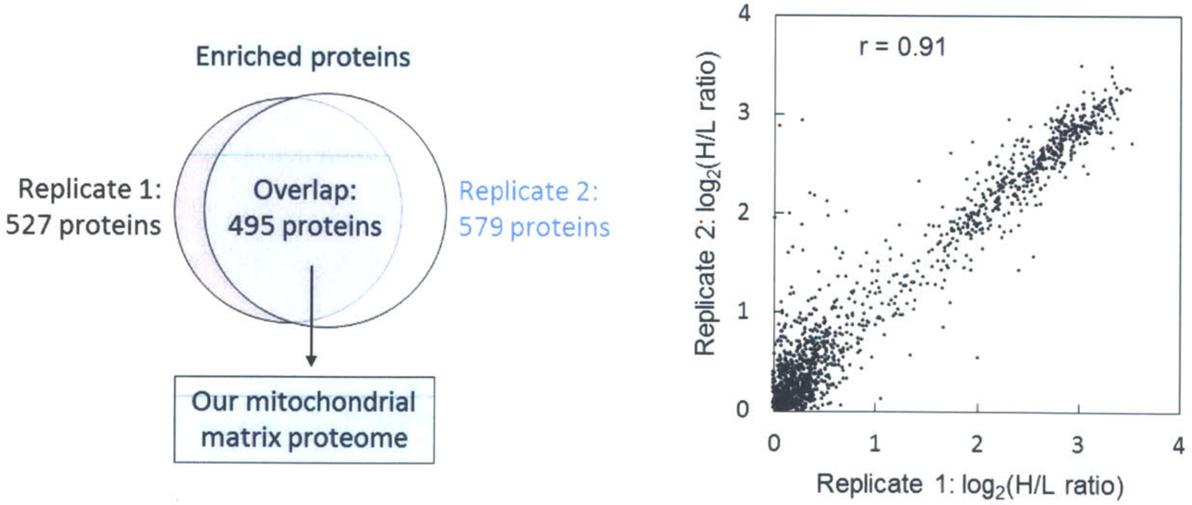


Figure 3-6. Defining our mitochondrial matrix proteome.
Left: our mitochondrial matrix proteome is defined as the intersection of the two enriched populations in both replicates. Right: good correlation between SILAC ratios for replicates 1 and 2.

Specificity analysis of mitochondrial matrix proteome

In the previous chapter, we assessed the spatial specificity of promiscuous protein labeling with APEX using a series of imaging assays. Those experiments suggested that the labeling is restricted to the same side of the membrane where the enzyme is targeted. Another measure of its spatial specificity would come from analysis of our matrix proteome.

First of all, we assessed the specificity of enrichment for mitochondrial proteins. Taking advantage of existing knowledge about mitochondria biology from the literature, we have assembled a list of proteins with previous mitochondrial annotation from Gene Ontology database (*10*), MitoCarta (*4*), and specific literature research. We found that 94% of our 495 proteins can be found in this list, compared to merely 8% of proteins in the entire human proteome (Figure 3-7). The 31 enriched proteins not on the list were called "mito orphans". While the localization of some of these simply has not been studied, the others have been previously known to localize outside mitochondria. Indeed, as shown later in Figure 3-12, immunofluorescence study on six of these "mito orphans" revealed complete or partial overlap with mitochondrial markers, strongly suggesting mitochondrial association.

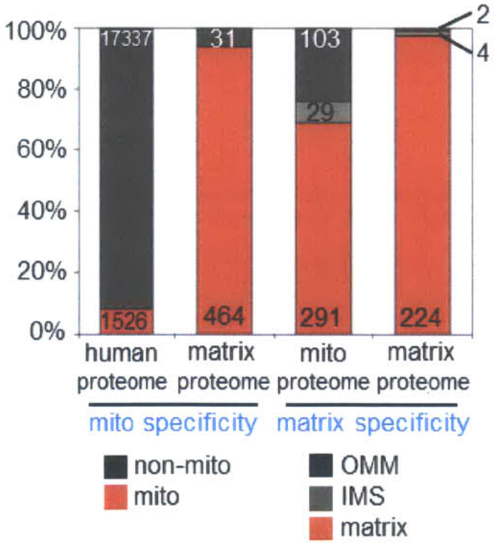


Figure 3-7. Our proteome is specific to mitochondrial matrix proteins.
Left two columns show the fraction of proteins in the entire human proteome (column 1) or in our matrix proteome (column 2) with prior mitochondrial annotation in the Gene Ontology Cell Component database (*10*), MitoCarta (*4*), or literature. Right two columns show the distribution of proteins (with prior sub-mitochondrial localization information) throughout the mitochondria, for all known mitochondrial proteins (column 3), and for our matrix proteome (column 4). OMM is the outer mitochondrial membrane. IMS is the inter-membrane space. The numbers of proteins, rather than fractions, in each category is indicated in the column.

We next assessed the specificity of enriching mitochondrial matrix proteins. As shown in Figure 3-7, of the 230 proteins in our list with prior information regarding sub-mitochondrial localization (matrix, IMS, or outer membrane (OMM)), 97% are known to be matrix proteins, compared to 69% for all characterized mitochondrial proteins. Our dataset is therefore strongly enriched for mitochondrial matrix proteins. Note that we excluded inner membrane (IMM) proteins from this specificity analysis because some of them can be labeled by matrix-targeted enzyme if they are exposed to the matrix side. In this way, their labeling is not as indicative for spatial specificity as IMS and OMM proteins.

To further assess the specificity of our matrix proteome, we examined several protein complexes of the inner mitochondrial membrane that have been structurally and/or topologically characterized. Figure 3-8 illustrates complexes I-IV and ATP synthase of the electron transport chain. Only subunits that come into contact with the matrix space and are sterically exposed appear in our proteome (shaded red). Subunits that are embedded in the membrane or face the inter-membrane space (IMS) were not detected.

A similar analysis of the proteins of the TOM/TIM/PAM protein import pathway shows that, again, only matrix-exposed proteins are labeled (Figure 3-9). Our data clearly indicate that biotinylation catalyzed by matrix-localized APEX did not reach beyond the inner mitochondrial membrane.

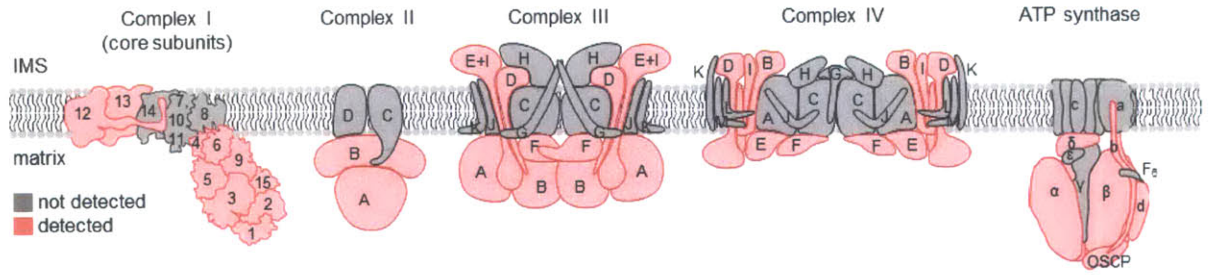


Figure 3-8. Analysis of labeling specificity for inner mitochondrial membrane protein complexes. The protein subunits of Complexes I-IV and ATP synthase, for which structural information is available, are illustrated (based on PDB IDs 1ZOY, 1L0L, 1OCC and references (*11,12*)). Subunits detected in our mitochondrial matrix proteome are shaded red; those not detected are shaded grey. Note that structural information is not available for all 45 subunits of Complex I, so some subunits that appear exposed here may not be exposed in the complete complex.

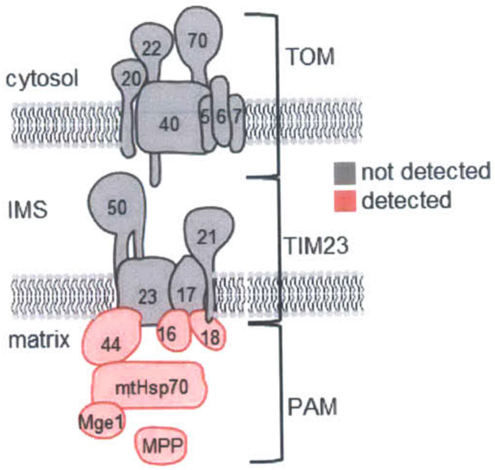


Figure 3-9. Analysis of labeling specificity for mitochondrial translocase complexes.
Same analysis as in Figure 3-8, for proteins of the TOM/TIM23/PAM mitochondrial protein import machinery (*13*) that spans the outer and inner mitochondrial membranes.

Depth of coverage analysis of mitochondrial matrix proteome

To analyze our depth of coverage, we searched our matrix proteome for five groups of well-established matrix proteins (Figure 3-10). Within each group, 80-90% of the proteins were detected. Interestingly, we found that nearly identical subsets of proteins within each group were detected in replicate 1 and replicate 2. As shown in Figure 3-10, each mitochondrial matrix protein in these five categories has similar levels of enrichment in two replicates. This argues that the coverage/sensitivity of our method is in fact quite high, but only for ~85% of proteins.

We sought to understand why certain matrix proteins were consistently not detected. We quickly ruled out lack of or low expression levels (Figure 3-11). Lack of surface-exposed tyrosines is also unlikely to be the cause, because we earlier estimated that >90% of proteins have one or more surface-exposed tyrosine (see argument in Section 2, Chapter 2, and reference (*14*), the caveat of this analysis is that it is based on sampling crystallographic structures of proteins from various sources, yet the surface tyrosine exposure may be different for the subset of mitochondrial matrix proteins). It is still possible that these non-labeled matrix proteins are sterically buried within macromolecular complexes in the context of living cells.

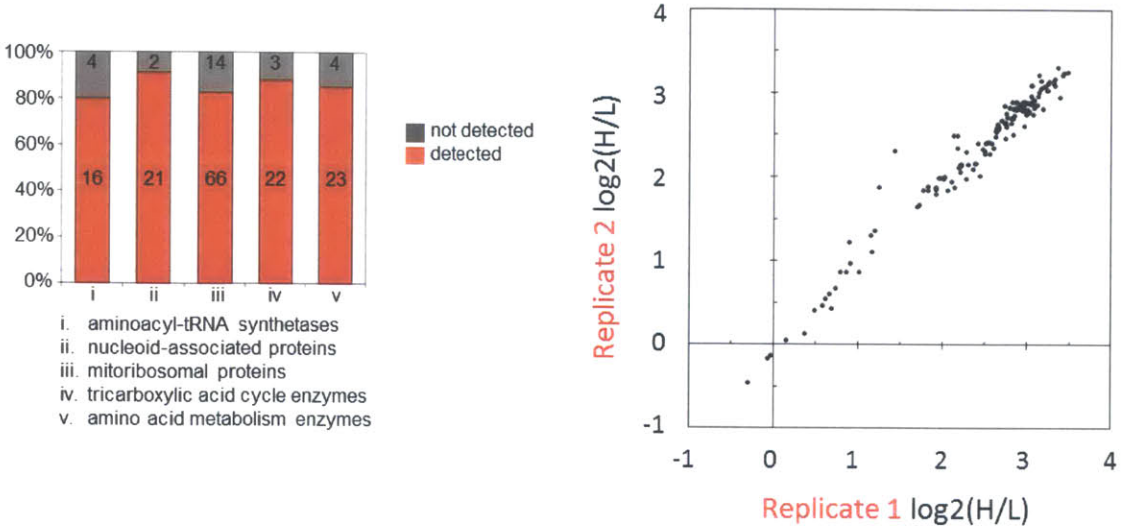


Figure 3-10. Depth of coverage of our mitochondrial matrix proteome.
Left: five groups of well-established mitochondrial matrix proteins were analyzed (i-v). For each group, 80-91% of proteins were detected in our matrix proteome. Right: correlation analysis shows that nearly identical subsets of proteins within each group were detected in the replicate 1 and replicate 2 experiments.

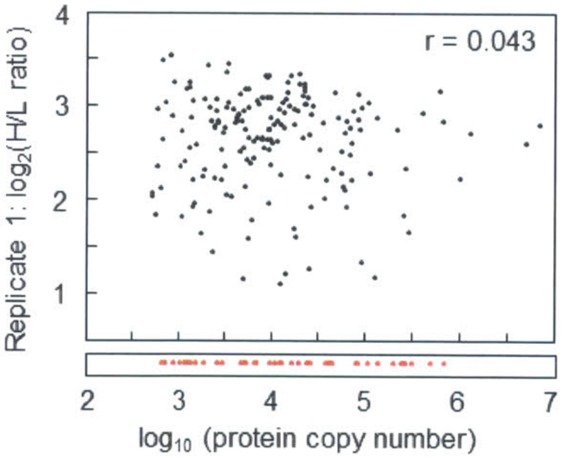


Figure 3-11. Correlation between SILAC ratio and protein expression level in U2OS cells.
We limited our analysis to soluble matrix proteins. SILAC ratios are from replicate 1. Protein expression level data are from reference (*15*).The bottom strip, in red, shows protein expression levels for matrix proteins from Figure 3-10 that are *not* detected in our matrix proteome. Enrichment ratio is not correlated with protein copy number (r = 0.043).

Insights from the proteomic dataset

Our mitochondrial matrix proteome provides a number of novel insights. First, 6% of our list, or 31 proteins, have not previously been linked to mitochondria, including the purine biosynthesis enzyme PAICS and several proteins related to RNA processing and modification (RPUSD3, TRUB2, PUSL1 related to tRNA pseudouridine synthase and METT5D1, METTL8, TRMT61B, TRMT11, NSUN3 related to tRNA methyltransferase). Nine proteins in our orphan list fall just below the cut-off point for MitoCarta (*4*), and another five proteins have predicted mitochondrial targeting sequences. Two orphans lack any prior annotation or characterization (C2orf69 and C15orf61). Due to the extremely high specificity of our proteome, and by the fact that many of our orphans have high H/L ratios in both replicates, corresponding to false-positive rates near 0%, it is highly likely that these orphans are *bona fide* mitochondrial matrix proteins, at least in HEK293T cells.

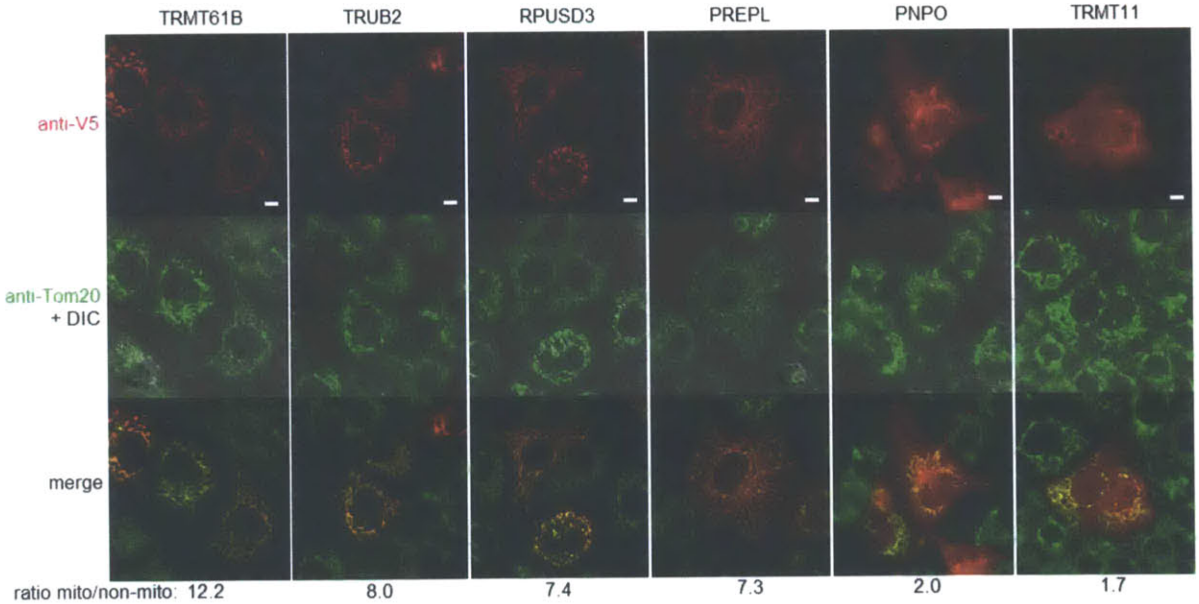


Figure 3-12. Immunofluorescence imaging analysis of a subset of "mito orphan" proteins. Genes encoding for these proteins were obtained from Broad Institute and confirmed by sequencing. Proteins are tagged with V5 epitope tag at the C-terminus. Anti-Tom20 immunostain is used here as a marker of mitochondria. The extent of mitochondrial localization was quantified by taking the intensity ratio of fluorescence signal that falls in the mitochondrial regions over that outside these regions.

Using immunofluorescence imaging, we did follow up studies on some of these proteins. The extent of mitochondrial localization was quantified by taking the ratio of fluorescence intensity in the mitochondrial regions over that outside these regions. We found that TRMT61B, TRUB2, RPUSD3, and PREPL completely co-localize with mitochondrial marker, while PNPO and TRMT11 partially overlaps with the marker (Figure 3-12). This study strongly supports our claim that these orphans are true mitochondrial proteins in HEK293T cells.

Second, we found that 240 proteins in our proteome have previously been associated with mitochondria but have not been known to localize to the matrix. Unsurprisingly, these include proteins related to DNA replication and repair (5 proteins), gene transcription (6 proteins), RNA processing (16 proteins), protein synthesis and folding (24 proteins), and proteolysis (8 proteins). In addition, we observed 25 metabolic enzymes (related to heme, folate, ubiquinone, and lipid production) and 15 proteins involved in redox homeostasis. Notably, 11 proteins related to apoptosis were found in our list.

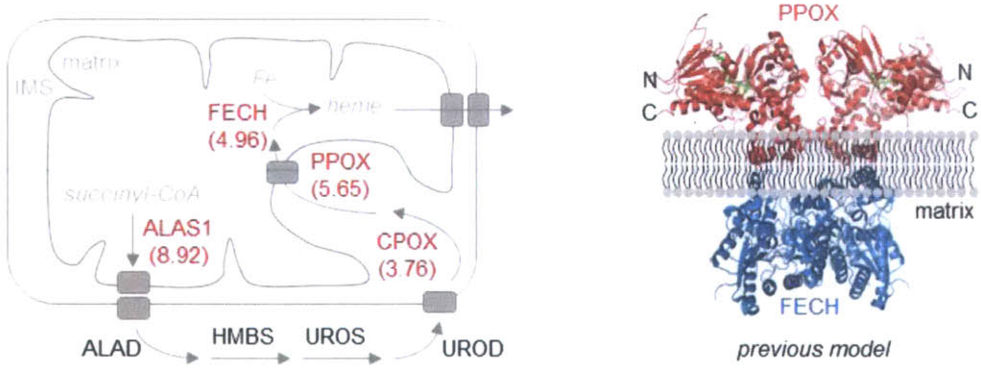


Figure 3-13. Literature model of CPOX and PPOX localization within the mitochondria.
Left: sub-mitochondrial localization of the eight core enzymes involved in heme biosynthesis (*18,19*). Four of these enzymes are detected in our matrix proteome and are colored red (with H/L ratios from replicate 1 shown in parentheses). We do not detect the four cytosolic enzymes. Drawing adapted from (*18*). Right: previously proposed model (*19,20*) based on computation and protease-accessibility assays (*16*) showing the docking of a PPOX dimer and a FECH (ferrochelatase) dimer through the inner mitochondrial membrane. N- and C-terminal ends of PPOX are labeled. Our data contradict this model because the PPOX-APEX EM images in (C) show that the C-terminus of PPOX is in the matrix, not the IMS. PPOX and FECH structures are from PDB IDs 1SEZ and 2QD1.

Third, our matrix proteome contains two proteins previously annotated as outer mitochondrial membrane proteins (COASY, SAM50) and four proteins previously annotated as residing in the IMS (CPOX, PPOX, PNPT1, CHCHD3). It is possible that these proteins are dual-localized within mitochondria, or that previous annotations of sub-mitochondrial localization are incorrect. We followed up on two of these proteins, coproporphyrinogen-III oxidase (CPOX) and protoporphyrinogen oxidase (PPOX), due to their importance in heme biosynthesis (Figure 3-13). Protease-treatment of purified mitochondria has previously suggested that both CPOX and PPOX are located in the IMS (*16,17*), but we strongly detect both proteins in our matrix proteome (H/L ratios > 3.7 in both replicates).

To resolve the discrepancy between our findings and previous studies, we fused both enzymes directly to APEX for EM analysis. Figure 3-14 shows that PPOX-APEX gives clear staining in the matrix. Since the diaminobenzidine polymer generated by APEX does not cross membranes (*7*), we conclude that the C-terminus of PPOX is located in the mitochondrial matrix. In contrast, both fusions of APEX to CPOX give no detectable matrix stain and instead highlight clusters within some regions of the IMS. Therefore, the N- and C-terminal ends of CPOX do not appear to localize to the matrix; our detection of CPOX in the matrix may be due to an internal region of CPOX being matrix-exposed.

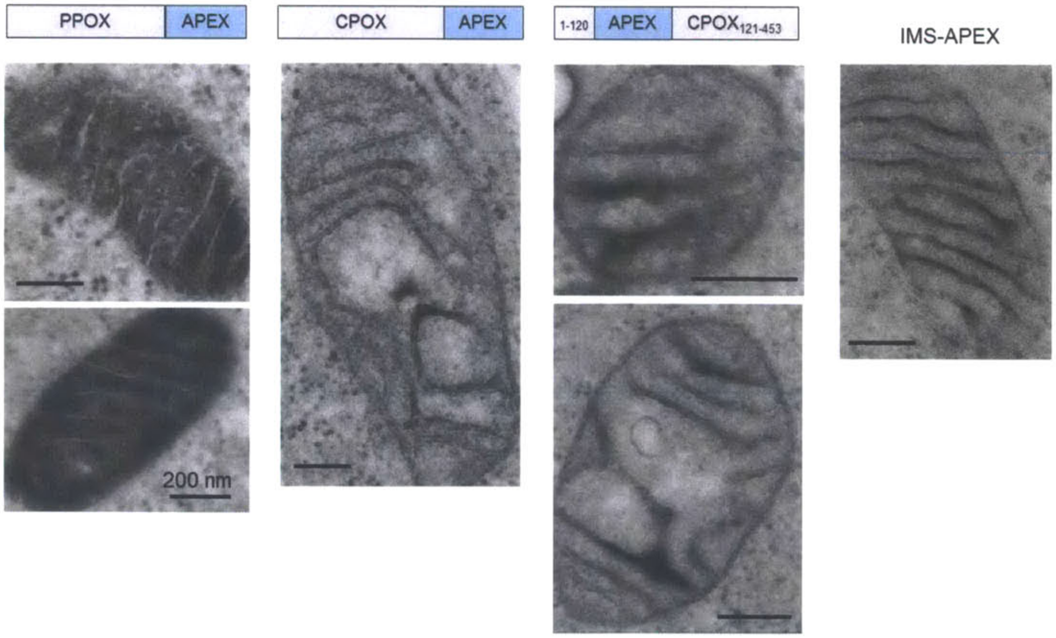


Figure 3-14. EM study of sub-mitochondrial localization of the heme biosynthesis enzymes CPOX and PPOX.
This figure was generated by Jeffrey Martell and Dr. Eliza Vasile at Koch Institute. APEX was fused to the C-terminus of full-length PPOX and CPOX (*21,22*). In the third construct, APEX was inserted into full-length CPOX after the mitochondrial leader sequence (*22*). EM contrast was generated from APEX with diaminobenzidine and OsO_4. As a control, an EM image is shown for APEX targeted to the IMS with the 68 amino acid N-terminal signal peptide of LACTB (*23*). All scale bars, 200 nm.

Our proteomic results on endogenous PPOX and our EM imaging of recombinant PPOX-APEX are not consistent with a previous literature model in which PPOX interacts with ferrochelatase (FECH), the last enzyme of heme biosynthesis, through the inner mitochondrial membrane (Figure 3-13) (*20*). If PPOX resides in the matrix, as we believe it does, then this localization has implications for the nature of its interactions with FECH and CPOX, as well as the mechanism by which the substrate of PPOX (protoporphyrinogen IX) is transported across the inner mitochondrial membrane (perhaps via a novel transporter) (*19*).

Detection of biotin-labeled peptides

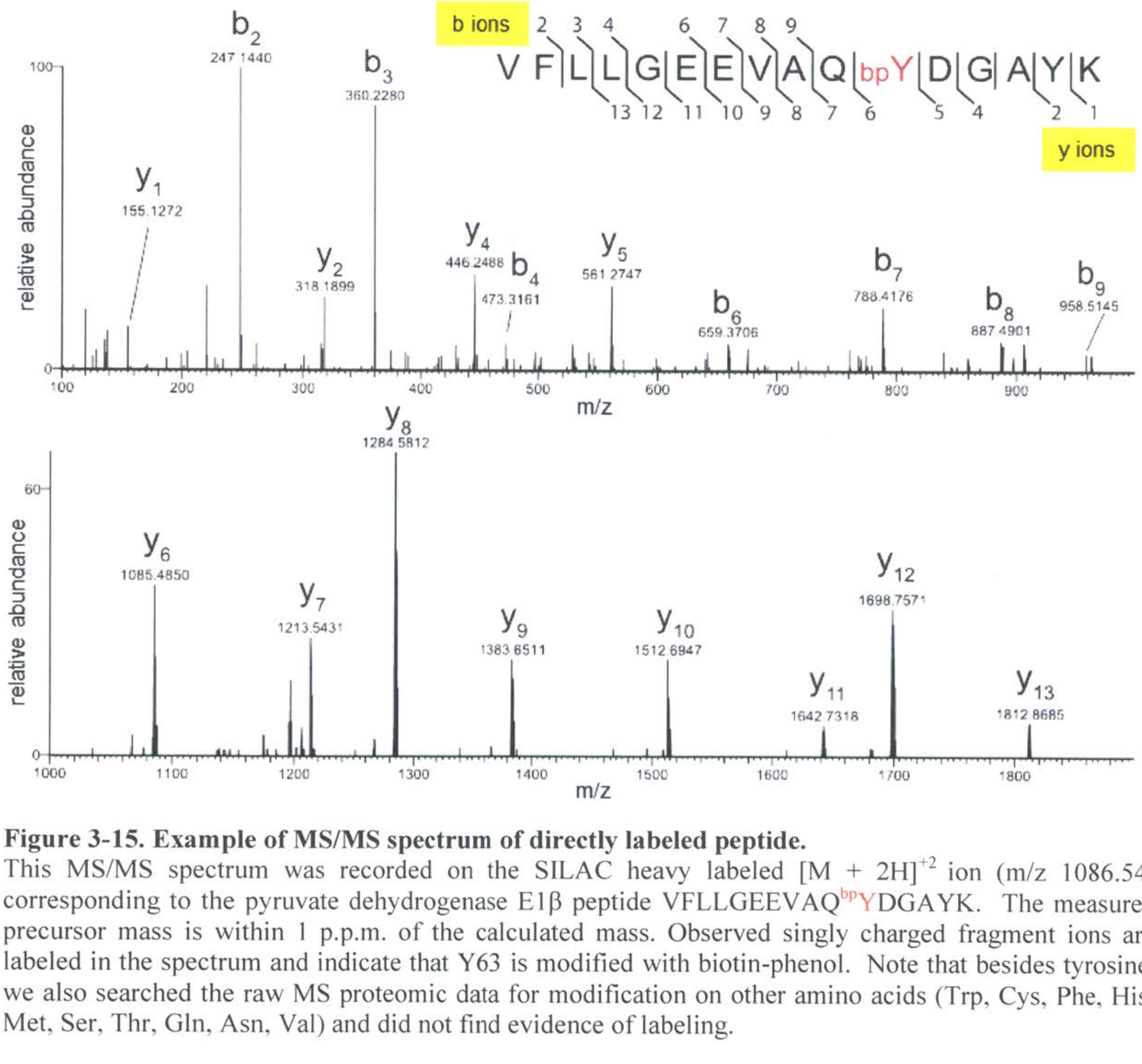


Figure 3-15. Example of MS/MS spectrum of directly labeled peptide.
This MS/MS spectrum was recorded on the SILAC heavy labeled $[M + 2H]^{+2}$ ion (m/z 1086.54) corresponding to the pyruvate dehydrogenase E1β peptide VFLLGEEVAQ^{bp}YDGAYK. The measured precursor mass is within 1 p.p.m. of the calculated mass. Observed singly charged fragment ions are labeled in the spectrum and indicate that Y63 is modified with biotin-phenol. Note that besides tyrosine, we also searched the raw MS proteomic data for modification on other amino acids (Trp, Cys, Phe, His, Met, Ser, Thr, Gln, Asn, Val) and did not find evidence of labeling.

Taking advantage of the sequencing power of MS/MS data, we were able to detect 167 directly biotinylated peptides and to verify that the labeling site is on a tyrosine residue. These peptides are derived from a subset of proteins in our matrix proteome. One example of MS/MS analysis is depicted in Figure 3-15.

We analyzed the peptides that map to soluble proteins with known structure, or membrane proteins with known topology. Figure 3-16 shows that peroxidase-catalyzed biotinylation favors surface exposed tyrosine sites for the soluble proteins: the biotinylation sites map to surface-exposed rather than buried tyrosine residues in 58 of 60 cases. In addition, we found 15 biotinylated peptides that are derived from 11 transmembrane proteins. The biotinylation site maps to the matrix-exposed side of the protein rather than the IMS side in 14 out of 15 cases. Three examples are shown in Figure 3-17.

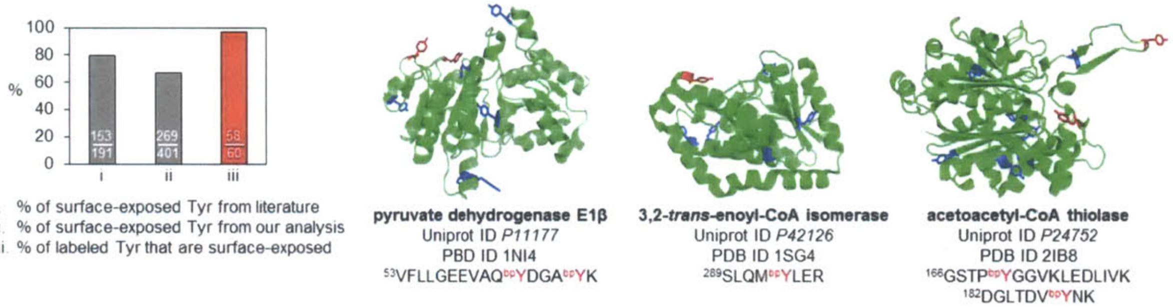


Figure 3-16. Mapping biotinylated peptides to soluble matrix proteins.
Left: occurrence of surface-exposed tyrosine residue in soluble proteins with known structure. (i) is from reference (*14*). (iii) is based on analysis of the 60 biotinylated peptides we detect that map to soluble proteins with known structure (39 proteins). (ii) is based on all the tyrosines (biotinylated or not, 401 in total) within these 39 proteins. The criterion for "surface-exposed" is having >10 Å^2 solvent-accessible area (*14*). Right: protein structures showing the positions of five of the biotinylated tyrosines (colored red, bp = biotin-phenol modification). Non-labeled tyrosines within each protein are colored blue. Sequences of detected biotinylated peptides are shown beneath each protein structure.

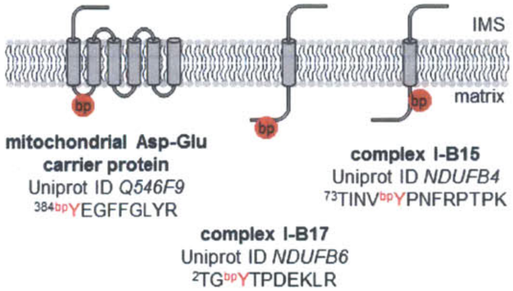


Figure 3-17. Mapping biotinylated peptides to transmembrane proteins.
Shown here are three specific examples. Biotinylation sites are depicted by the red circle. Sequences of detected biotinylated peptides are given underneath.

3.3 Secretory pathway proteome

We next applied the proximity-dependent promiscuous labeling technique to the endoplasmic reticulum lumen. Because endoplasmic reticulum lumen is an oxidative environment, we have the option of using either HRP or APEX. We reasoned that HRP is a better choice for its higher activity. Furthermore, because HRP is inactive in the cytosol, this automatically eliminates any potential background labeling outside the secretory pathway due to (a small fraction of) mis-targeted enzyme.

Promiscuous protein labeling in the endoplasmic reticulum lumen with HRP

HRP was genetically targeted to the endoplasmic reticulum lumen of U2OS cells by attaching a *C*-terminal KDEL targeting sequence. Biotin-phenol labeling was performed in the same way as for the matrix proteome experiment. Super-resolution fluorescence imaging showed tightly correlated localizations between the enzyme and biotinylated targets (Figure 3-18).

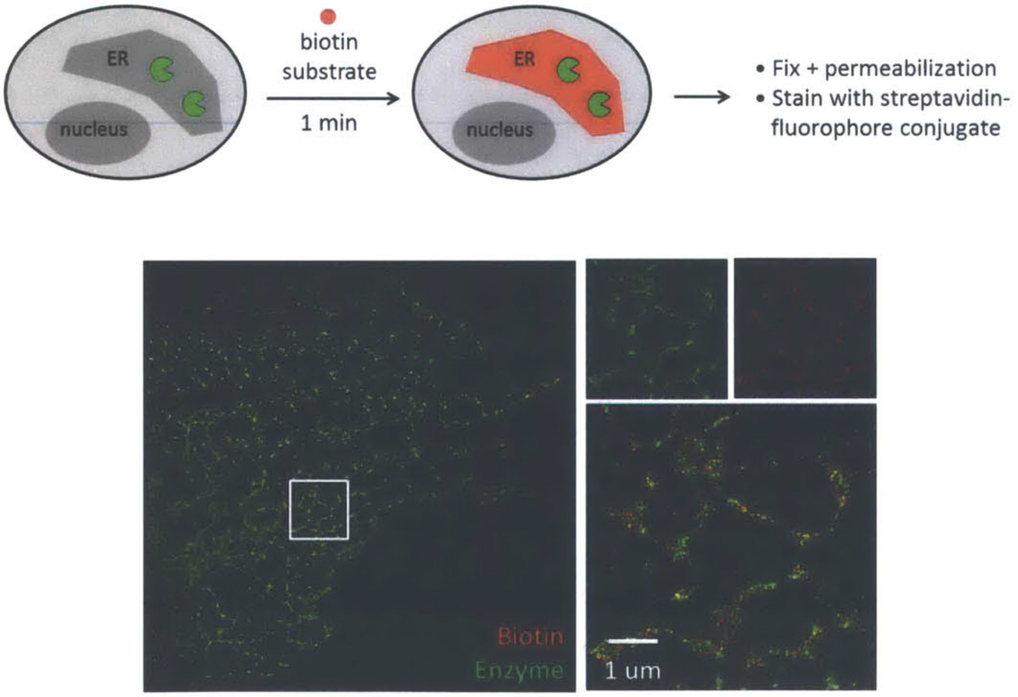


Figure 3-18. Promiscuous protein labeling in the endoplasmic reticulum lumen.
U2OS cells expressing endoplasmic reticulum-targeted HRP were labeled with biotin-phenol and H_2O_2. After cell fixation, permeabilization and blocking, biotinylated targets were labeled with streptavidin-dye conjugate and endoplasmic reticulum-HRP, with V5 epitope tag, was labeled with anti-V5 antibody. Top: labeling scheme. Bottom: STORM super-resolution image of HRP enzyme (in green) and biotinylated targets (in red).

Enrichment and subsequent LC-MS/MS analysis of the labeled proteome followed the same protocol as described in the previous section. Figure 3-19 shows the distribution of protein enrichment ratio H/L in log scale. The cutoff point was determined to be 1.13 in $\log_2$ unit. A total of 309 proteins were enriched above this cutoff point.

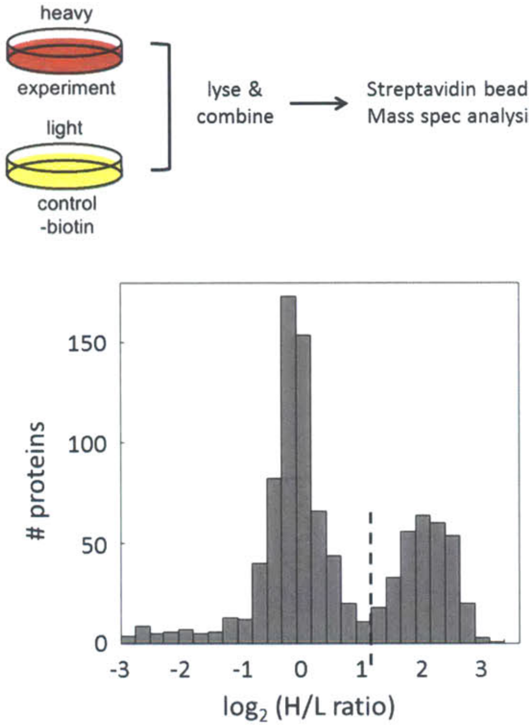


Figure 3-19. Distribution of protein enrichment ratio.
Top: SILAC labeling scheme. The light state is negative control omitting biotin-phenol reagent. Bottom: distribution of SILAC ratio (H/L) in $\log_2$ unit. The graph clearly shows a bimodal distribution, with an H/L cutoff point at around 1.13 in $\log_2$ unit.

Specificity analysis of our secretome

Interestingly, despite the fact that we targeted the HRP specifically to the endoplasmic reticulum lumen, and that our imaging characterization so far showed correct endoplasmic reticulum pattern for both the enzyme and the biotinylated targets, this proteomic dataset is not specific to endoplasmic reticulum proteins. Endoplasmic reticulum proteins account for only half of the enriched proteome. Proteins in the remaining half were known to be extracellular proteins, integral membrane proteins, or proteins localized to the Golgi apparatus. It appears that the entire secretory pathway was labeled! We reason that because the endoplasmic reticulum is the first stage in the secretory pathway, HRP in the endoplasmic reticulum lumen has the potential to label any passing by proteins in this pathway. We broke up this secretome into five localization categories (Figure 3-20).

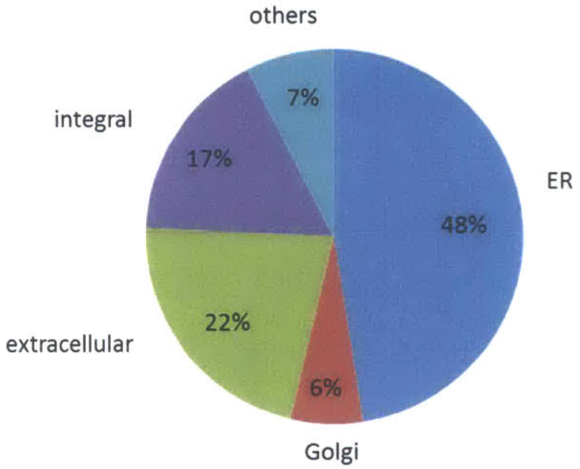


Figure 3-20. Distribution of protein localizations in our secretome dataset.
Nearly half of proteins are in the endoplasmic reticulum. This is followed by 22% of proteins secreted into the extracellular matrix, and 17% of proteins on the cell surface. Golgi proteins comprise another 6%. The remaining 7% have either unknown localization or located elsewhere.

Analysis of endoplasmic reticulum membrane proteins

As is shown in case of the mitochondrial matrix proteome, the biotin-phenoxyl radical labeling is limited to membrane bounded space. A protein is labeled only if a portion of it is exposed to the space where the enzyme is targeted. Analysis of the endoplasmic reticulum membrane proteins in our secretome further supported this observation. Figure 3-21 lists the top 20 most abundant endoplasmic reticulum membrane proteins (*24*), ranked by their abundance in U2OS cells (*15*). Fourteen proteins (70%) in this list are enriched in our dataset. Some of these proteins have known topology and the location of transmembrane domain. As expected, these proteins all have a large portion exposed to the endoplasmic reticulum lumen. For the remaining six proteins below the cutoff point of enrichment, four of them are single-transmembrane proteins with large cytoplasmic domains and virtually no lumen-exposed portion, and another one (DAD1) has three transmembrane domains with fewer than 10 lumen-exposed amino acid residues (*25*).

Rank	Protein name	Lumen exposure	Cytosol exposure
1	VAPA	–	+
2	Calnexin	+	+
3	Bap31	NA	NA
4	Ribophorin I	+	+
5	Climp63	+	+
6	Ribophorin II	+	+
7	Sec61 beta	–	+
8	Ribosome-binding protein 1 (p180)	–	+
9	SRP Receptor beta	NA	NA
10	VAPB	–	+
11	Dad1	–	+
12	Malectin	+	+
13	Surfeit locus protein 4	NA	NA
14	Sec61 alpha	NA	NA
15	TRAP delta	+	+
16	TRAP alpha	+	+
17	Magt1 (OST3/OST6 family)	NA	NA
18	STT3	NA	NA
19	Signal peptidase complex subunit 3	NA	NA
20	Transmembrane protein 97	NA	NA

Figure 3-21. Analysis of 20 abundant endoplasmic reticulum transmembrane proteins in U2OS cells. For each entry, a plus sign in lumen exposure or cytosol exposure means at least 10 amino acid residues are present in these spaces. Otherwise, it is indicated as a minus sign. "NA" means analysis not available.

The above analysis suggests that this method is potentially useful for mapping the topology of membrane proteins. However, to be a truly useful method, a larger number of biotinylated peptides need to be identified and then mapped to protein sequences. Currently, this is technically challenging because we detected only a small number of directly labeled peptides due to their high hydrophobicity which causes problems with poor elution from the LC column and poor detection sensitivity on the MS.

3.4 Conclusion

In summary, we have developed a novel method for mapping the proteomic composition of cellular organelles, using genetically-targetable peroxidase enzymes that catalyze the generation of short-lived, highly-reactive, promiscuous, and membrane impermeant radicals in living cells. With a temporal resolution of 1 minute, labeled proteins were harvested and identified by MS using well-established techniques. We note that in addition to its simplicity, the method has no noticeable toxicity and requires far less material than conventional organellar proteomics (1 – 4 mg of crude protein required prior to streptavidin enrichment, compared to entire animal organs (*26,27*)). Another feature of our methodology is that the same peroxidase, APEX, used for proteomic mapping in live cells can also be used for EM visualization (*7*), which is useful for verifying the localization of APEX in cells and for follow-up studies on proteomic hits.

Our initial demonstration on the human mitochondrial matrix proteome shows that specificity is exceptionally high, due to the fact that labeling is performed in living cells while membranes and other structures are still intact. Depth of coverage is also high for the majority of proteins – likely those that are sterically accessible to the phenoxyl radical. Demonstration of this labeling method to the secretory pathway also shows high spatial specificity.

We envision many future applications of this technology, spanning cell biology and neuroscience. Other organelles and sub-cellular compartments such as the IMS and synaptic cleft could be proteomically mapped. With improvements in our ability to recover directly biotinylated peptides, and perhaps an expansion of the amino acid reactivity of our radicals, we could also potentially use the method to perform large-scale mapping of the topology of membrane proteins or analyze the three-dimensional architecture of macromolecular complexes in living cells.

3.5 Experimental methods

Labeling of the mitochondrial matrix proteome

Culturing of HEK293T cells in SILAC media over 8 days was performed as previously described (*28*). On day 0, P8 HEK293T cells were split into two T25 flasks at 10–15% confluence. One flask was cultured in light SILAC media: a custom preparation from Caisson Laboratories containing L-arginine (Arg0), L-lysine (Lys0), 10% dialyzed fetal bovine serum (Sigma), penicillin, streptomycin, glutamine, and 4.5 g/L glucose in DMEM. The second flask was cultured in heavy SILAC media, with the same composition as above, except Arg0 and Lys0 were replaced by L-arginine-$^{13}C_6$, $^{15}N_4$ (Arg10) and L-lysine-$^{13}C_6$,$^{15}N_2$ (Lys8) (Sigma Isotech). Every two days, before cells reached confluence, the heavy and light SILAC cultures were split into fresh SILAC heavy and light media, respectively. After three passages in T25 flasks, on day 6, light and heavy SILAC cultures were expanded into T75 flasks at 20% confluence and cultured for two more days. We prepared two T75 flasks (H1, H2) for heavy SILAC cells and three T75 flasks (L1, L2, L3) for light SILAC cells.

On day 9, the HEK293T cells (H1, H2, L2, and L3) in T75 flasks (~7-8 million cells per flask) were transfected with mito-APEX (15 μg plasmid DNA) using 75 μL Lipofectamine 2000 in 30ml DMEM media. 4 hours after transfection, the media was replaced with fresh light or heavy SILAC media, and the cells were incubated overnight. On day 10, 18 hours after transfection, a biotin-phenol solution in 20 mL SILAC media was added to SILAC flasks H1, H2, L1, and L3 (final concentration = 500 μM). Cells were incubated for 30 min at 37 °C before addition of H_2O_2 (final concentration = 1 mM) for 1 minute at room temperature. After 1 minute labeling, samples were immediately quenched by aspirating off the probe/H_2O_2 solution and replacing with quencher solution (10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in 20 mL DPBS) for 1 min at room temperature. To ensure complete quenching and removal of residual biotin-phenol, cells were washed twice with quencher solution and twice with DPBS. For the L2 flask (negative control with biotin-phenol and H_2O_2 omitted), cells were washed three times with DPBS. After the final wash, 7 mL of quencher solution was added to each flask, and cells were collected by gentle pipetting followed by centrifugation at 200 rpm for 5 min at room temperature. Supernatant was discarded, and the cell pellet was stored at -80 °C.

On day 11, cell pellets were lysed as follows. Pellets were thawed on ice, then 300 μL RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1%

Triton X-100, 1x protease cocktail (Sigma Aldrich), 1 mM PMSF, 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox) was added to each pellet. Lysates were clarified by centrifugation at 13,000 rpm for 10 min at 4 °C. Protein concentration in the supernatant was measured using the Pierce 660 nm Protein Assay kit, with BSA as the reference standard.

The following optimizations were performed to determine the final live cell labeling conditions used. First, we wished to restrict the H_2O_2 exposure time to 1 min, to minimize toxicity to cells and maximize the temporal resolution of our method. Given this temporal restriction, we found that lowering the H_2O_2 concentration to <1 mM significantly reduced the biotinylation signal, assessed by imaging. We suspect that this is the case despite the low K_m of APX for H_2O_2 (~20 μM, (*29*)) and the high membrane permeability of H_2O_2, because of the endogenous catalase activity in mammalian cells. For biotin-phenol, we tested loading for 1 min, 3 min, 10 min, and 30 min and found much stronger labeling with the 30 min pre-incubation prior to H_2O_2 addition for 1 min; this may be because the biotin-phenol probe has limited membrane permeability. After labeling, we also found that it was critical to wash and lyse cells in buffer containing quenchers (Trolox, ascorbate, and sodium azide). Without addition of quenchers, we observed labeling of the outer mitochondrial membrane protein TOM20 by matrix-localized APEX, by both MS and Western blot (data not shown), suggesting that peroxidase-mediated biotinylation continues after dissolution of the mitochondrial membranes. Upon addition of quenchers to the wash and lysis buffers, labeling of TOM20 was eliminated.

Enrichment of biotinylated proteins, in-gel protein digestion and extraction

SILAC-labeled whole cell lysates prepared as described above were mixed in a 1:1 ratio (by protein concentration): H1 cells were mixed with L1 cells, and H2 cells were mixed with L2 cells. In addition, L3 cells were spiked in at 5% (by protein concentration) to each H1/L1 and H2/L2 pair to facilitate subsequent quantitation. Approximately 500 μL of streptavidin-coated magnetic beads (Pierce cat. no. 88817) were prepared by washing twice with RIPA lysis buffer (formulation given above). 4 mg of each cell lysate mixture (at 4-5 mg/mL protein concentration) was combined with the streptavidin bead slurry. The suspension was incubated at room temperature for 1 hour with gentle rotation. Thereafter, beads were washed with 2× 1 mL RIPA lysis buffer, once with 1 mL of 2M urea in 10 mM Tris-HCl pH 8.0, and again with 2× 1 mL RIPA lysis buffer. Finally, biotinylated proteins were eluted from the streptavidin beads by incubation with 60 μL of 1× NuPAGE LDS Sample Buffer (Invitrogen) supplemented with 20 mM DTT and 2 mM biotin at 95 °C for 5 min..

Proteins eluted from streptavidin beads as described above were loaded onto a NuPAGE Novex Bis-Tris 4-12% gel and separated for 1 hour using 130V. The gel was stained overnight with Coomassie G-250 (Invitrogen) and subsequently destained with water. Lanes for each replicate were manually cut into 16 gel bands. Gel bands were destained in 50:50 acetonitrile:100 mM ammonium bicarbonate pH 8.0 for several hours followed by dehydration with acetonitrile. Dehydrated gel bands were swelled with 100 μL of 10 mM DTT in 100 mM ammonium bicarbonate for one hour while shaking. The DTT solution was subsequently removed and 100 μL of 55 mM iodoacetamide in 100 mM ammonium bicarbonate was added to each gel band and allowed to react for 45 min in the dark. The iodoacetamide solution was removed and bands were dehydrated with acetonitrile. Approximately 10-50 μL of 20 ng/μL sequencing-grade trypsin (Promega) was added to each sample and digestion was completed overnight with shaking at room temperature.

After the overnight digestion, excess trypsin solution was removed from each sample and bands were incubated with 20 μL of extraction solution (60% acetonitrile/0.1% TFA). The extraction solution was collected after 10 min and this process was repeated 2 additional times. The final peptide extraction was completed by incubating gel bands with 100% acetonitrile for 1-2 min. Extracted peptides were dried to completeness in a vacuum concentrator, then reconstituted in 100 μL of 0.1% formic acid and loaded onto C18 StageTips (*30*), washed twice with 50 μL of 0.1% formic acid, and finally eluted with 50% acetonitrile/0.1% formic acid. StageTip eluates were dried in a vacuum concentrator.

Determination of the cut-off point for our matrix proteome

Protein level information was obtained from the MaxQuant Protein Groups table. Leading protein groups identified as reverse hits or contaminants were removed from the data set. The replicate 1 sample (based on the H1/L1 SILAC pair, where L1 lacks APEX) and the replicate 2 sample (based on the H2/L2 pair, where L2 lacks biotin-phenol and H_2O_2) were analyzed separately. For each dataset, only protein groups identified by three or more distinct peptides with quantified ratios were retained for further analysis. Protein SILAC ratios were normalized such that the median value is 1.

In order to determine the cut-off SILAC ratio for the dataset, proteins were categorized into two classes: (1) "mito" for those previously known to be localized to mitochondria; and (2) "false-positive" for those found in a "false positive" list of 2410 non-mitochondrial proteins compiled by Vamsi Mootha's lab (*4*). Histogram distributions of SILAC ratios in log2 units were generated for these two classes, with bin size of 0.1. For each bin, the probability of finding a protein in each class, namely P(SILAC ratio, if in this class), was calculated as the ratio of the count in each bin and the total count of the class. The false-positive rate (FPR, or the probability of finding a protein in the false-positive class versus the true-positive class) was defined as the ratio of two probabilities:

$$FPR = \frac{\mathrm{P(SILAC\ ratio, if\ false\ positive)}}{\mathrm{P(SILAC\ ratio, if\ mito)}}$$

The cut-off point for each dataset is defined as the SILAC ratio at which the FPR reaches 0.1 (i.e. it is 10 times more likely to have a true-positive than a false-positive hit). The cut-off points (in log2 units) for replicates 1 and 2 were 1.05 and 1.25, respectively, which gives 527 enriched protein groups for replicate 1 and 579 enriched protein groups for replicate 2. The overlap of these two enriched datasets contains 495 protein groups, which we take to be our mitochondrial matrix proteome. The numbers of detected peptides, detected proteins, and enriched proteins for each replicate, as well as the overlap between the two replicates results are summarized in Table 3-1.

Table 3-1. Summary of MS analysis.

	Replicate 1	Replicate2
# detected peptides	50032	54744
overlap	39947	
# detected proteins	3430	3599
overlap	3147	
# enriched proteins	527	579
overlap	495	

Structural mapping of biotinylated peptides

Biotinylated peptide information was obtained from the MaxQuant Evidence table. All peptides identified as reverse or contaminant hits were excluded, and only biotinylated peptides with score >50 were retained for this analysis. For each modified peptide sequence associated with multiple MS/MS scans, the highest score for these scans was recorded.

For analysis of solvent accessible tyrosine residues, protein structure models were obtained through searching UniProt accession IDs in the Swiss Model Repository, and only models with sequence homology >85% were included in the analysis. Solvent accessible surface areas for individual tyrosine residues were calculated in PyMOL. Membrane topology information for each entry was obtained from the UniProt database.

Reference

1 Harris, D. A. & Das, A. M. Control of mitochondrial ATP synthesis in the heart. *Biochem J* **280 (Pt 3)**, 561-573 (1991).

2 Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309-1312 (1998).

3 Picard, M. *et al.* Mitochondrial structure and function are disrupted by standard isolation methods. *PLoS One* **6**, e18317 (2011).

4 Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123 (2008).

5 Simons, K. & Ikonen, E. Cell biology - How cells handle cholesterol. *Science* **290**, 1721-1726 (2000).

6 Schroder, M. & Kaufman, R. J. The mammalian unfolded protein response. *Annu Rev Biochem* **74**, 739-789 (2005).

7 Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically-encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143-1148 (2012).

8 Niers, J. M., Chen, J. W., Weissleder, R. & Tannous, B. A. Enhanced in vivo imaging of metabolically biotinylated cell surface reporters. *Anal Chem* **83**, 994-999 (2011).

9 Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

10 http://www.geneontology.org/.

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

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

13 Gebert, N., Ryan, M. T., Pfanner, N., Wiedemann, N. & Stojanovski, D. Mitochondrial protein import machineries and lipids: a functional connection. *Biochim Biophys Acta* **1808**, 1002-1011 (2011).

14 Bordo, D. & Argos, P. Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J Mol Biol* **217**, 721-729 (1991).

15 Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549 (2011).

16 Grandchamp, B., Phung, N. & Nordmann, Y. The mitochondrial localization of coproporphyrinogen III oxidase. *Biochem J* **176**, 97-102 (1978).

17 Ferreira, G. C., Andrew, T. L., Karr, S. W. & Dailey, H. A. Organization of the terminal two enzymes of the heme biosynthetic pathway. Orientation of protoporphyrinogen oxidase and evidence for a membrane complex. *The Journal of biological chemistry* **263**, 3835-3839 (1988).

18 Nilsson, R. *et al.* Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab* **10**, 119-130 (2009).

19 Hamza, I. & Dailey, H. A. One ring to rule them all: trafficking of heme and heme synthesis intermediates in the metazoans. *Biochim Biophys Acta* **1823**, 1617-1632 (2012).

20 Koch, M. *et al.* Crystal structure of protoporphyrinogen IX oxidase: a key enzyme in haem and chlorophyll biosynthesis. *EMBO J* **23**, 1720-1728 (2004).

21 Morgan, R. R., Errington, R. & Elder, G. H. Identification of sequences required for the import of human protoporphyrinogen oxidase to mitochondria. *Biochem J* **377**, 281-287 (2004).

22 Dailey, T. A., Woodruff, J. H. & Dailey, H. A. Examination of mitochondrial protein targeting of haem synthetic enzymes: in vivo identification of three functional haem-responsive motifs in 5-aminolaevulinate synthase. *Biochem J* **386**, 381-386 (2005).

23 Brown, T. A. *et al.* Superresolution fluorescence imaging of mitochondrial nucleoids reveals their spatial range, limits, and membrane interaction. *Mol Cell Biol* **31**, 4994-5010 (2011).

24 Shibata, Y. *et al.* Mechanisms determining the morphology of the peripheral ER. *Cell* **143**, 774-788 (2010).

25 Kelleher, D. J. & Gilmore, R. DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase. *Proc Natl Acad Sci U S A* **94**, 4994-4999 (1997).

26 Huber, L. A., Pfaller, K. & Vietor, I. Organelle proteomics: implications for subcellular fractionation in proteomics. *Circ Res* **92**, 962-968 (2003).

27 Schnaitman, C. & Greenawalt, J. W. Enzymatic properties of the inner and outer membranes of rat liver mitochondria. *The Journal of cell biology* **38**, 158-175 (1968).

28 Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc* **1**, 2650-2660 (2006).

29 Dalton, D. A., Diaz del Castillo, L., Kahn, M. L., Joyner, S. L. & Chatfield, J. M. Heterologous expression and characterization of soybean cytosolic ascorbate peroxidase. *Arch Biochem Biophys* **328**, 1-8 (1996).

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

Part B

Fluorescence imaging analysis of protein localization and trafficking

Chapter 4 Introduction to fluorescence imaging analysis of protein trafficking

4.1 Introduction

The protein composition of cellular compartments is maintained by the continual flux of entering and exiting proteins. The resulting protein trafficking among various compartments is crucial for many aspects of cell physiology. Protein trafficking into the nucleus allows transcriptional control of gene expression, and proteins destined for secretion are first synthesized in the endoplasmic reticulum, and then modified in the Golgi, sorted into secretory vesicles, and eventually released into the environment. One prominent example of protein trafficking is clathrin-mediated endocytosis of cell surface receptors (*1*). While internalization of nutrient receptors is crucial for taking up ligand from the environment, internalization of signaling receptors provides a means of signal attenuation and desensitization, and subsequent targeting from endosome to the lysosome achieves protein degradation.

Because of its high spatial resolution and high detection sensitivity, fluorescence microscopy has become an indispensable tool for studying protein localization and trafficking. This has been aided by the development various protein labeling techniques during the past decades. This chapter introduces these chemical and physical tools.

4.2 Existing methods for fluorescent labeling of specific proteins

Specific protein labeling methods can be broadly categorized into two classes: the "tag-less" approaches versus those requiring fusion to protein or peptide tags. The "tag-less" approach is exemplified by the use of antibodies to target endogenous proteins. Because the large size of antibodies precludes their entry across the cell membrane, for live cell applications, this method has been limited to cell surface proteins. Even so, their large size often interferes with ligand binding and/or receptor trafficking, not to mention their often weak targeting affinity and tendency to crosslinking underlying receptors (monovalent Fab fragments can be used to avoid crosslinking problem). In the special case of cell surface receptor proteins, specific labeling has been traditionally achieved by using a fluorescently labeled ligand (*2,3*). One caveat with this method is the activation of underlying receptor, and therefore it is not possible to study the receptor in its basal, ligand-free state. Although the number of antibodies against endogenous protein targets has been steadily increasing over the past decades, not every protein has an antibody available.

Alternatively, a peptide or protein tag can be fused to the protein of interest in order to facilitate specific labeling. The most prominent example of protein tag is the development of fluorescent proteins (*4*). The ability to genetically encode these fluorescent labels ensures perfect targeting specificity and has revolutionized the field of protein labeling and imaging. During the past decades, fluorescent proteins have been engineered to span the palette of the entire visible spectrum (*5*), including the near infrared range (*6*), or to possess unique photophysical properties that afford advanced imaging modalities such as super-resolution microscopy (*7*). One limitation of fluorescent proteins is their lack of brightness and photostability, especially when comparing to commonly used organic fluorophores. To overcome this problem, many novel protein tags have been developed for site-specific protein labeling with small, organic fluorophores. These include HaloTag (*8*), SNAP/CLIP tag (*9*), and DHFR tag (*10*).

HaloTag was engineered from bacterial dehalogenase whose natural function is to catalyze the hydrolysis of chloroalkanes or bromoalkanes into alcohols, thereby protecting the bacteria from toxicity of these xenobiotics. In this enzymatic catalysis, a covalent adduct of alkyl chain with aspartic acid side chain is transiently formed in the active site as an intermediate. The engineered version of HaloTag is capable of trapping this intermediate and utilizes it as a means of covalently modifying the protein. Similarly, SNAP tag and CLIP tag were both engineered from human O^6-alkylguanine-DNA alkyltransferase. The natural function of the enzyme is to transfer the alkyl group from O^6-alkyl-guanine onto itself, thus protecting DNA from damaging by alkylating agents. The result is covalent modification of the enzyme at its active site cysteine residue. Protein engineering alters the substrate specificity of the enzyme: while SNAP tag recognizes O^6-benzylguanine derivatives, CLIP tag recognizes O^6-propargylguanine derivatives (*9*). These techniques are attractive to biologists due to the covalent nature of labeling and their fast kinetics. The second-order rate constants of SNAP/CLIP are on the order of $10^3 - 10^4$ M^{-1} s^{-1}, and the rate constant of HaloTag approaches 10^6 M^{-1} s^{-1}, which is close to diffusion-controlled rate limit. Unlike these enzyme variants that feature covalent modification, the DHFR tag utilizes strong, yet non-covalent interactions between the protein scaffold and its ligand. Trimethoprim was designed as a competitive enzyme inhibitor, with a binding affinity of 10^{-8} M towards dihydrofolate reductase (DHFR) (*10*).

The major drawback with all the labeling methods mentioned above is the large tag size, which typically exceeds 100 amino acids. This can interfere with protein function, block interaction with other proteins, or alter trafficking properties. The use of small peptide tag is therefore advantageous over these bulky protein fusions. The introduction of epitope tags suits for this purpose and can be viewed as an extension of antibody-based approach (*11*). It retains the benefit of high specificity of antibody recognition with the added advantage of being able to target any proteins of interest regardless of their immunogenicity. However, it also shares all the limitations with the antibody. Site-specific protein labeling techniques aim to directly modify the protein of interest with a probe, thus bypassing the requirement of antibodies. A number of methods have been developed, which include poly His (12-14), tetra-cysteine tag for FlAsH/ReAsH (15), SorTag (16), aldehyde tag (*17*), AcpS/Sfp (*18,19*), biotin acceptor peptide (*20*), and lipoic acid acceptor peptide (*21-26*).

The poly His tag utilizes the coordination complex formed between tandem imidazole groups on the histidine side chain, a probe containing nitrilotriacetic acid moiety, and a transition metal ion such as nickel (II) or zinc (II). Originally developed for purifying recombinant proteins, this technique has been adapted for protein labeling in living cells by using (*12,13*). FlAsH/ReAsH techniques developed by Tsien *et al.* (*15*) is another example of utilizing strong interaction between a peptide motif and the probe. The tetra-cysteine (-CCxxCC-) structure in the peptide tag tightly associates with the biarsenical moiety of small molecule probes, with dissociation constant as low as 10^{-11} M.

The other peptide tags can all be recognized for post-translational modification by respective enzymes. The SorTag is a hexapeptide motif (LPETGG) at the C-terminus of protein. It is recognized by a prokaryotic transpeptidase, sortase, for covalent conjugation with a probe-containing pentaglycine peptide. The aldehyde tag is a pentapeptide sequence CxPxR, in which the cysteine residue can be selectively oxidized into formylglycine by formylglycine generating enzyme. The aldehyde group in formylglycine can then be targeted with aminooxy- or hydrazide-functionalized probes (*27*). In the case of AcpS/Sfp, phosphopantetheinyl transferase enzyme recognizes these sequences and transfer modified coenzyme A onto the serine residue of the peptide. *E. coli* biotin ligase, BirA, specifically recognizes a 15-amino acid peptide sequence (acceptor peptide) and catalyzes the formation of amide bond between biotin carboxyl group and the lysine side chain amino group. Likewise, the lipoate protein ligase (LplA), also from *E. coli*, recognizes a 13-amino acid peptide sequence and modifies its lysine residue with lipoic acid. Recently, LplA has been engineered to recognize a variety of small molecule probes, including coumarin derivatives (*21-23*), aryl azide (*24*), alkyl azide and alkyne (*25*), and aryl-aldehyde or hydrazine (*26*). These LplA-based labeling methods are collectively termed "probe incorporation mediated by enzyme", or PRIME.

Together, these various protein labeling techniques are valuable for studying protein localization and trafficking in the context of living cells.

4.3 Fluorescence microscopy for studying protein localization and trafficking

The high spatial resolution and compatibility with live cell experiment of fluorescence microscopy make it a powerful tool for examining the localization and trafficking of proteins within the living cells. The development of digital recording with sensitive camera greatly facilitates quantitative analysis of images, either individually or in time-series. During the past decades, various advanced imaging methods have been developed to improve the optical resolution. Confocal and multi photon microscopy achieve optical sectioning either by rejecting out-of-focus light (*28*) or by limiting the excitation volume via non-linear optical absorption (*29*); total internal reflection fluorescence (TIRF) microscopy utilizes evanescent field illumination to limit the excitation field to a thin layer near the glass substrate, typically less than 100 nm thick (*30*); super-resolution imaging techniques, such as stimulated emission depletion (STED), photo-activatable localization microscopy (PALM), and stochastic optical reconstruction (STORM), have achieved spatial resolution down to 25 nm and temporal resolution as fast as 0.5 second in live cells (*31-33*).

Additionally, various microscopic methods have been developed to monitor protein trafficking process in live cells. Fluorescence recovery after photobleaching (FRAP) allows examining the diffusion of biomolecules (proteins, lipids, etc.) inside the cell in real time. In this method, biomolecules of interest are fluorescently labeled in live cells, and an intense laser beam is directed to the sample to transiently photobleach fluorophores in a selected region, such as the nucleus. This is immediately followed by sequentially imaging the cell to watch how fluorescently labeled molecules redistribute into the photobleached region. From this time-series analysis, the diffusion coefficient can be deduced. A variation of this technique is fluorescence loss in photobleaching (FLIP), in which the photobleaching laser beam is kept on. Due to diffusion of fluorophores, the fluorescence intensity outside the photobleaching region will decrease over the time, which provides information about diffusion coefficient. An alternative approach is to use photo-activatable protein fusions and selectively photo-activate a sub-cellular region, such as the Golgi apparatus, and study how this cohort of fluorescent protein traffics inside the cell (*7*).

Single-particle tracking with quantum dots is by far the most powerful method for studying protein trafficking on the cell surface. As in the case of super-resolution imaging techniques, fitting the point spread function of single-particle fluorophore gives localization accuracy of tens of nanometers. Due to the brightness and photostability of the fluorescent probe, the trajectories of individually labeled molecules are recorded in real time, with a temporal resolution on the order of milliseconds. Compared to ensemble fluorescence measurement of molecular diffusion, single-particle tracking experiment has the potential of revealing the heterogeneity within the sample. For example, trajectory analysis in single-particle tracking is able to distinguish between the mobile and immobile populations within the same cell, whereas ensemble analysis of the same sample only shows the average diffusion properties. In addition, by analyzing the trajectory in different cellular region, single-particle tracking could also provide information about the spatial heterogeneity of the cell surface, such as the presence of diffusion barriers.

The first single-particle tracking experiment dates back to 1982, when Webb *et al.* used fluorescently labeled low density lipoprotein (LDL) particle to observe the dynamics of its receptor (LDLR) on the surface of cultured fibroblast cells (*34*). LDL is suitable for such studies because its hydrophobic core of cholesterol esters can be easily loaded with multiple organic dye molecules such as the carbocyanine lipid analog DiI. This affords the brightness necessary for identifying individual LDL particles under the microscope.

Since then, various nanoparticles have been utilized for single-particle tracking. Colloidal gold particles with diameters ranging between 15 – 40 nm are strong light scatter sources, and thus have been employed for single-particle tracking in bright-field microscopy. Fluorescent labels are favored because of the dark background in imaging. For example, fluorescent microspheres with diameters ranging between 30 – 100 nm have been used to track membrane anchored proteins (*35,36*). Phycobiliproteins, such as commercially available R-phycoerythrin (R-PE), contain more than 30 fluorophores per protein molecule, and therefore are extremely bright at the single molecule level (*37*). Similar to the LDL particle, certain virus particles can also be labeled with lipophilic fluorophores. This has allowed direct visualization of viral entry into mammalian cells (*38*).

Another important class of nanoparticles for single-particle tracking is the quantum dots. These semiconductor nanoparticles feature core-shell binary structure, with their band gap, and thus emission spectra, tunable by the particle size. Quantum dots have broad excitation spectrum, large extinction coefficients (10^5 – 10^6 M^{-1} cm^{-1} at the first exitonic absorption peak) and high fluorescence quantum yield (as high as 0.8), which make them ideal for single-particle tracking probes (*39*). In addition, quantum dots are highly photostable, being able to withstand continual illumination for minutes without photobleaching. Due to occasional photo-ionization, these particles do exhibit blinking during single-particle tracking. On the one hand, this is a drawback that complicates data analysis; on the other hand, it can be used as a means to validate the presence of single quantum dot, instead of a cluster of particles. The narrow emission bandwidth (full-width at half maxima 30 – 90 nm) of QDs also makes them useful for multiplex imaging (*39*).

4.4 Conclusion

Studying protein localization and trafficking requires the coordinated efforts of chemical biology and physical instrumentation. Here, we have reviewed various techniques for fluorescently labeling proteins of interest and microscopy methods for observing their localization as a function of time with high spatial and temporal resolutions. Each of these labeling methods comes with its own strength and weakness. One should be judicious when choosing among these tools for addressing a specific biological question. By carefully combining these techniques, it would be possible to achieve multiplex labeling and imaging within the same sample, which makes it possible to study the spatial and temporal correlation between several proteins of interests.

The biotin ligase/acceptor peptide approach features small peptide tag size, stable protein-fluorophore linkage, rapid labeling kinetics and ease of operation. In addition, it is also capable of selectively labeling the cell surface. This added advantage makes it especially attractive for studying cell surface receptor proteins. Therefore, in the next two chapters, we focus on the application of this labeling method in both ensemble fluorescence imaging and single-particle tracking experiments.

References

1 Kirchhausen, T. Clathrin. *Annu Rev Biochem* **69**, 699-727 (2000).

2 Reynolds, G. D. & St Clair, R. W. A comparative microscopic and biochemical study of the uptake of fluorescent and 125I-labeled lipoproteins by skin fibroblasts, smooth muscle cells, and peritoneal macrophages in culture. *Am J Pathol* **121**, 200-211 (1985).

3 Ghosh, R. N. & Webb, W. W. Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules. *Biophys J* **66**, 1301-1318 (1994).

4 Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224 (2006).

5 Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Methods* **2**, 905-909 (2005).

6 Filonov, G. S. *et al.* Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat Biotechnol* **29**, 757-761 (2011).

7 Lippincott-Schwartz, J. & Patterson, G. H. Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol* **19**, 555-565 (2009).

8 Los, G. V. *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373-382 (2008).

9 Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* **15**, 128-136 (2008).

10 Calloway, N. T. *et al.* Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *Chembiochem* **8**, 767-774 (2007).

11 Jarvik, J. W. & Telmer, C. A. Epitope tagging. *Annu Rev Genet* **32**, 601-618 (1998).

12 Hauser, C. T. & Tsien, R. Y. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc Natl Acad Sci U S A* **104**, 3693-3697 (2007).

13 Peneva, K. *et al.* Exploiting the nitrilotriacetic acid moiety for biolabeling with ultrastable perylene dyes. *Journal of the American Chemical Society* **130**, 5398-+ (2008).

14 Roullier, V. *et al.* High-affinity labeling and tracking of individual histidine-tagged proteins in live cells using Ni2+ tris-nitrilotriacetic acid quantum dot conjugates. *Nano Lett* **9**, 1228-1234 (2009).

15 Martin, B. R., Giepmans, B. N., Adams, S. R. & Tsien, R. Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* **23**, 1308-1314 (2005).

16 Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **3**, 707-708 (2007).

17 Wu, P. *et al.* Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc Natl Acad Sci U S A* **106**, 3000-3005 (2009).

18 George, N., Pick, H., Vogel, H., Johnsson, N. & Johnsson, K. Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* **126**, 8896-8897 (2004).

19 Sunbul, M., Yen, M., Zou, Y. & Yin, J. Enzyme catalyzed site-specific protein labeling and cell imaging with quantum dots. *Chem Commun (Camb)*, 5927-5929 (2008).

20 Howarth, M. & Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* **3**, 534-545 (2008).

21 Uttamapinant, C. *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* **107**, 10914-10919 (2010).

22 Cohen, J. D., Thompson, S. & Ting, A. Y. Structure-guided engineering of a Pacific Blue fluorophore ligase for specific protein imaging in living cells. *Biochemistry* **50**, 8221-8225 (2011).

23 Jin, X., Uttamapinant, C. & Ting, A. Y. Synthesis of 7-aminocoumarin by Buchwald-Hartwig cross coupling for specific protein labeling in living cells. *Chembiochem* **12**, 65-70 (2011).

24 Baruah, H., Puthenveetil, S., Choi, Y. A., Shah, S. & Ting, A. Y. An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* **47**, 7018-7021 (2008).

25 Fernandez-Suarez, M. *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **25**, 1483-1487 (2007).

26 Cohen, J. D., Zou, P. & Ting, A. Y. Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation. *Chembiochem* **13**, 888-894 (2012).

27 Rabuka, D., Rush, J. S., deHart, G. W., Wu, P. & Bertozzi, C. R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat Protoc* **7**, 1052-1067 (2012).

28 Conchello, J. A. & Lichtman, J. W. Optical sectioning microscopy. *Nat Methods* **2**, 920-931 (2005).

29 Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73-76 (1990).

30 Axelrod, D. Total internal reflection fluorescence microscopy in cell biology. *Traffic* **2**, 764-774 (2001).

31 Jones, S. A., Shim, S. H., He, J. & Zhuang, X. Fast, three-dimensional super-resolution imaging of live cells. *Nat Methods* **8**, 499-508 (2011).

32 Chojnacki, J. *et al.* Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* **338**, 524-528 (2012).

33 Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642-1645 (2006).

34 Barak, L. S. & Webb, W. W. Diffusion of Low-Density Lipoprotein-Receptor Complex on Human-Fibroblasts. *Journal of Cell Biology* **95**, 846-852 (1982).

35 Hicks, B. W. & Angelides, K. J. Tracking movements of lipids and Thy1 molecules in the plasmalemma of living fibroblasts by fluorescence video microscopy with nanometer scale precision. *J Membr Biol* **144**, 231-244 (1995).

36 Fein, M. *et al.* Lateral mobility of lipid analogues and GPI-anchored proteins in supported bilayers determined by fluorescent bead tracking. *J Membr Biol* **135**, 83-92 (1993).

37 Wilson, K. M., Morrison, I. E. G., Smith, P. R., Fernandez, N. & Cherry, R. J. Single particle tracking of cell-surface HLA-DR molecules using R-phycoerythrin labeled monoclonal antibodies and fluorescence digital imaging. *J Cell Sci* **109**, 2101-2109 (1996).

38 Brandenburg, B. & Zhuang, X. Virus trafficking - learning from single-virus tracking. *Nat Rev Microbiol* **5**, 197-208 (2007).

39 Resch-Genger, U., Grabolle, M., Cavaliere-Jaricot, S., Nitschke, R. & Nann, T. Quantum dots versus organic dyes as fluorescent labels. *Nature Methods* **5**, 763-775 (2008).

Chapter 5 Ensemble fluorescence imaging analysis of receptor trafficking

Except for Section 5.4, work described in this chapter has been published in: P. Zou and A. Y. Ting. Imaging LDL receptor oligomerization during endocytosis using a co-internalization assay. *ACS Chemical Biology* 2011, 6, 308-313. This chapter is adapted from this publication.

5.1 Introduction

In this chapter, we examine the clathrin-mediated endocytosis of the LDL receptor using a novel fluorescence imaging assay based on biotin ligase/acceptor peptide labeling. Internalization is often a highly regulated process, and there have been many well-studied cases where the oligomerization state of a receptor is closely correlated with its internalization activity (reviewed in ref. (*1*)). Oligomerization could promote internalization either by directly triggering cross-reactivity between monomers, as is in the case of cross phosphorylation of the epidermal growth factor (EGF) receptor, or indirectly by increasing the local concentration of internalization motif which recruits the formation of the endocytic machinery to these loci at the plasma membrane (*2*). For this reason, equipped with this live cell imaging assay, we seek to understand whether receptor oligomerization plays a role in the endocytic process.

Clathrin-mediated endocytosis

Clathrin-mediated endocytosis is one of the most well studied endocytic internalization mechanism. In this process, ligand-bound receptors are targeted to clathrin-coated pits, which are cage-like structure formed by clathrin polymer near the cell surface. Then the cell membrane around the pits invaginates into the cell and eventually pinches off to trap the ligand-receptor complex into a clathrin-coated vesicle. As the vesicle moves away from the plasma membrane, its clathrin shell disassembles and recycles back to the cell surface, leaving only a membrane-bound vesicle. After entering the cell, these uncoated vesicles then fuse with each other to form endosomes, where a sorting mechanism takes place to separate ligand from its receptor. Certain type of receptors, such as the low-density lipoprotein (LDL) receptor and transferrin receptor, are recycled back to the plasma membrane, while other receptors and most ligands are retained in the endosome and eventually targeted to the lysosome for degradation (*3*). The whole process is extremely fast and efficient. In the case of the LDL receptor, it is constitutively internalized and recycled once every ~12 min, completing a total of ~150 round trips in its life span of ~30 hours (*4*).

The LDL receptor is an integral membrane protein with a large extracellular domain (residues 1 – 767) that binds to the LDL particle, a single transmembrane domain (TMD, residues 768 – 789), and a short cytoplasmic tail (residues 790 – 839) bearing an NPxY motif (^{804}NPVY) that is recognized for clathrin-mediated endocytosis (*5*,*6*). It binds to LDL at neutral pH but releases it at endosomal pH (~ 5) (*7*). Mutations in the NPxY motif, especially at the Tyr807, often lead to a significant decrease in receptor internalization, as revealed in patients afflicted with familial hypercholesterolemia (FH683 with W792Stop mutation and J.D. with Y807C mutation) (*8-10*).

Existing methods for measuring receptor oligomerization state

Many cell surface receptors form dimers or higher-order oligomers. Oligomerization state often plays an important role in regulating receptor function, such as ion conductance (*11*), enzymatic activity (*12*), and effector recruitment during signal transduction (*13*). Receptor oligomerization can be dynamically regulated in both temporal and spatial dimensions. In time, oligomerization state may be altered by molecular events such as ligand binding (*12,13*). In space, the extent of oligomerization may vary in different cellular compartments – such as at the cell surface, in the endoplasmic reticulum, or in endocytic vesicles (*14*). Receptor dimerization and internalization are usually investigated separately using different assays. A new method for analyzing the oligomerization state during the process of internalization would benefit mechanistic studies of receptor function.

Most methods used to characterize receptor oligomerization require removing the receptor from the cellular context. For example, oligomerization can be probed via co-immunoprecipitation from cell lysate (*15*), or via direct measurements of receptor size using gel-filtration chromatography or electrophoresis (*16*). However, these methods do not replicate physiological conditions, and the requirement for receptor solubilization raises the possibility of missing transiently formed oligomers or aggregating receptors due to incomplete solubilization (*16*). Chemical crosslinking can also be used to probe receptors in living cells, but it has limited efficiency and sensitivity (*17*); in addition, its non-specific nature could lead to false positives by crosslinking via nearby proteins. Recently, a strategy called RASSL (receptor activated solely by a synthetic ligand) (*18*) has been applied to study the co-internalization of mutant and wild-type beta-2-adrenergic receptor (*19*). However, this requires chemical synthesis of selective ligands and avoids receptor activation.

Compared with traditional biochemical assays, light microscopy offers high spatial and temporal resolution. Recently, numerous fluorescence imaging methods have been developed to probe receptor oligomerization in living cells. These include single-molecule subunit counting

(*20*), bimolecular fluorescence complementation (*15*), fluorescence resonance energy transfer (*16*), and fluorescence recovery after photobleaching (*21*). Limited by the fluorescence labeling methods used in these assays, many of these approaches do not distinguish between specific cohorts of receptors. For these reasons, we sought to develop a new methodology, based in single living cells, to selectively probe the oligomerization state of the receptor subpopulation that is undergoing endocytic internalization.

5.2 Fluorescence imaging assay for receptor internalization

Cell surface fluorescence labeling of LDL receptor

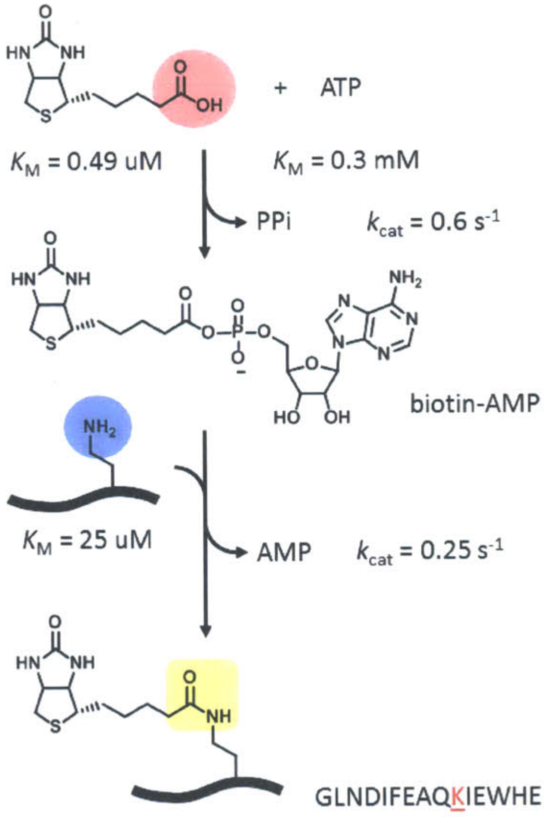


Figure 5-1. BirA-mediated site-specific biotinylation of AP peptide.
In the first step, BirA catalyzes the formation of biotin-adenylate (biotin-AMP) ester from biotin and ATP, where inorganic phosphate (PPi) is released as a byproduct. The resulting adenylate ester is tightly held at the substrate binding site until binding of the acceptor peptide (AP) substrate, and an amide bond is formed with the concomitant release of adenosine monophosphate (AMP). The central lysine residue is highlighted in red.

We selected the low density lipoprotein receptor (LDLR) to demonstrate our methodology. As described in the previous section, LDL receptor is a constitutively internalized and recycled receptor, and it mediates the clearance of LDL cholesterol from plasma. To develop the internalization assay, we need to specifically label the cell surface population of LDL receptor. The N-terminus of mature LDLR was fused to a 15-amino-acid "acceptor peptide" (AP) (fusion after the leader sequence), which is specifically biotinylated by an *E. coli* biotin ligase BirA (Figure 5-1). To selectively label the cell surface population of LDLR with fluorescent probe, BirA is expressed in the endoplasmic reticulum, so that AP-tagged receptor is efficiently biotinylated as it travels through the endoplasmic reticulum in the secretory pathway (Figure 5-2). Cell surface pool of biotinylated AP-LDLR is then labeled with AlexaFluor568-conjugated monovalent streptavidin (mSA-Alexa568), which is a 54 kDa protein. This labeling is fast (2 minutes for the mSA step), highly stable ($K_d \sim 5\times10^{-14}$ M between biotin and mSA), and does not artificially induce receptor crosslinking (*22*). Figure 5-3 shows that BirA/mSA labeling of AP-LDLR is site-specific and requires co-expression of endoplasmic reticulum-targeted BirA.

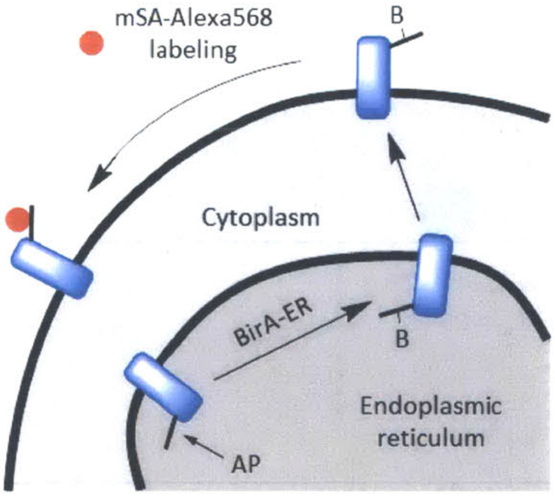


Figure 5-2. Scheme for selective labeling of cell surface AP-LDLR.
Biotin ligase is targeted to the endoplasmic reticulum (BirA-ER), where it site-specifically conjugates biotin onto AP-LDLR. As the biotinylated receptor traffics to the cell surface, it is tagged with fluorophore-conjugated monovalent streptavidin (mSA) (14). Because mSA is membrane-impermeant, only the surface population of the receptor is fluorescently labeled. B = biotin.

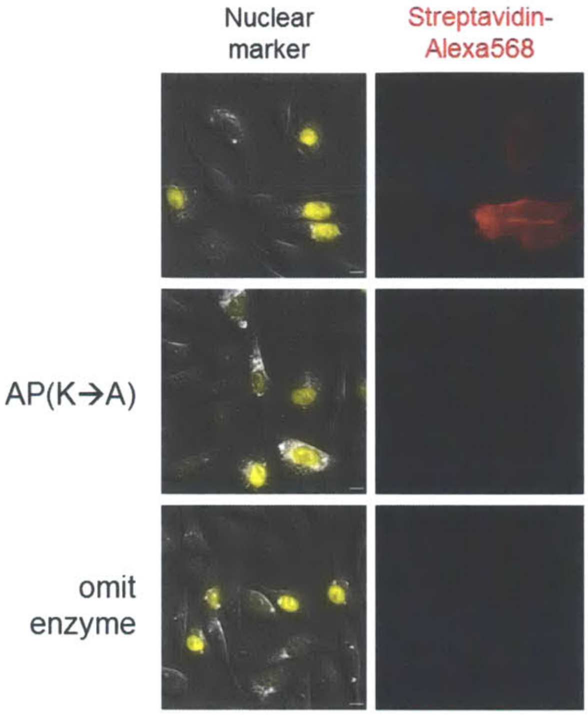


Figure 5-3. Specific labeling of AP-LDLR.
T-REx CHO A7 cells expressing AP-LDLR, BirA, and a nuclear YFP transfection marker (in a separate plasmid) were incubated with 10 μM biotin overnight. Biotinylated AP-LDLR on the cell surface was labeled with mSA-Alexa568 at 4 °C for 10 min. Negative controls are shown with AP-LDLR replaced by a K→A point mutant in AP, and with BirA omitted (bottom). Scale bars = 10 μm.

Cell surface fluorescence quenching

We wish to quantify the fluorescence intensity that is contributed by the cell surface receptor only. During endocytic internalization, this intensity is expected to decrease over the time. By observing the same field of view and taking the ratio of remaining intensity over the total intensity at time zero, we are able to calculate the level of internalization for each cell. Image-based segmentation is potentially useful for discerning the membrane fluorescence intensity versus intracellular signal, but it is not clean, especially when internalized vesicles are close to the plasma membrane.

An alternative approach is to selectively quench the fluorescence of labeled receptors at the cell surface. An ideal quencher must be membrane-impermeant, and should have broad absorbance, high extinction coefficient, and low fluorescence quantum yield. We compared Trypan blue (*23*), QSY dyes (Invitrogen), and BHQ dyes (Biosearch Technology) for this purpose. QSY21 was the most promising for quenching mSA-Alexa568 fluorescence in solution, but it had limited water solubility and gave incomplete quenching at the cell surface. We therefore prepared a conjugate of QSY21 and spermine, to increase polarity and electrostatic attraction to the phospholipid bilayer. *In vitro* measurements showed that 40 μM QSY21-spermine quenched 80 nM mSA-Alexa568 fluorescence by 94% (Figure 5-4).

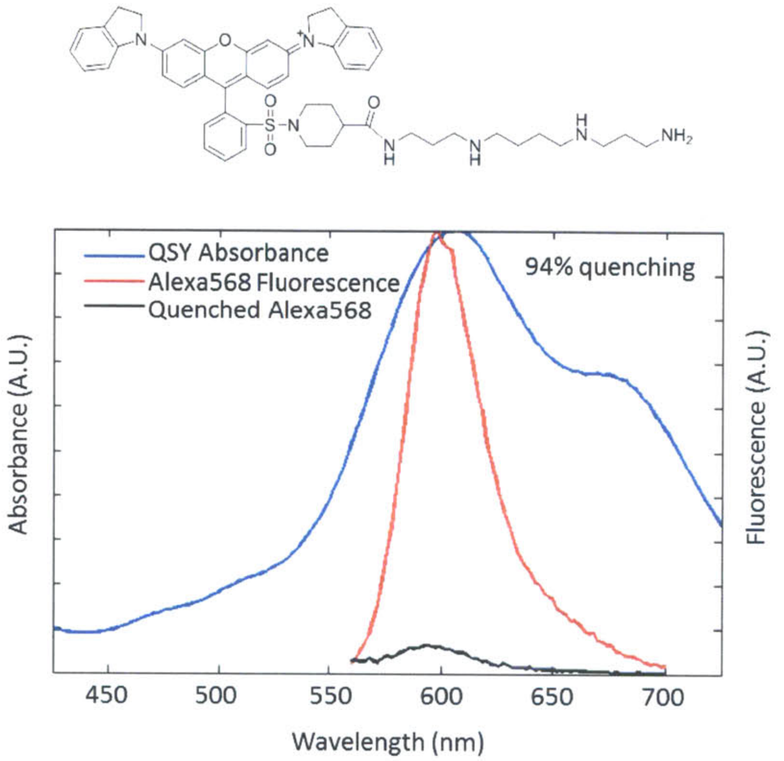


Figure 5-4. Characterization of QSY21-spermine.
Top: structure of QSY21-spermine. Bottom: absorption spectrum of QSY21-spermine in PBS pH 7.4 (blue). Fluorescence emission spectra of 80 nM mSA-Alexa568 in PBS pH 7.4 before (red) and after (black) the addition of 40 μM QSY21-spermine.

To test cell membrane-permeability of QSY21-spermine, we incubated cells with resorufin acetate, which readily enters cells by diffusing across the plasma membrane. Inside cells, resorufin acetate is hydrolyzed by endogenous esterases to yield a red fluorescent product, resorufin. While QSY21-spermine efficiently quenches resorufin fluorescence in solution, we verified that this reagent does not quench intracellular fluorescence in cells loaded with resorufin (Figure 5-5). On live cells, we observed 91% quenching of mSA-Alexa568 labeled surface AP-LDLR by 40 μM QSY21-spermine within 1 min (Figure 5-6).

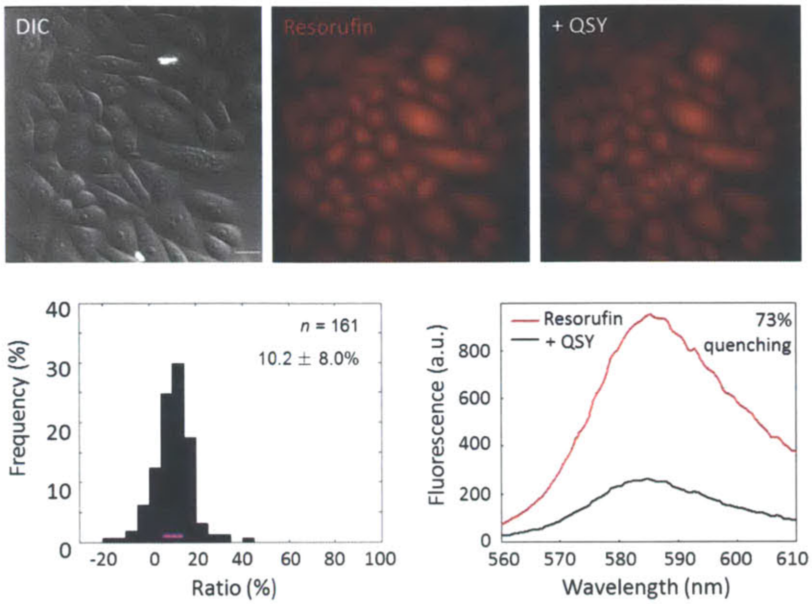


Figure 5-5. Testing membrane permeability of QSY21-spermine.
Top: representative images taken immediately before and after the addition of QSY21-spermine to T-REx CHO A7 cells loaded with resorufin acetate (a gift from Daniel S. Liu). Scale bar = 20 μm. Bottom left: distribution of quenching extents, for single cells from images on the top. Mean is shown ± s.d. Bottom right: fluorescence quenching of 100 μM resorufin by 160 μM QSY21-spermine in PBS pH 7.4.

Internalization assay

Before applying these labeling and surface quenching protocols to image LDLR internalization, we utilized the T-REx system (Invitrogen) to control AP-LDLR expression level in CHO ldlA-7 (T-REx CHO A7), a cell line that lacks endogenous LDLR (*24*). Details are described in the Experimental methods. We proceeded to image wild-type LDLR internalization in single cells. For quantitation, background-corrected total Alexa568 intensities were recorded for each labeled cell before (I_B) and after (I_A) treatment with QSY21-spermine. We used the ratio I_A / I_B to express the percent of internalized receptor for single cells. The frequency distribution histograms of this ratio and characteristic images are shown in Figure 5-6.

Immediately after labeling, cells displayed a ring-like fluorescence pattern. This fluorescence was nearly completely quenched after addition of QSY21-spermine. However, if the labeled cells were incubated at 37 °C for 5 minutes, fluorescence puncta were observed inside cells, suggesting internalization of labeled AP-LDLR into endosomes. Indeed, QSY21-spermine added at this time point quenched only fluorescence from the cell surface, but not fluorescence from internal puncta. During the 5-minute incubation at 37 °C, the percent of internalized receptors increased from 9.0 ± 3.3% to 43.7 ± 19.2% (mean ± s.d.). This internalization rate is comparable to previously reported values (~ 55% internalization for 5-minute incubation) obtained using radioactively labeled LDL (*5,25*), but the notable advantage of our assay is its single cell resolution.

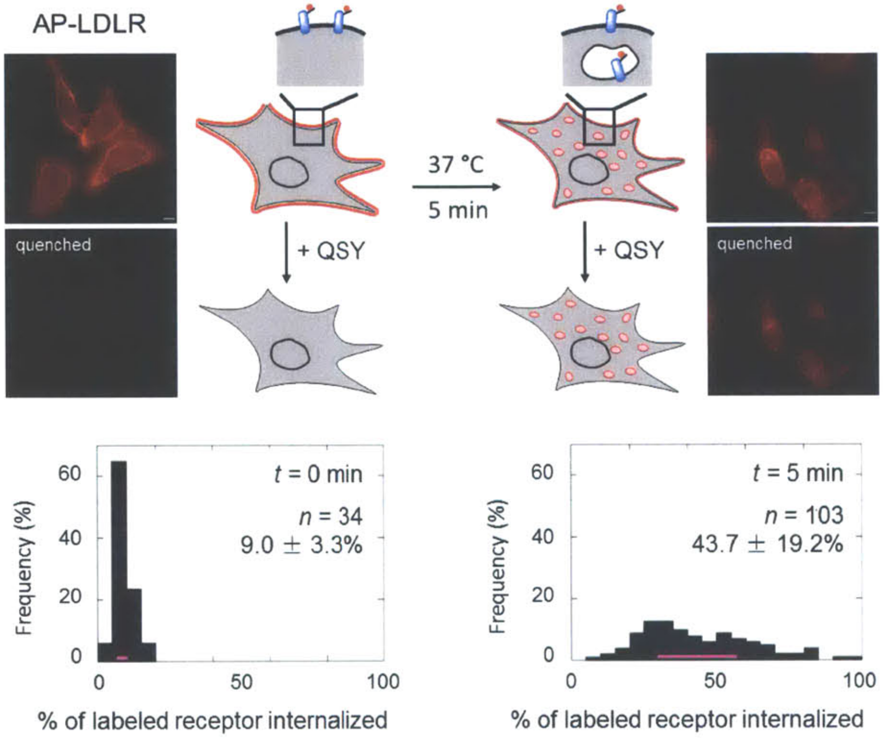


Figure 5-6. Internalization analysis time course of wild-type LDL receptor.
T-REx CHO A7 cells (lacking endogenous LDLR) expressing BirA and AP-LDLR were labeled with mSA-Alexa568. These cells were then incubated at 37 °C for 5 min. Representative images are shown immediately before and after surface fluorescence quenching with QSY21-spermine. Percent of internalized receptor was quantified for single cells, and the distributions of these percentages are shown in histograms, with pink lines denoting the interquartile range (25-75%). Data are given as mean ± s.d. Scale bars = 10 μm.

We used the same assay to analyze the internalization of three LDLR mutants. Previous studies have shown that Y807 mutants (Y807C and Y807A) and C-terminally truncated mutant (Δ792) of LDLR are defective in endocytic internalization, due to lack of targeting to clathrin-coated pits (*17*). Indeed, our assay shows that all three LDLR mutants remained on the cell surface during the 5-minute incubation period (Figure 5-7), which contrasts with the rapid internalization of wild-type LDLR.

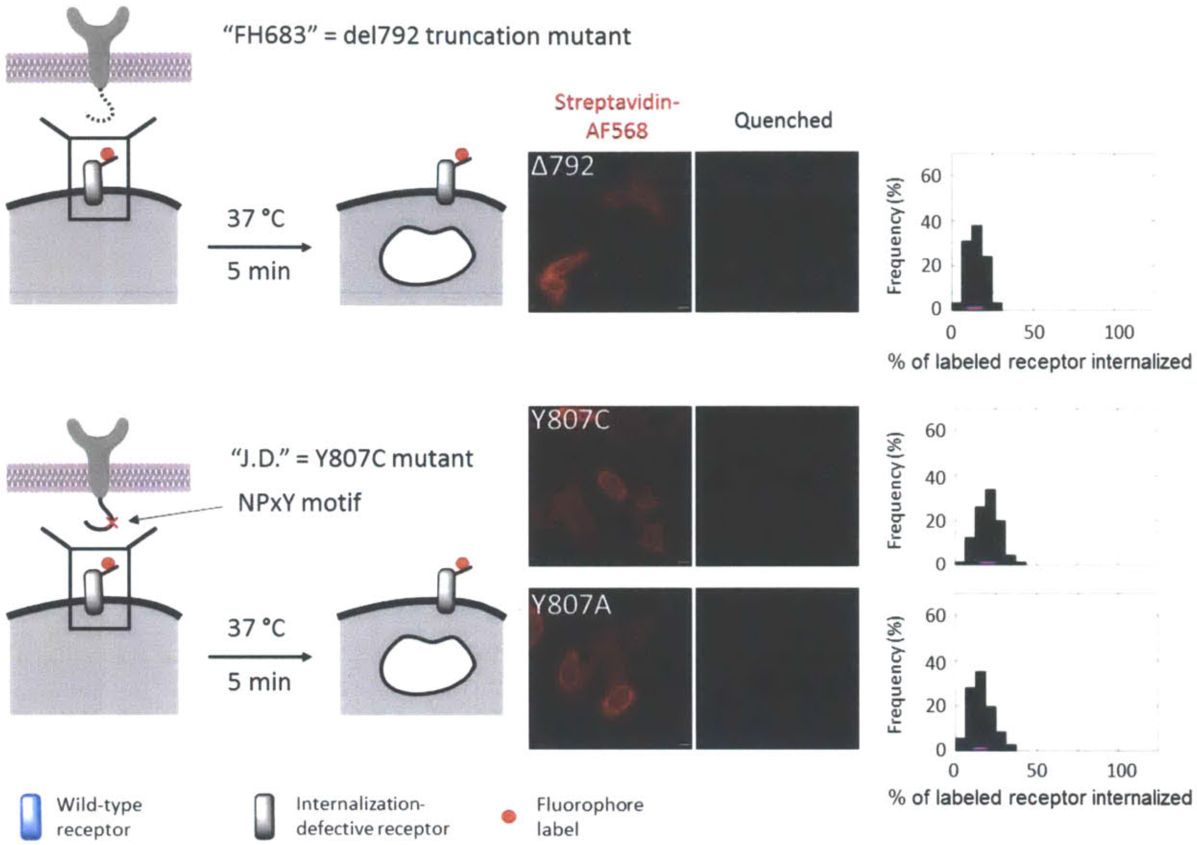


Figure 5-7. Internalization analysis of mutant LDLR.
T-REx CHO A7 cells expressing BirA and the indicated LDLR construct(s) were labeled with mSA-Alexa568, as described in Figure 5-6, and then incubated at 37 °C for 5 min. Representative images are shown immediately before and after surface fluorescence quenching with QSY21-spermine. Percent of internalized receptor was quantified for single cells, and the distributions of these percentages are shown in histograms, with pink lines denoting the interquartile range (25-75%). Data are given as mean ± s.d. Scale bars = 10 μm.

5.3 Determine receptor oligomerization state during endocytic internalization

Chemical crosslinking experiments have detected LDLR dimers in both purified samples and in cells (*17*). In addition, LDL particles have been visualized on fixed cells by electron microscopy, and have been observed to form dimers on the cell surface and within clathrin-coated pits (*26*). It is unclear, however, if the underlying receptors are also dimeric, if they are dimeric in the ligand-free state, and if cell fixation alters protein distribution. New methodology is needed to study this fundamental aspect of LDLR biochemistry more completely and with greater accuracy.

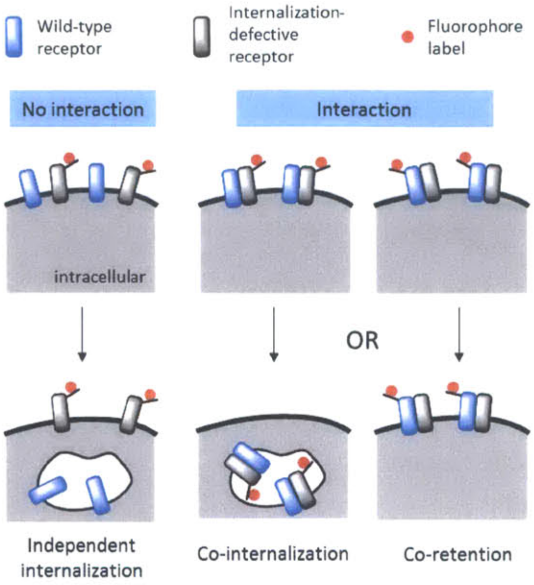

Wild-type receptor
Internalization-defective receptor
Fluorophore label
No interaction
Interaction
intracellular
OR
Independent internalization
Co-internalization
Co-retention

Figure 5-8. Scheme for oligomerization assay.
Cells co-expressing wild-type and mutant LDLR are labeled with mSA-Alexa568 and subsequently incubated at 37 °C to allow WT receptor to internalize. If internalizing receptors remain as monomers, then labeled AP-mutant receptor will remain on the cell surface (first row). Receptor oligomerization is revealed as co-internalization with the WT receptor (second row) and/or co-retention with the mutant (third row). The extent of internalization is quantified for each cell, by surface fluorescence quenching with QSY21-spermine after the incubation period.

Oligomerization assay

Building upon the internalization assay described in the previous section, we established a fluorescence imaging method to probe LDLR oligomerization state during the endocytic process in living cells. As shown in Figure 5-8, this method utilizes internalization-defective LDLR mutants and is composed of two symmetric assays. The first assay tests whether co-expression of wild-type LDLR facilitates internalization of these mutants. Conceptually, this strategy is similar to co-immunoprecipitation; the interaction is detected by "co-internalization" of mutant with the wild-type receptor into the cell. The second assay tests whether co-expressed mutant receptors negatively affect the internalization of wild-type LDLR. In other words, the interaction is manifested via "co-retention" of the wild-type with the mutant receptor on the cell surface. Taken together, these two assays provide complementary information of how receptors associate with each other and whether there is a dominant effect on internalization from either the wild-type or the mutant.

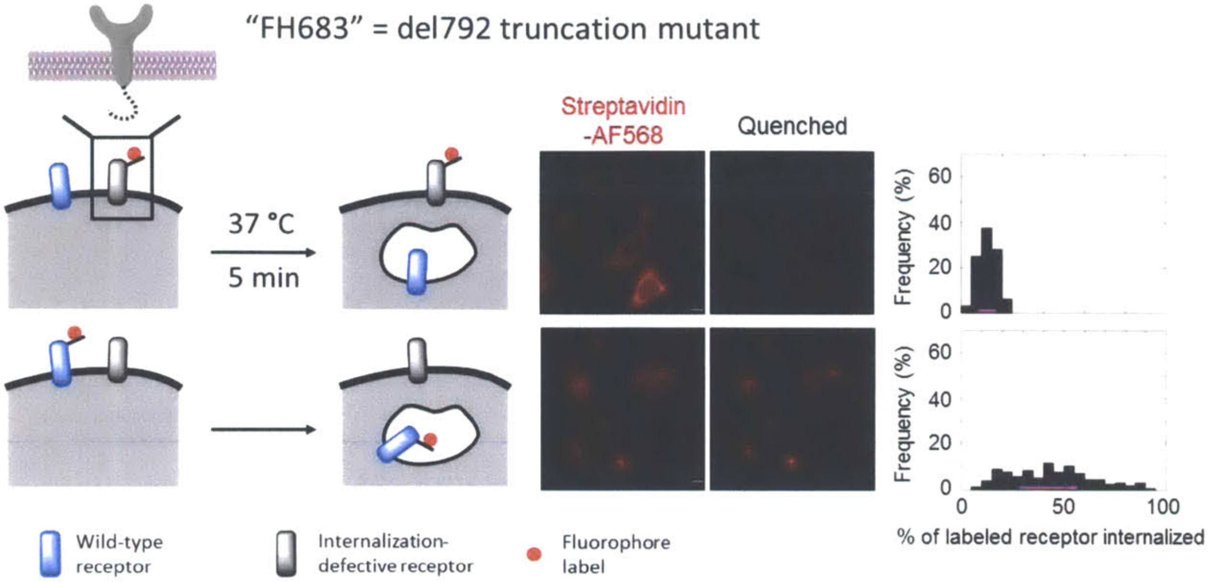


Figure 5-9. Co-internalization analysis of Δ792 LDLR.
T-REx CHO A7 cells (lacking endogenous LDLR) expressing BirA and the indicated LDLR construct(s) were labeled with mSA-Alexa568. Top: lack of co-internalization of mutant with the wild-type. Bottom: lack of co-retention of the wild-type with the mutant. Representative images are shown immediately before and after surface fluorescence quenching with QSY21-spermine. Percent of internalized receptor was quantified for single cells, and the distributions of these percentages are shown in histograms, with pink lines denoting the interquartile range (25-75%). Data are given as mean ± s.d. Scale bar = 10 μm.

To test for oligomerization via this co-internalization assay, we co-expressed wild-type LDLR with each of the three LDLR mutants in T-REx CHO A7 cells, and quantified the extent of internalization. In the case of truncation mutant Δ792 LDLR, we co-expressed AP-tagged mutant with untagged wild-type receptor. Our assay shows that Δ792 LDLR remains internalization-defective whether or not the wild-type is co-expressed ($p = 0.71$). In the complementary assay, we co-expressed AP-tagged wild-type and untagged mutant LDL receptor. We found that the internalization of the wild-type receptor does not slow down with Δ792 LDLR co-expression (Figure 5-9). These experiments suggest that Δ792 LDLR does not interact with the wild-type receptor during endocytic internalization. We note that the same amount of plasmid DNA of mutant and wild-type receptors was used to transfect cells, and that efficient co-expression was achieved in about 90% of transfected cells for all these experiments, as measured by immunofluorescence staining (see Experimental methods, Figure 5-24). However, one caveat of this analysis is that we did not experimentally check for the relative expression level.

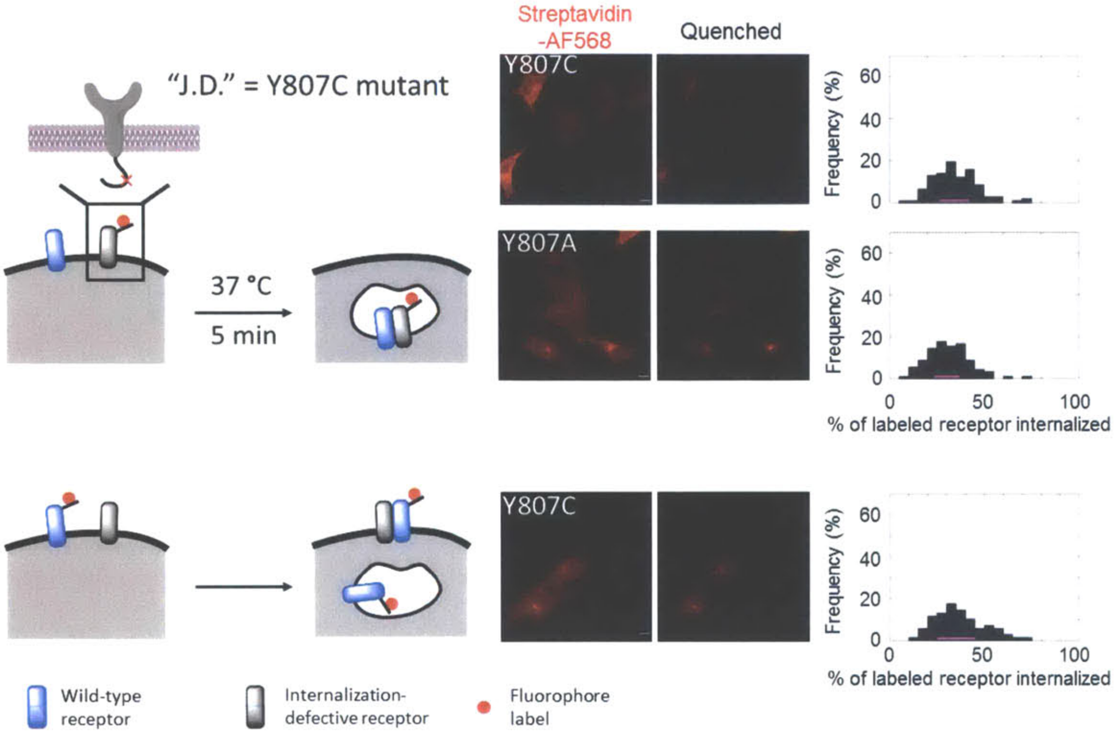


Figure 5-10. Internalization analysis of Y807 LDLR mutants.
T-REx CHO A7 cells expressing BirA and the indicated LDLR construct(s) were labeled with mSA-Alexa568, as described in Figure 5-6, and then incubated at 37 °C for 5 min. Top: co-internalization of mutant with the wild-type. Bottom: co-retention of the wild-type with the mutant. Representative images are shown immediately before and after surface fluorescence quenching with QSY21-spermine. Percent of internalized receptor was quantified for single cells, and the distributions of these percentages are shown in histograms, with pink lines denoting the interquartile range (25-75%). Data are given as mean ± s.d. Scale bars = 10 μm.

In the case of Y807 mutants, we performed the same set of experiments. We observed that both Y807C and Y807A entered cells more rapidly in the presence of wild-type LDLR ($p < 0.001$, Figure 5-10). At the same time, we found that the internalization of the wild-type receptor slows down significantly when it is co-expressed with Y807C mutant (comparing to Figure 5-6). These results suggest that Y807 mutants oligomerize and co-internalize with the wild-type LDLR.

As a negative control for this analysis, we co-expressed transferrin receptor (TfnR) with AP-Y807C LDLR. Like LDLR, the internalization of TfnR is mediated by clathrin-dependent pathway (*27*), yet these two receptors do not associate with each other. Enhanced green fluorescent protein (EGFP) was fused to the C-terminus of TfnR to facilitate examination of its expression level. As expected, the internalization of AP-Y807C LDLR remained low and was uncorrelated with the expression level of TfnR-EGFP ($R^2 < 0.001$) (Figure 5-11).

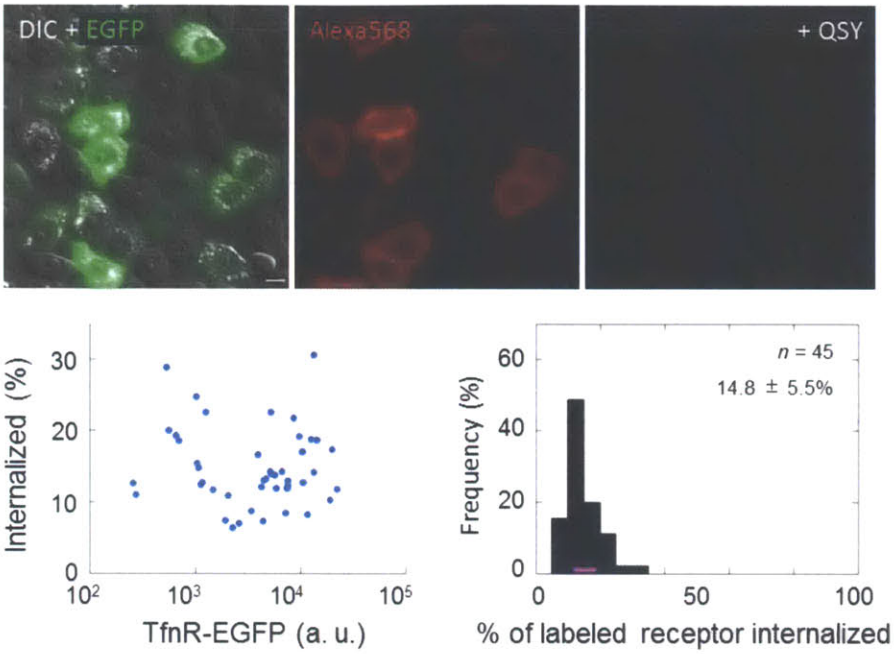


Figure 5-11. Internalization analysis of Y807C LDLR co-expressed with transferrin receptor (TfnR) as a negative control.
T-REx CHO A7 cells were co-transfected with AP-Y807C LDLR, BirA, and transferrin receptor fused to EGFP (TfnR-EGFP). Internalization after 5 min at 37 °C was imaged and analyzed. Top: representative images of TfnR expression (EGFP channel), and mSA-Alexa568 labeling before and after quenching with QSY21-spermine. Scale bar = 10 μm. Bottom left: plot of the internalization extent for single cells against TfnR expression level. $R^2 = 0.0007$. Bottom right: histogram showing the distribution of internalization extents. The pink line denotes the interquartile range (25–75%).

Why do internalization-defective mutants behave differently toward wild-type co-expression? We believe that in the case of Δ792 LDLR, the region required for LDLR oligomerization may be deleted, and thus co-internalization is not possible. Indeed, previous work has suggested that the cytoplasmic domain of LDLR is required for its oligomerization (*17*). An interesting direction for future work would be to extend this co-internalization analysis to LDLR mutants with various truncated regions. This would allow us to identify the location of this hypothesized oligomerization domain.

To support our fluorescence imaging analysis of receptor oligomerization, we sought to detect such interactions using traditional method: co-immunoprecipitation (co-IP) assay. We have designed two sets of co-IP experiments. In the first set, wild-type LDL receptors are differentially fused to AP-tag and HA epitope tag, and are co-expressed in T-REx A7 cells at controlled expression level. In another set, the AP-tagged LDL receptor is a truncation mutant Δ792, which has been suggested to be oligomerization-incompetent in our imaging analysis. These configurations are identical to our previous imaging experiments.

Figure 5-12 shows the SA-HRP blot and anti-HA western blot analysis of co-IP fractions. SA-HRP blot (panel A) confirms successful pull down of AP-WT LDLR from the lysate. Anti-HA Western blot (panel B) demonstrates successful expression of HA-WT LDLR. However, the majority of HA-WT LDLR (> 95%) is not co-immunoprecipitated with AP-LDLR, as indicated by the strong signal in the unbound fraction and almost non-existent signal in the elute fraction in panel B. We tried to push the limit of detection sensitivity by loading maximum amount of sample on the gel and by applying longer exposure time when developing the blot. Under these conditions, we were able to detect weak signals from co-IP experiments. The bottom of panel B shows that HA-WT LDLR was co-immunoprecipitated with AP-WT LDLR, but not with AP-Δ792 LDLR.

These results are consistent with our imaging assays. Indeed, the Δ792 mutation disturbs receptor oligomerization, and at the same time, it destroys the co-internalization. It is likely that LDL receptor monomers oligomerize only transiently, such that at any given time, the interacting subpopulation constitutes only a small fraction of the total receptor pool which cannot be captured efficiently by co-IP and detected with western blot analysis. Additionally, LDL receptor oligomers might not be stable enough to survive cell lysis conditions. This demonstrates that our imaging analysis is more sensitive in detecting oligomer formation than traditional biochemical assays.

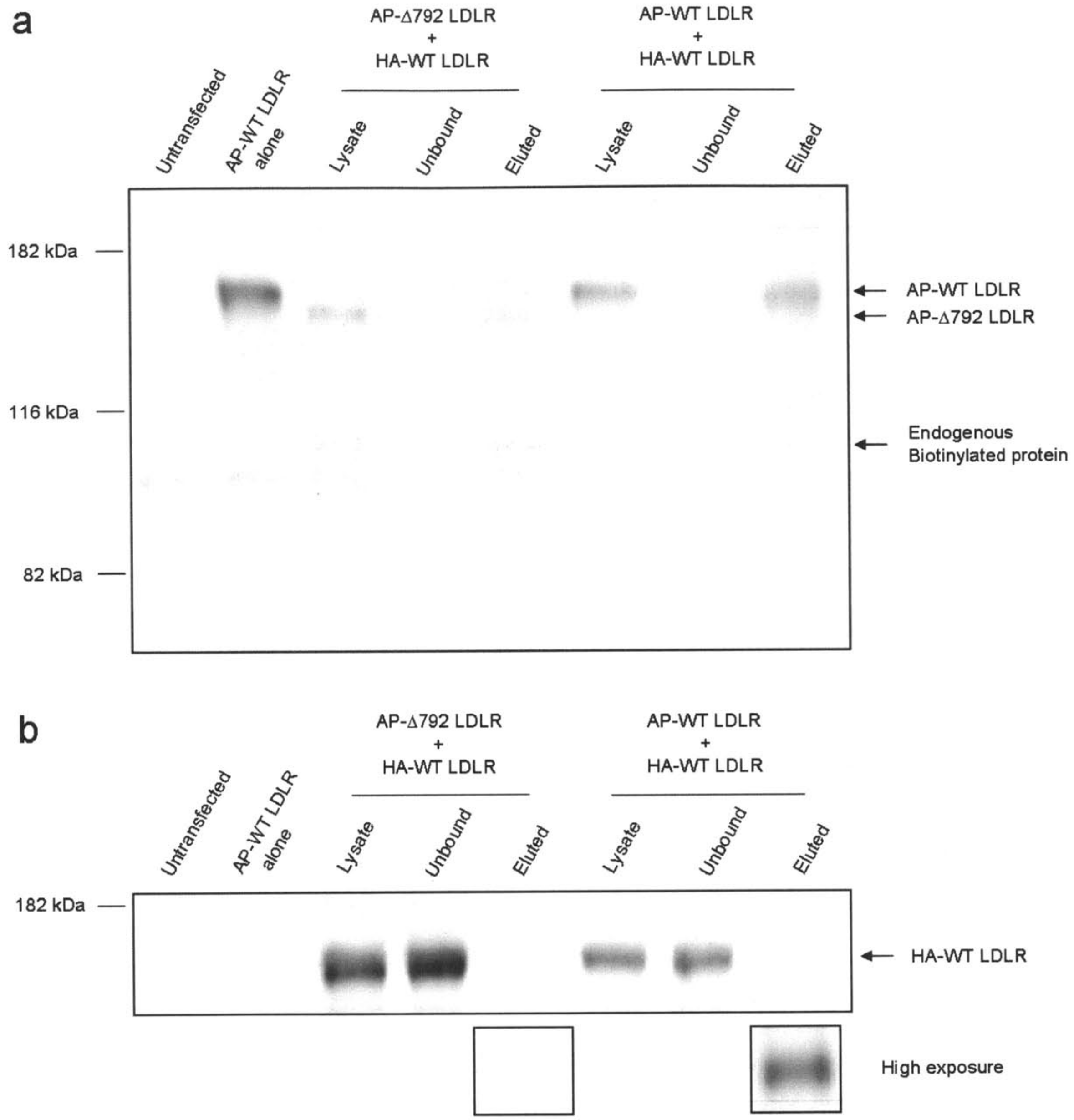

a
AP-Δ792 LDLR + HA-WT LDLR
AP-WT LDLR + HA-WT LDLR
Untransfected
AP-WT LDLR alone
Lysate
Unbound
Eluted
Lysate
Unbound
Eluted
182 kDa
116 kDa
82 kDa
AP-WT LDLR
AP-Δ792 LDLR
Endogenous Biotinylated protein
b
AP-Δ792 LDLR + HA-WT LDLR
AP-WT LDLR + HA-WT LDLR
Untransfected
AP-WT LDLR alone
Lysate
Unbound
Eluted
Lysate
Unbound
Eluted
182 kDa
HA-WT LDLR
High exposure

Figure 5-12. Co-immunoprecipitation analysis of LDLR oligomerization.
T-REx CHO A7 cells expressing BirA and the indicated LDLR construct(s) were lysed and analyzed by (a) streptavidin-HRP blot and (b) anti-HA blot. Whole lysate was analyzed ("Lysate"), along with the streptavidin-enriched fraction ("Eluted"), and the unbound supernatant ("Unbound"). Control lanes are shown with untransfected cells, and cells expressing only AP-WT LDLR. AP-Δ792 LDLR is ~5 kDa smaller than AP-WT LDLR.

LDL ligand does not affect oligomerization state

To test whether the addition of ligand could affect LDL receptor oligomerization, we repeated the above panel of experiments with LDL added into the medium at saturating concentrations, and we obtained the same results (Figure 5-13). Since each LDL particle contains only one copy of ApoB-100 protein, which binds to LDLR in a 1:1 molar ratio (*28*), it is unlikely that LDL binding would promote oligomerization of LDLR. Consistent with these observations, our mutation analysis suggests that LDLR oligomerization is mediated by its cytoplasmic tail, which is inaccessible to LDL.

Regarding whether LDL promotes receptor internalization, two models have been proposed in the past: the escalator model and the elevator model. The escalator model states the receptor is constitutively internalized whether or not the ligand is present. In this model, the bound ligand is passively endocytosed into the cell, in the same way as passengers carried by the escalator. To support this model, electron microscopy studies have shown that LDLR is spontaneously localized in clathrin-coated pits and internalized even in the absence of its ligand (*29*). The elevator model, on the other hand, states that that receptor internalization is dependent on ligand binding. That is, bound-ligand actively triggers receptor internalization (via conformational change), in the same way as passengers press button to call for elevator. Previous experiments using monensin, a carboxylic ionophore that inhibits receptor recycling, indicated that half of the LDLR population on the cell surface is internalized only if LDL is added into the medium (*25*), thus supporting the elevator model. Our experiment helps resolve this issue by showing that the extent of internalization for wild-type LDLR is the same whether or not it is bound to LDL ($p > 0.4$, Figure 5-14).

We note that the same batch of LDL proteins have been shown to bind to cells expressing LDL receptor when they were added at 100 μg/mL (see Experimental methods, Figure 5-22).

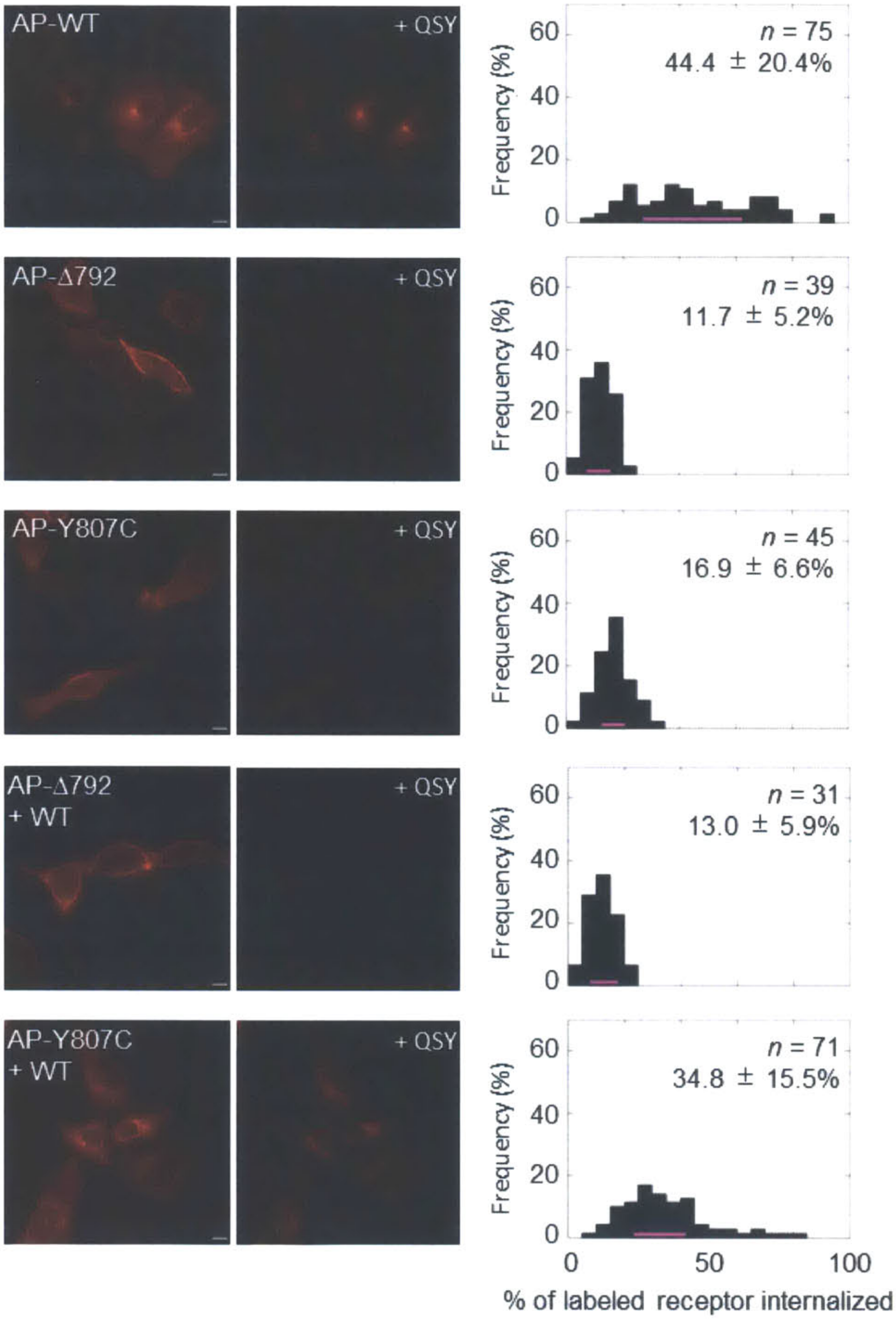


Figure 5-13. Internalization analysis of wild-type and mutant LDLR in the presence of LDL. T-REx CHO A7 cells expressing BirA and the indicated receptor(s) were labeled with mSA-Alexa568 in the presence of 100 μg/mL LDL, and then incubated at 37 °C for 5 minutes with 10 μg/mL LDL. Representative images are shown immediately before and after surface quenching with QSY21-spermine. The distribution of these ratios is shown in histograms, and pink lines denote the interquartile range (25-75%). Data are given as mean ± s.d. Scale bar = 10 μm.

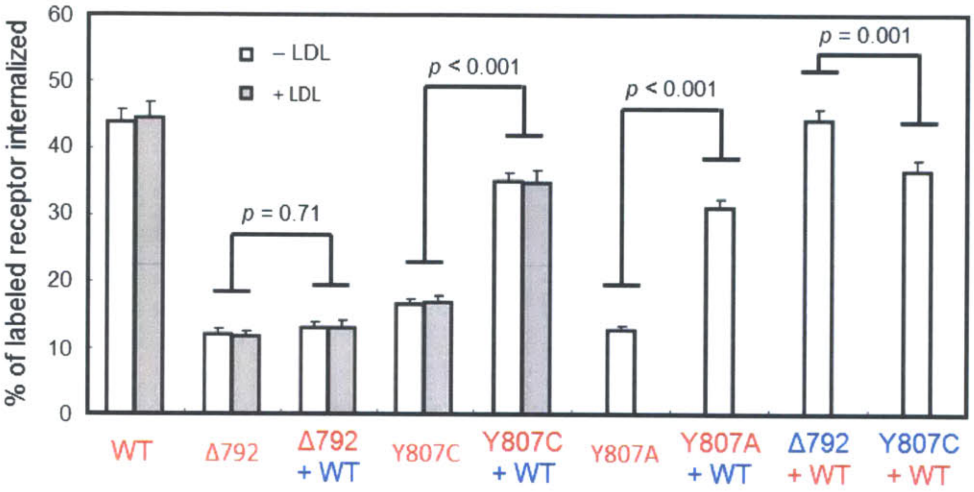


Figure 5-14. Summary of LDLR internalization analysis.
The presence of LDL does not affect receptor internalization ($p > 0.4$, Kolmogorov-Smirnov test). Unlike the truncation mutant Δ792 LDLR, the internalization of Y807C LDLR mutant increases significantly when co-expressed with wild-type LDLR. Meanwhile, the internalization of wild-type LDLR slows down significantly only when it is co-expressed with Y807C mutant, but not with Δ792 LDLR. Data are given as mean ± s.e.m.

5.4 Enzymatic incorporation of unnatural probe desthiobiotin

In addition to its natural biotin substrate, BirA also recognizes an analogue, desthiobiotin (DTB), and efficiently ligates it to AP peptide (*30*). Due to structural similarities, both DTB and biotin bind to the same site on streptavidin, but the absence of sulfur atom in DTB causes a reduction of binding affinity by approximately 100-fold, which makes its binding to streptavidin effectively reversible (*31*). Taking advantage of this difference in reversibility, it is possible to distinguish between biotinylated versus desthiobiotinylated targets. In tissue imaging, this would be useful for removing background streptavidin-fluorophore labeling from endogenous biotinylated protein.

We started with testing the extent of reversibility of streptavidin binding to DTB. To maximize DTB labeling signal, we prepared the N-hydroxysuccinimide ester of DTB (DTB-NHS) and used it to non-specifically label primary amines in cells. Figure 5-15 shows that desthiobiotinylated cells are readily labeled with streptavidin-dye conjugate, and this labeling is completely abolished when cells are pre-blocked with unlabeled streptavidin. Further, incubating blocked cells with 1 mM biotin at 37 °C for 1 hour causes streptavidin to dissociate from DTB targets and rescues the subsequent labeling signal. Encouraged by this initial result, we moved on to improving enzymatic DTB ligation to acceptor peptides.

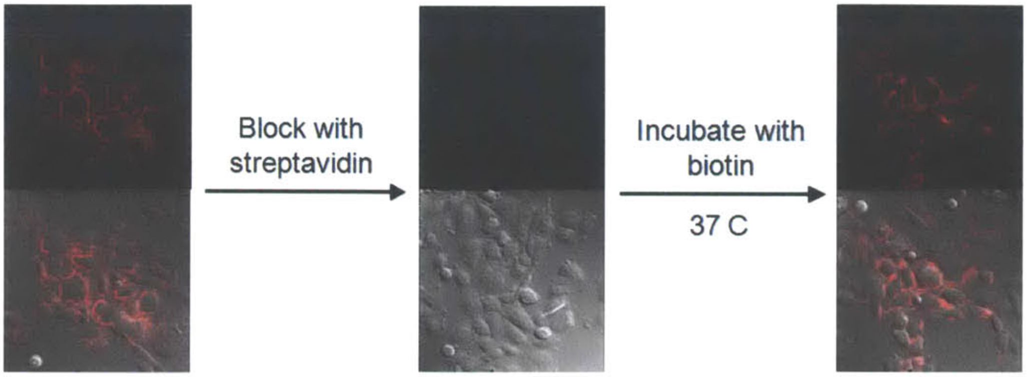


Figure 5-15. Reversible binding of streptavidin to DTB.
CHO cells were non-specifically desthiobiotinylated with DTB-NHS. Cells were likely dead due to high toxicity of this reagent, but they are not permeabilized. DTB on the cell surface could be easily detected with streptavidin-Alexa568 conjugate. This labeling was completely abolished if cells were pre-blocked with streptavidin. Because the binding of streptavidin to DTB is reversible, the blocking could be relieved by incubating cells with 1 mM biotin at 37 °C for 1 hr. Top: fluorescence image; bottom: overlay with DIC.

Enzyme-mediated DTB ligation to AP peptide

We started with in vitro ligation experiments using purified BirA enzyme. Figure 5-16 confirms that BirA efficiently mediates the ligation of DTB to AP peptide. Omitting either ATP or the enzyme gives no product formation. Reaction products were purified by reversed-phase HPLC and subsequently confirmed by MS. Under the same reaction condition (1 hour incubation at 30 °C), quantitative conversion was observed for both biotinylation and desthiobiotinylation of the AP peptide. We then tested whether site-specific DTB labeling of plasma membrane proteins could be achieved using BirA targeted to the endoplasmic reticulum. However, this proved to be unsuccessful, even with the presence of 1 mM DTB in the cell culture medium. In comparison, 10 μM biotin in the cell culture medium is sufficient for completely biotinylating the AP-fusion of cell surface protein (data not shown).

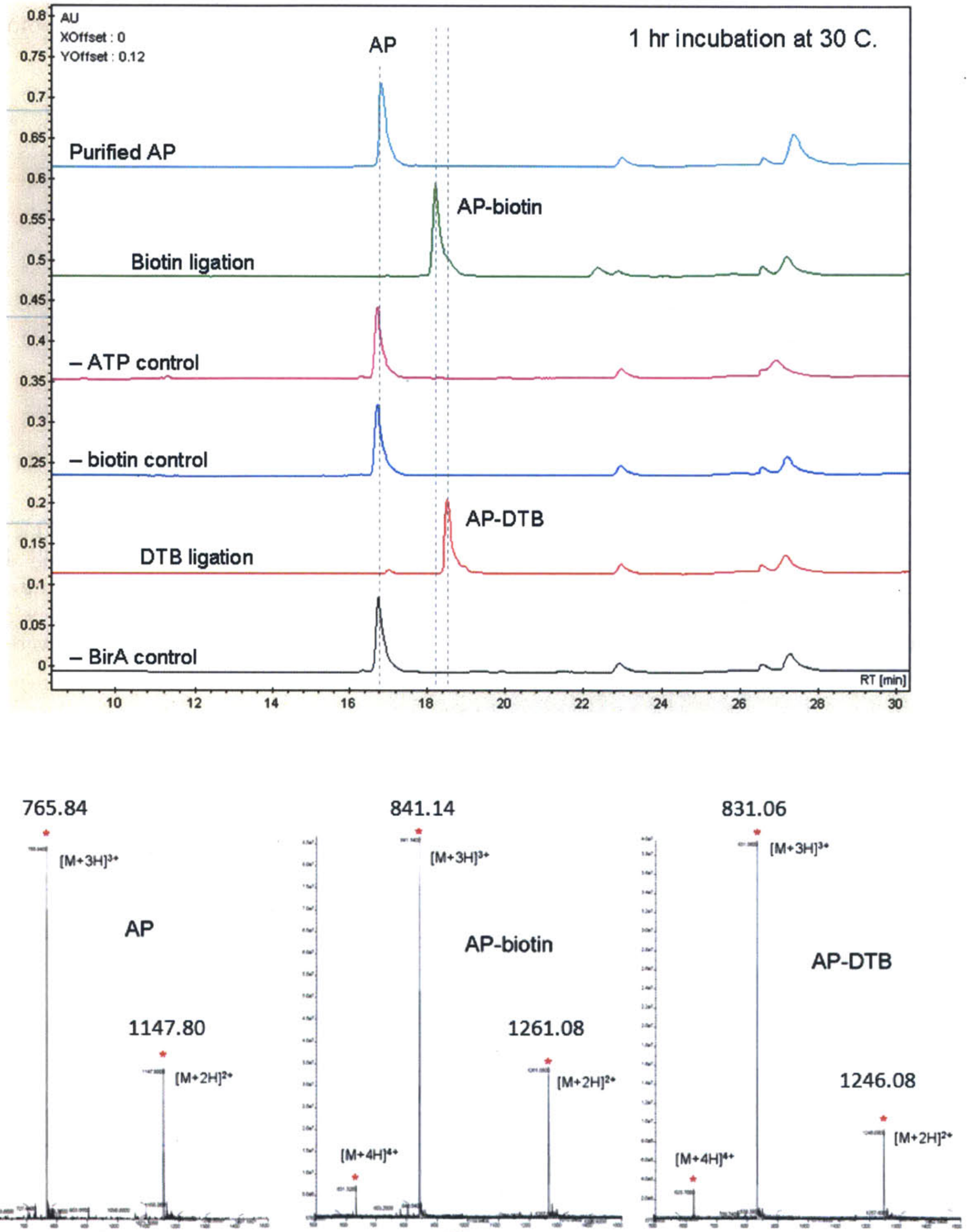


Figure 5-16. In vitro desthiobiotin ligation using purified BirA.
Top: HPLC analyses of ligation reaction. Purified AP peptide (100 μM) was incubated with 2 μM BirA, 1 mM DTB, 4 mM ATP, and 5 mM magnesium acetate in bicine buffer pH 8.3, at 30 °C for 1 hour. The reaction was quenched with 45 mM EDTA, and analyzed with reversed-phase HPLC. Bottom: mass- spec confirmation of ligation products.

The alternative approach is cell surface desthiobiotinylation of AP with purified BirA. It has been shown earlier that BirA purified from *E. coli* often tightly binds to biotin or biotin adenylate (bio-AMP). The existence of biotin or bio-AMP at sub-stiochiometric level is less a problem for in vitro ligation experiments, in which BirA is applied with catalytic amount (2 μM enzyme versus 100 μM AP substrate), than it is a problem for cellular labeling. A simple estimation showing that BirA enzyme is about 10^3- to 10^4-fold in excess of AP peptide in a standard labeling protocol. Contaminating biotin and/or bio-AMP can be removed by first incubating BirA with purified AP peptide and ATP and then dialyzing away small molecules from the reaction mixture. As expected, this recycled BirA gave no labeling in the absence of biotin (Figure 5-17). Further, the presence of 1 mM DTB in the reaction mixture gave specifically labeling of cells transfected with AP-CFP-TM. Compared to biotin, DTB ligation is a much slower reaction and requires high concentration of DTB probe (1 mM). This problem is even more severe for intracellular labeling. We reasoned that the intracellular DTB level was too low for efficient ligation, even with 1 mM DTB added to the culturing media.

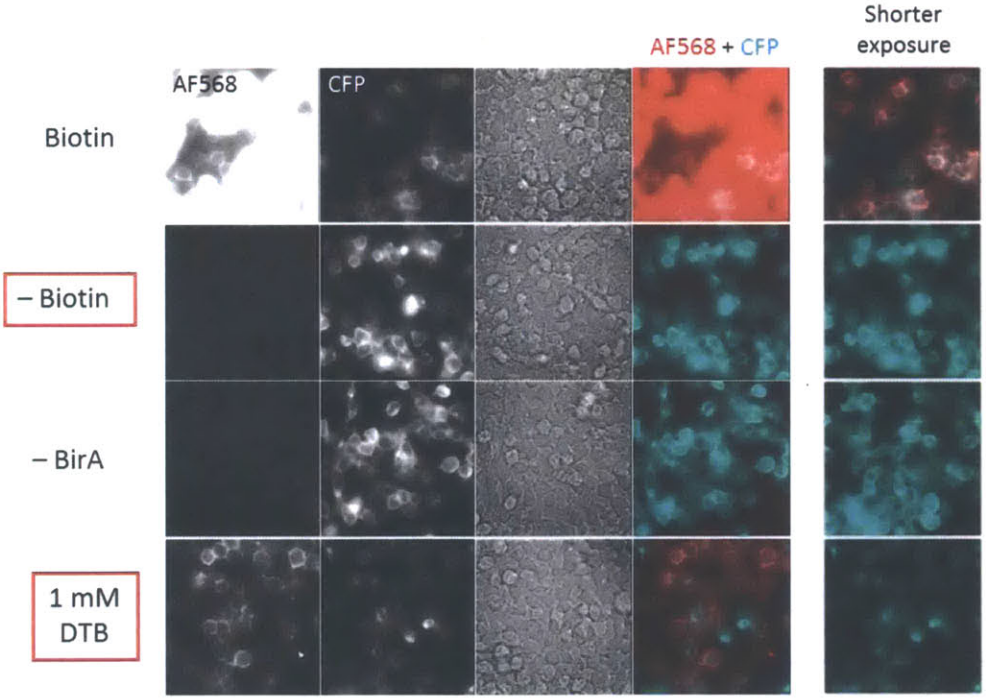


Figure 5-17. Cell surface desthiobiotinylation of AP-CFP-TM protein with exogenous BirA.
DTB on the cell surface was targeted with streptavidin- Alexa568. Cells were incubated with 0.3 μM BirA, 4 mM ATP, 5 mM $MgCl_2$, and 1 mM DTB in DPBS at room temperature for 30 min. 10 μM biotin was used instead of DTB as positive control for labeling (first row). Omitting DTB or biotin in the reaction gave no labeling.

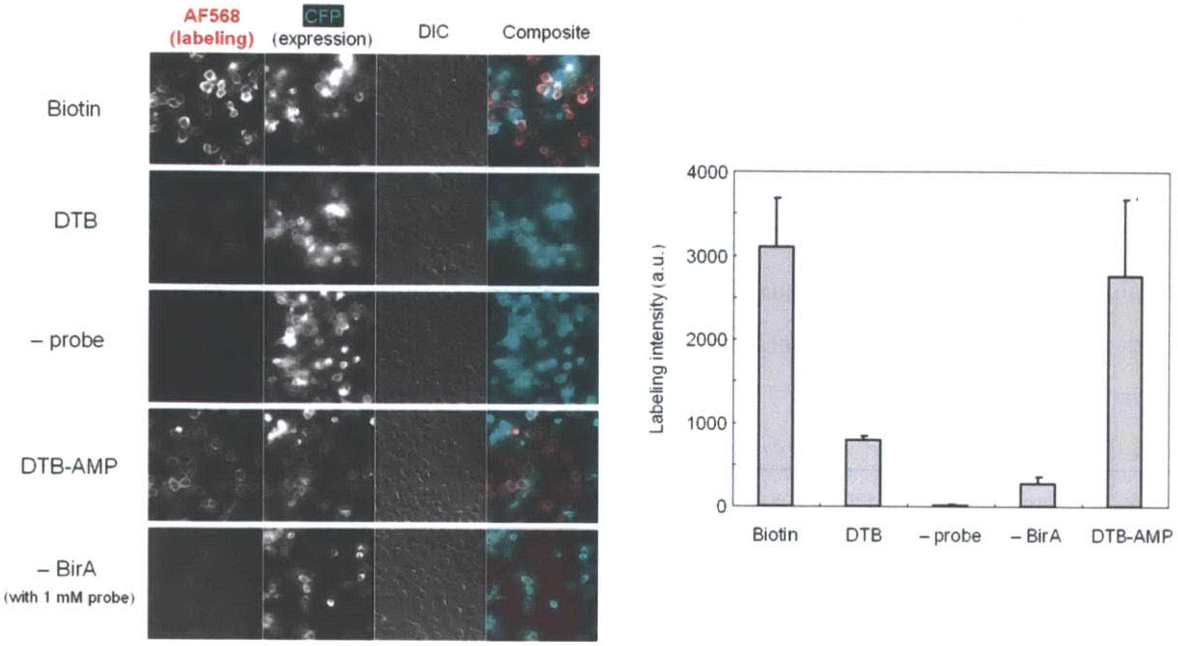

AF568
(labeling)
CFP
(expression)
DIC
Composite
Biotin
DTB
– probe
DTB-AMP
– BirA
(with 1 mM probe)
4000
3000
2000
1000
0
Labeling intensity (a.u.)
Biotin
DTB
– probe
– BirA
DTB-AMP

Figure 5-18. Wild-type BirA efficiently ligates DTB-AMP to AP peptide on the cell surface.
Left: fluorescence images comparing DTB adenylate labeling with biotinylation and DTB ligation on AP-CFP-TM. Right: quantitation of streptavidin-AlexaFluor568 (AF568) labeling intensity in these images, with background subtracted.

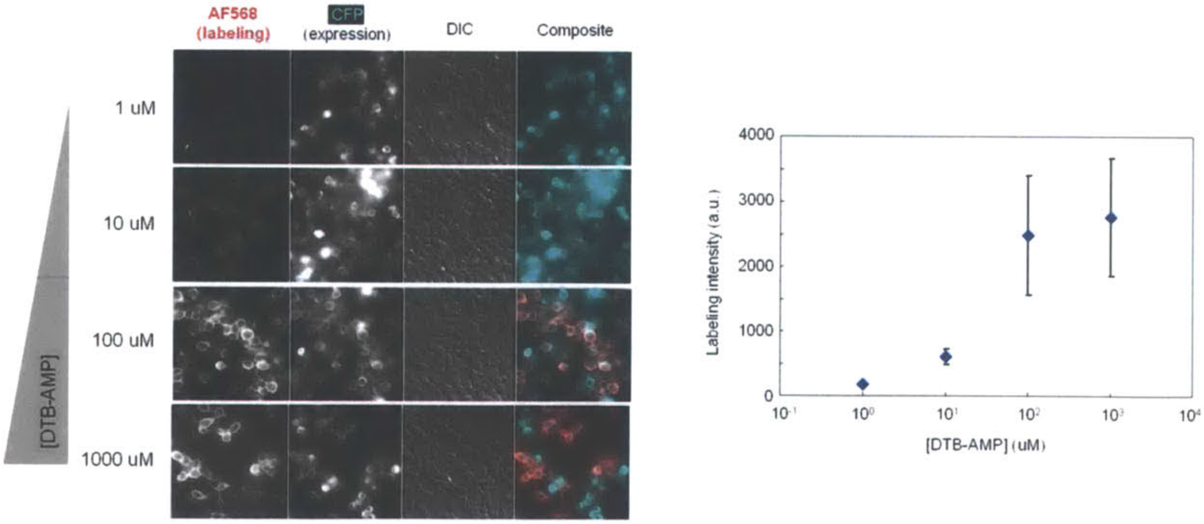

AF568
(labeling)
CFP
(expression)
DIC
Composite
1 uM
10 uM
100 uM
1000 uM
[DTB-AMP]
4000
3000
2000
1000
0
Labeling intensity (a.u.)
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
[DTB-AMP] (uM)

Figure 5-19. Titration of DTB adenylate probe and quantitation.
Cells were incubated with WT BirA and DTB adenylate (DTB-AMP) probe for 10 min at room temperature. Background was subtracted from quantiations.

We set out to boost desthiobiotin ligation efficiency both on the cell surface and inside cells. On the cell surface, we reasoned that we could improve the kinetics by supplying the enzyme with its activated intermediate, desthiobiotin adenylate. This bypasses the first catalytic step of carboxyl group activation, which turned out to be the rate-limiting step of catalysis. Indeed, we found that cell surface ligation is significantly improved by using desthiobiotin adenylate. The level of labeling with 100 μM probe for 10 min is comparable to biotinylation, presumably saturating all the AP sites on the cell surface (Figure 5-18 and Figure 5-19).

To facilitate intracellular desthiobiotin labeling, we reasoned that cellular loading with desthiobiotin is likely a slow process, due to the negatively charged carboxyl group. This could be improved by using the acetoxymethyl ester of desthiobiotin (DTB-AM ester) which readily permeate through the plasma membrane. Once inside the cell, DTB-AM ester could be hydrolyzed by endogenous esterases to unmask the proper desthiobiotin. We observed that, with 100 μM DTB-AM supplied into the media, labeling of AP-actin with cytosolic BirA completed in about 1 hour (Figure 5-20).

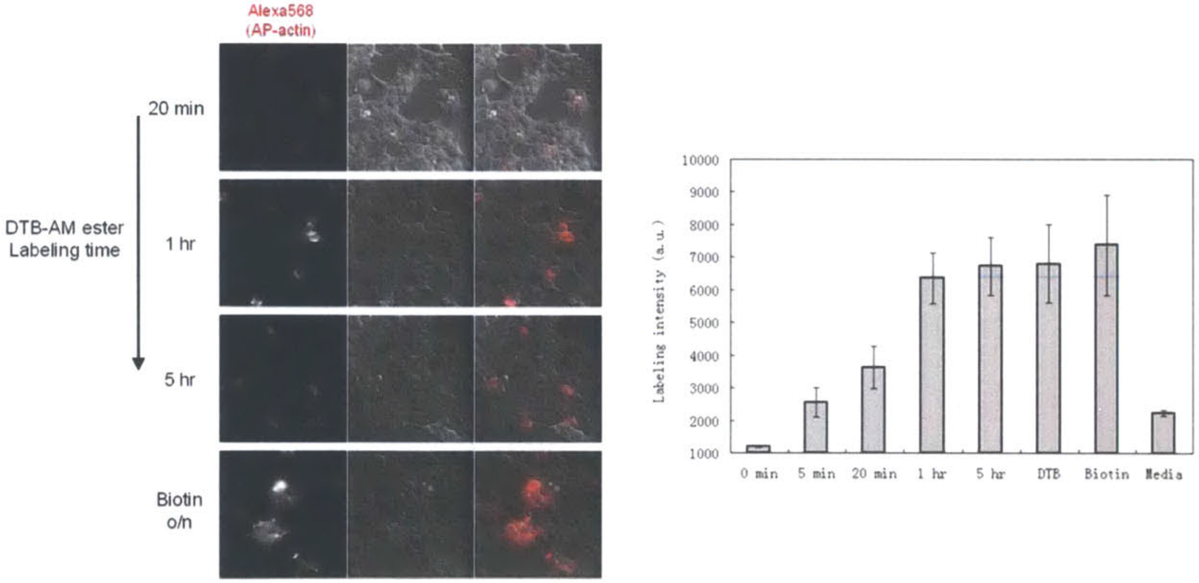


Figure 5-20. Intracellular DTB ligation using AM ester.
Left: fluorescence images showing the time course analysis of DTB-AM ester labeling. Right: quantitation of time course analysis.

Applications of intracellular DTB labeling

Our goal is to develop a method to detect desthiobiotin (DTB) targets with streptavidin, while omitting background from endogenous biotin. This would be especially useful for fixed tissue imaging. After cell surface labeling with DTB, cellular fixation and membrane permeabilization, cells were pre-blocked with streptavidin. Washing with biotin selectively dissociate streptavidin from DTB-labeled targets, but not endogenous biotinylated proteins (Figure 5-21).

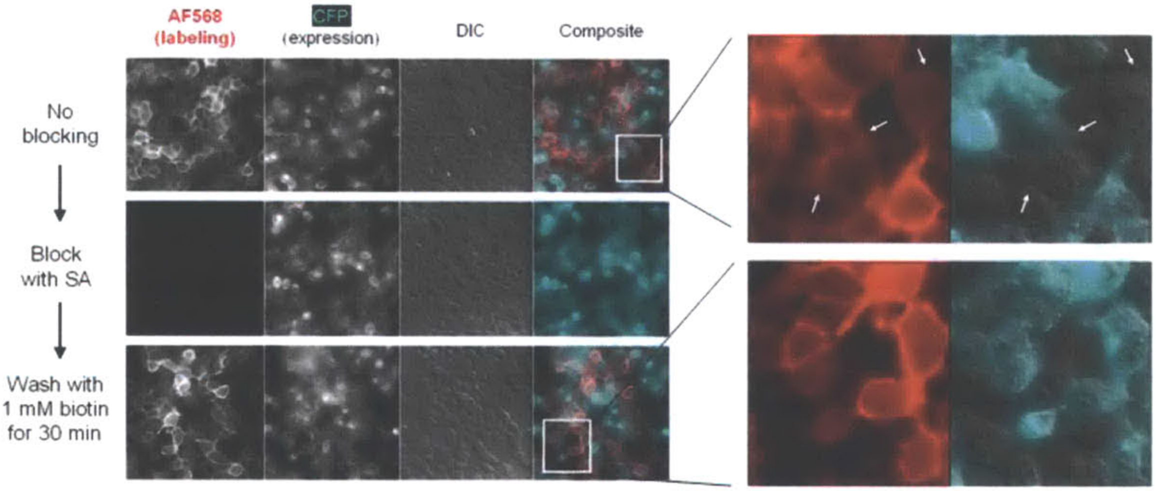


Figure 5-21. Blocking endogenous biotin signal using enzyme-mediated destiobiotinylation. There was a high level of endogenous biotin signal. This background, together with DTB label, could be almost completed blocked with streptavidin. Incubation with 1 mM biotin at 30 °C for 30 min selectively rescued DTB label on the cell surface. Arrows indicate labeling of endogenous biotin in untransfected cells.

5.5 Conclusion

In summary, we have developed a fluorescence imaging assay to probe LDLR oligomerization during internalization. The endocytosis of both wild-type and internalization-defective receptor mutants were analyzed in the context of receptor co-expression in living cells. If receptor monomers interact with each other, co-expressed wild-type and mutant receptors are expected to mutually affect each other's internalization rates.

Using site-specific protein labeling mediated by biotin ligase and monovalent streptavidin, as well as cell surface fluorescence quenching with a novel reagent, we quantified LDLR internalization under various conditions. We found that, during endocytosis, LDLR is capable of oligomerizing in a ligand-independent manner, and that this interaction is likely mediated by its cytoplasmic domain. LDLR dimerization was suggested in previous electron microscopy studies and chemical crosslinking experiments but has never been observed directly in live cells, free of potential artifacts from cell lysis and fixation. An important feature is that our assay does not require the use of ligand, making it possible to study receptor oligomerization in both ligand-bound and ligand-free states. This allowed us to discover that the rate of receptor internalization is not altered by the binding of low density lipoprotein ligand.

The assay described here utilizes known mutations that cause defects in receptor internalization. We chose the LDLR as a demonstration of this method, but we envision that this assay could be applied to other receptors as well. One potential limitation of such extension is that the internalization-defective mutant receptor has to be carefully chosen so as to avoid disrupting monomer interaction and/or receptor function.

Instead of using the natural substrate biotin, we also demonstrated that BirA efficiently catalyzes the ligation of desthiobiotin onto acceptor peptide, and that the kinetics of this reaction can be improved by using DTB adenylate ester for cell surface labeling or acetoxymethyl ester for intracellular applications. DTB labeling allows subtraction of endogenous biotinylation background from fixed cell sample, which is important for tissue imaging.

5.6 Experimental methods

Reagents and constructs

Enzymes used in cloning and restriction digestions were purchased from New England BioLabs. Oligonucleotides used in cloning and mutagenesis were purchased from Invitrogen. CHO ldlA-7, a Chinese hamster ovary cell line defective in endogenous LDLR, was a kind gift from Prof. Monty Krieger (MIT) (*24*). HeLa and CHO-K1 cells were purchased from ATCC. Dulbecco's modified Eagle medium (DMEM), Dulbecco's phosphate buffered saline (DPBS) were purchased from Gibco. Fetal bovine serum (FBS) was purchased from PAA Laboratories. Bovine serum albumin (BSA) was purchased from Amresco, L-proline was from USB Corp., sodium pyruvate and resorufin were from TCI America, and biotin was a gift from Tanabe USA. Spermine and 37% (v/v) formaldehyde were purchased from Alfa Aesar. Tween 20 was purchased from Sigma-Aldrich. Blasticidin, AlexaFluor568 succinimidyl ester, QSY21 succinimidyl ester, rabbit α-HA, and goat α-rabbit conjugated to AlexaFluor647 were purchased from Invitrogen.

Monovalent streptavidin (mSA) and its conjugate with Alexa Fluor 568 (Invitrogen) were prepared as previously described (*32*). By measuring the absorbance spectrum, the purified conjugate had a concentration of 7.9 μM and an average dye-to-protein ratio of 2.9. This conjugate was stored at –80 °C.

QSY21-spermine was prepared by mixing QSY21 succinimidyl ester (Invitrogen) with spermine in DMSO/DMF solution. 1.0 mg (1.3 μmol) QSY21 succinimidyl ester was dissolved in 50 μL of anhydrous DMSO. To this solution was added 1.25 mg of spermine (6.2 μmol) in 25 μL DMF, and the reaction mixture was incubated at room temperature in the dark for 4 hours. The resulting solution was used directly in subsequent quenching assays (typically at 1:400 dilution) without purification. ESI-MS: m/z 866.40 (calculated for $C_{51}H_{60}N_7O_4S$ $[M^+]$ m/z 866.44). The starting material, QSY21 succinimidyl ester, was detected on the mass spectrometer (m/z 779.46, calculated $[M^+]$ m/z 779.25) at the beginning of the reaction, but this peak disappeared at the end of the reaction.

Fluorescently labeled LDL particles (DiI-LDL and DiD-LDL) were prepared by incubating LDL solution with DiI or DiD overnight, followed by dialysis against PBS pH 7.4 to remove excess dye molecule. Low density lipoprotein (LDL) at a concentration of 7.6 mg/mL was obtained from Prof. Monty Krieger (MIT). 200 μL of this solution was mixed first with an equal volume of PBS pH 7.4, and then 2.0 μL of DiI (30 mg/mL) in DMSO. The mixture was incubated at 37 °C for 4 hours. At the end of incubation, the solution was centrifuged at 12,000 g for 10 min to pellet aggregates, and the supernatant was further filtered with a 0.22 μm centrifugal filter (Amicon). The clarified DiI-LDL solution was dialyzed against PBS pH 7.4 and stored in the dark at 4 °C. The same protocol was used for preparing DiD-LDL.

The AP-WT LDLR construct was created by PCR-amplifying the AP-WT LDLR gene (*33*) with flanking restriction sites, *Nhe* I and *Blp* I, and then ligating into digested pcDNA4-mCherry-LplA vector (a gift from Dr. Katharine A. White (*34*)). The resulting construct places AP-LDLR expression under control of an inducible promoter, featuring two tetracycline operator 2 sites located downstream of a CMV immediate-early promoter. All AP-LDLR mutants were created by site-directed mutagenesis using QuikChange (Stratagene).

HA-WT LDLR was created by first inserting the HA epitope tag into AP-WT LDLR (AP-HA-WT LDLR) via inverse PCR (*35*), and then removing the AP tag via QuikChange (Stratagene). YFP-tagged histone H2B (H2B-YFP) has been used as a nuclear marker of transfection (*33*). BirA-KDEL in pDisplay vector (*33*) originally contained an HA epitope tag immediately downstream of its Ig κ-chain leader sequence. In the present study, this epitope tag was removed by QuikChange. Transferrin receptor-EGFP fusion construct (TfnR-EGFP) was a gift from Prof. Mark Howarth (University of Oxford).

Fluorescence measurement

A Tecan Sapphire platereader and 96-well polystyrene microplates with clear flat-bottom (Greiner) were used to measure the absorption spectrum of QSY21-spermine and the fluorescence emission spectrum of mSA-Alexa568 and resorufin. For the absorption measurement, 100 μL of 160 μM QSY21-spermine solution in PBS pH 7.4 was prepared. For the fluorescence measurement, 100 μL of 80 nM mSA-Alexa568 (or 100 μM resorufin) in PBS pH 7.4 was prepared. The solution was excited at 530 nm, and its fluorescence spectrum was recorded from 560 nm to 700 nm. The same excitation and emission range was used to measure the quenched fluorescence from a 100 μL solution of 80 nM mSA-Alexa568 (or 100 μM resorufin) in the presence of 40 μM (or 160 μM) QSY21-spermine. Background absorption and fluorescence were measured for PBS pH 7.4, and subtracted from the above measurements. Each measurement was performed in triplicate. To quantify the extent of fluorescence quenching, the area under the Alexa568 emission curve was integrated for both quenched (A_Q) and unquenched (A_0) cases. The quenching efficiency was reported as $(A_0 - A_Q) / A_0$.

Testing membrane permeability of QSY21-spermine

T-REx CHO A7 cells at ~80% confluence were incubated with 100 µM resorufin acetate (a gift from Daniel S. Liu) in DPBS at room temperature for 5 min. Thereafter, cells were washed with ice-cold DPBS five times, for 3 min each time, to remove excess dye. Immediately after taking a fluorescence image, the buffer was aspirated and 160 µM QSY21-spermine in ice-cold DPBS was added onto the monolayer of cells. Another fluorescence image of the same field of view was taken 1 min after adding the quencher. Quantitation of fluorescence quenching was performed as described for receptor internalization analysis.

Mammalian cell culture and transfection.

HeLa and T-REx CHO-K1 cells were cultured in growth medium (DMEM supplemented with 10% FBS, 70 mg/L L-proline and 220 mg/L sodium pyruvate) at 37 °C under 5% (v/v) CO_2. For the T-REx CHO A7 stable cell line, the medium was further supplemented with 10 µg/mL blasticidin. Cells were transfected using either Nucleofector (Amaxa) or Lipofectamine 2000 (Invitrogen).

The protocol for nucleofection is as follows. 1 million T-REx CHO A7 cells suspended in 100 µL nucleofection solution (120 mM KCl, 10 mM KH_2PO_4, 2 mM EGTA, 25 mM HEPES, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 5 mM oxidized glutathione, 2 mM ATP, pH 7.5) were mixed with 1.8 – 2.0 µg of DNA and nucleofected with the U-023 program. Cells were rescued with pre-warmed growth medium and plated onto the center of homemade 35 mm imaging dishes with glass (No. 1.5) bottoms at a density of 50,000 – 100,000 cells per dish.

The protocol for Lipofectamine 2000 transfection is as follows. 40,000 – 80,000 T-REx CHO A7 cells were plated onto the center of homemade 35 mm imaging dishes with glass (No. 1.5) bottoms. 12 – 24 hours after plating, these cells were transfected with 0.3 µg of DNA mixed with 1 µL Lipofectamine 2000 in 250 µL DMEM, following manufacturer's protocol. 4 – 5 hours after transfection, the medium was aspirated and complete culture medium was added to cells in the imaging dishes.

Unless stated otherwise, if cells were transfected with pcDNA4 constructs and BirA-KDEL, 10 μM biotin and 50 ng/mL tetracycline were supplied to the growth medium. Typically, cells were labeled and imaged 24 – 36 hours after transfection, at 70 – 90% confluence.

Fluorescence microscopy and data analysis

Typically, cells were labeled with mSA-Alexa568 24 ~ 36 hour after transfection, at 70 – 90% confluence. To quench cell surface fluorescence signal, QSY21-spermine solution at 40 μM in ice-cold DPBS was added into the imaging dish.

Images were acquired with a fully motorized epi-fluorescence microscope (Axio Observer.Z1, Carl Zeiss) using a 40× oil-immersion lens (Carl Zeiss). The system was controlled using Slidebook 4.2 software (Intelligent Imaging Innovations). Digital images (16 bit) were obtained with a cooled EMCCD camera (QuantEM:512SC, Photometrics) with exposure times between 200 msec and 500 msec. A xenon lamp in a Lambda DG4 device (Sutter Instruments) was used to excite all fluorophores, and the filter set information is listed in Table 5-1. Nomarski-type differential interference contrast (DIC) was used for transmitted-light imaging of all samples.

Table 5-1. Filter set information

Fluorophore	Excitation	Emission	Dichroic
YFP and EGFP	493/16	525/30	502
Alexa568, DiI, resorufin	570/20	605/30	585
Alexa647 and DiD	630/20	680/30	660

Image analysis was performed using Slidebook 4.2 software. To quantify internalization, a mask for each labeled cell was individually created based on AlexaFluor568 signal in the pre-quench image, and then these masks were copied to the corresponding post-quench (+QSY) image. Occasionally, stage drift caused a shift of a few pixels between images taken before and after quenching, and this was corrected manually by shifting the masks. After subtracting background in the AlexaFluor568 channel, the percent of labeled receptor internalized was calculated for each mask by taking the ratio of fluorescence intensities before and after quenching.

The frequency distributions of internalization extent for two data sets were compared using the two-sample Kolmogorov-Smirnov test, and the *p*-values were reported.

Cell surface labeling of AP-receptors with mSA-Alexa568

T-REx CHO A7 cells expressing AP-LDLR (wild-type or mutant) and BirA were labeled at 4 °C as previously described (*33*). Briefly, cells were washed three times with ice-cold DPBS, and then incubated with 80 nM mSA-AlexaFluor568 in DPBS containing 1% (w/v) BSA at 4 °C for 10 min. At the end of incubation, excess labeling reagent was removed by gently washing cells three times with ice-cold DPBS. Thereafter, cells were either imaged directly or incubated at 37 °C for internalization experiments.

Control expression level of AP-LDLR in T-REx CHO A7 stable cell line

CHO ldlA-7 cells stably expressing the tetracycline repressor protein (T-REx CHO A7) were obtained by transfection with pcDNA6/TR (Invitrogen) using Lipofectamine 2000 (Invitrogen). The cells were then grown in 10 μg/mL blasticidin in growth medium for 10 days. Colonies of surviving cells were isolated by pipetting and expanded in the same blasticidin-containing growth medium. Candidate cell lines were tested for tetracycline-induced expression of pcDNA4-mCherry, using mCherry fluorescence as readout of expression level.

By measuring the uptake of fluorescently labeled LDL particles in cells transiently expressing AP-LDLR, we determined that 50 ng/mL tetracycline induces AP-LDLR expression to a level comparable to endogenous LDLR in wild-type CHO-K1 cells and HeLa cells (Figure 5-22). T-REx CHO A7 cells were transfected with AP-WT LDLR and BirA targeted to the endoplasmic reticulum. Cells were cultured in tetracycline-free growth medium overnight. 12 – 16 hours after transfection, the growth medium was supplemented with various concentrations of tetracycline (0, 20, 50, 100, 200, 500 ng/mL) to induce AP-LDLR expression. 24 – 36 hours after induction, cells were incubated with 100 μg/mL DiI-LDL for 1 hour at 37 °C. For comparison, untransfected HeLa and wild-type CHO-K1 cells expressing endogenous LDLR were analyzed in the same manner.

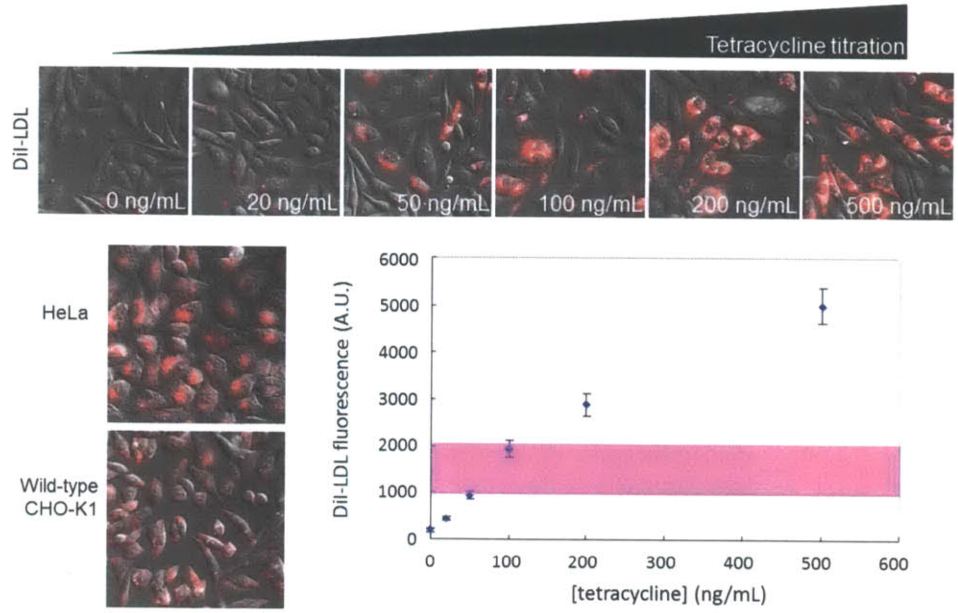


Figure 5-22. Titration of AP-LDLR expression level in T-REx CHO A7 cells.
Cells transiently transfected with AP-LDLR (wild-type) plasmid were cultured in growth medium containing various concentrations of tetracycline for 24 – 36 hours to induce AP-LDLR expression. Top: representative images of cells after incubation with 100 μg/mL DiI-LDL at 37 °C for 1 hour. Images show DiI fluorescence overlaid with DIC. Untransfected cells did not take up DiI-LDL. Bottom left: for comparison, untransfected HeLa and wild-type CHO-K1 cells expressing endogenous LDLR were labeled with DiI-LDL and imaged in the same manner. Bottom right: plot of whole-cell DiI-LDL signal (mean ± s.e.m.) against tetracycline concentration. The pink region denotes the DiI signal level in wild-type CHO-K1 and HeLa cells. 50 ng/mL tetracycline was used for subsequent experiments in this work.

Using the same ligand uptake assay, we also verified that receptors labeled with mSA-Alexa568 bind and internalize LDL to the same extent as unlabeled receptors, suggesting that the mSA label does not affect these functions (Figure 5-23). T-REx CHO A7 cells were transfected with AP-WT LDLR and BirA as described above. 24 – 36 hours after transfection, cells were first surface-labeled with 200 ng/mL mSA-Alexa568 for 10 min and then incubated with 10 μg/mL DiD-LDL for 1 hour at 4 °C. For comparison, mSA-AlexaFluor568 labeling was omitted in the control group. To assess LDL uptake, these cells were further incubated with 10 μg/mL unlabeled LDL at 37 °C for 1 hour and then re-imaged in the same manner. Separate experiments showed that the 10-min labeling protocol with mSA-AlexaFluor568 saturates cell surface AP-LDLR with both biotin and mSA; i.e., labeling was complete (data not shown).

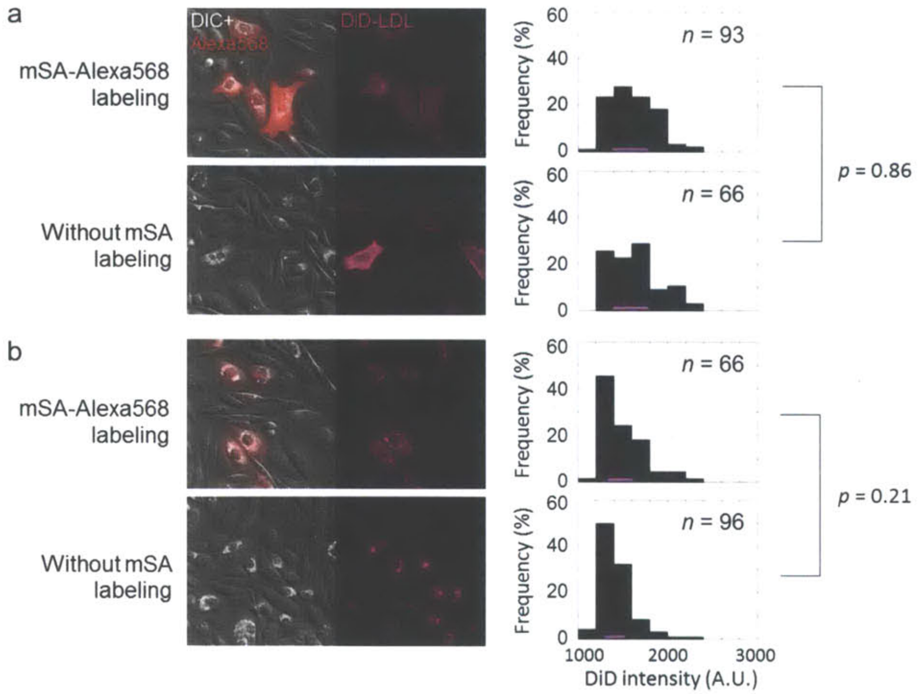


Figure 5-23. The mSA label on AP-LDLR does not affect LDL binding and internalization. T-REx CHO A7 cells transfected with AP-LDLR (wild-type) and BirA were labeled with mSA-Alexa568 at 4 °C for 10 min, incubated with 10 μg/mL DiD-LDL at 4 °C for 1 hour, and then imaged (a, top panel). Thereafter, cells were further incubated with 10 μg/mL unlabeled LDL at 37 °C for 1 hour, and then re-imaged (b, top panel). The bottom panels show the same analyses without mSA-Alexa568 labeling. The distributions of whole-cell DiD-LDL intensity for single cells are plotted in histograms. Pink lines indicate interquartile ranges (25–75%), and *p*-values are calculated using the Kolmogorov-Smirnov test. There is no significant difference between labeled and unlabeled receptors in LDL binding and uptake.

Receptor internalization assay

T-REx CHO A7 cells were transfected with BirA, H2B-YFP transfection marker, and wild-type and/or mutant LDLR (differentially tagged with either AP or HA, as indicated in each figure). 24 – 36 hours after transfection, cells were surface labeled with mSA-AlexaFluor568, and then these cells were rinsed twice with DPBS pre-warmed to 37 °C, before incubating at 37 °C for 5 min. At the end of the incubation, cells were chilled with ice-cold DPBS and imaged on the microscope. To assess the effect of LDL on receptor internalization, 100 μg/mL LDL was added to the buffer during mSA-Alexa568 labeling, and 10 μg/mL LDL was added to the buffer during the subsequent incubation.

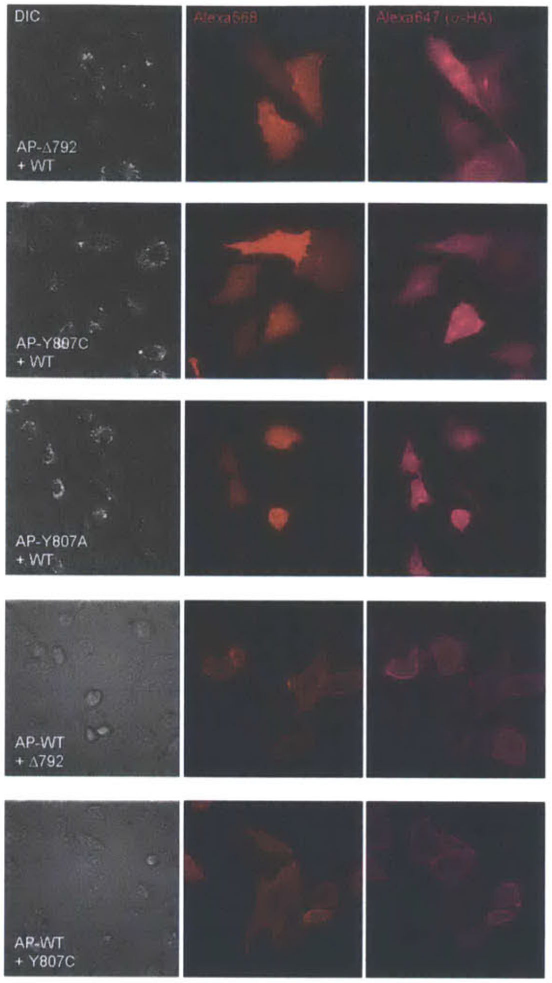


Figure 5-24. Confirmation of LDLR co-expression by epi-fluorescence imaging.
After labeling with mSA-AlexaFluor568, cells were fixed and stained with primary anti-HA antibody followed by AlexaFluor647-conjugated secondary antibody to detect co-expressed (non-AP-tagged) receptor. Representative images are shown for each combination. By counting fluorescently labeled cells, the efficiency of HA-WT co-expression with the AP-mutant was calculated to be: 94% for AP-Δ792 (n = 81), 89% for AP-Y807C (n = 82), and 92% for AP-Y807A (n = 65). The efficiency of AP-WT co-expression with HA-tagged mutant LDLR was calculated to be: 94% for HA-Δ792 (n = 62), and 95% for HA-Y807C (n = 74).

Prior to imaging, buffer was aspirated from the imaging dish and ~500 μL of 40 μM QSY21-spermine solution in ice-cold DPBS was added to cover the monolayer of cells. Within 1 min after quenching, fluorescence and DIC images were taken again for cells in the same field of view. We note that it was necessary to perform our live imaging assay for a 5-min incubation period only. Longer incubations gave significant internalization of the LDLR mutants, probably via non-specific pathways.

Evaluation of LDLR co-expression efficiency using immunostaining

T-REx CHO A7 cells were transfected with HA-tagged and AP-tagged LDLRs (as indicated in Figure 5-24), and BirA as described above. 30 hours after transfection, cells were washed with DPBS and labeled with mSA-Alexa568 at 4 °C for 15 min. After washing off excess labeling reagent, cells were fixed with 4% (v/v) formaldehyde in PBS for 10 min at room temperature. After washing three times with PBS, cells were permeabilized and blocked with blocking buffer (0.1% v/v Tween 20 + 3% w/v BSA in PBS) overnight at 4 °C. For immunostaining, cells were probed with rabbit α-HA (0.6 μg/mL in blocking buffer) for 1 hour, washed 4× 5 min with blocking buffer, labeled with goat α-rabbit conjugated with Alexa647 (4 μg/mL in blocking buffer) for 1 hour, and then washed again 4× 5 min with blocking buffer at room temperature.

Co-immunoprecipitation analysis of LDLR oligomerization

$3 - 4 \times 10^6$ T-REx CHO A7 cells were co-transfected with HA-WT LDLR, BirA, and either AP-WT LDLR or AP-Δ792 LDLR. 24 – 36 hours after transfection, cells were treated with lysis buffer (125 mM Tris-maleate pH 6, 2 mM $CaCl_2$, 160 mM NaCl, 1% (v/v) Triton X-100, and 1 mM PMSF) and then centrifuged to yield a clear lysate supernatant. Biotinylated proteins were pulled down from this supernatant using streptavidin magnetic particles (Roche Applied Science), pre-blocked using 0.1% (w/v) bovine serum albumin. The supernatant was collected and designated as the "unbound" fraction. The bound fraction was dissociated from the beads by heat denaturation, and designated as the "eluted" fraction. Lysate of cells without transfection was used as a negative control. Lysate of cells transfected with only AP-WT LDLR was used as a positive control for the SA-HRP blot, and a negative control for the anti-HA blot.

These fractions, as well as the whole lysate, were analyzed by both SA-HRP blot and anti-HA western blot. After transferring the protein bands from SDS-PAGE gel to nitrocellulose membrane, the membrane was blocked with blocking buffer (3% w/v BSA in PBS, with 0.1% v/v Tween-20) at 4 °C overnight. For the SA-HRP blot, the membrane was incubated with 0.3 μg/mL SA-HRP (Thermo Scientific) at room temperature for 30 min, and then rinsed with blocking buffer 4× 5 min. For the anti-HA western blot, the membrane was first incubated with 0.8 μg/mL rabbit anti-HA primary antibody (Invitrogen) for 1 hour, then washed with blocking buffer 4× 5 min, then incubated with goat anti-rabbit secondary antibody-HRP conjugate (Bio-Rad Laboratories, 1:5000 dilution in blocking buffer) for 30 min. The membrane was finally washed 4× 5 min with blocking buffer. The blot was developed with SuperSignal West Pico Reagents (Thermo Scientific) following the manufacturer's protocol, and imaged with a digital camera.

References

1 Terrillon, S. & Bouvier, M. Roles of G-protein-coupled receptor dimerization - From ontogeny to signalling regulation. *EMBO Reports* **5**, 30-34 (2004).

2 Liu, A. P., Aguet, F., Danuser, G. & Schmid, S. L. Local clustering of transferrin receptors promotes clathrin-coated pit initiation. *The Journal of cell biology* **191**, 1381-1393 (2010).

3 Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* **11**, 567-573 (2004).

4 Brown, M. S., Anderson, R. G. & Goldstein, J. L. Recycling receptors: the round-trip itinerary of migrant membrane proteins. *Cell* **32**, 663-667 (1983).

5 Brown, M. S. & Goldstein, J. L. A Receptor-Mediated Pathway for Cholesterol Homeostasis. *Science* **232**, 34-47 (1986).

6 Chen, W. J., Goldstein, J. L. & Brown, M. S. NPXY, A Sequence Often Found in Cytoplasmic Tails, Is Required for Coated Pit-Mediated Internalization of the Low Density Lipoprotein Receptor. *J.Biol.Chem.* **265**, 3116-3123 (1990).

7 Rudenko, G. *et al.* Structure of the LDL receptor extracellular domain at endosomal pH. *Science* **298**, 2353-2358 (2002).

8 Davis, C. G. *et al.* The J.D. mutation in familial hypercholesterolemia: amino acid substitution in cytoplasmic domain impedes internalization of LDL receptors. *Cell* **45**, 15-24 (1986).

9 Davis, C. G., van Driel, I. R., Russell, D. W., Brown, M. S. & Goldstein, J. L. The low density lipoprotein receptor. Identification of amino acids in cytoplasmic domain required for rapid endocytosis. *The Journal of biological chemistry* **262**, 4075-4082 (1987).

10 Chen, W. J., Goldstein, J. L. & Brown, M. S. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *J Biol Chem* **265**, 3116-3123 (1990).

11 Ulbrich, M. H. & Isacoff, E. Y. Subunit counting in membrane-bound proteins. *Nat Methods* **4**, 319-321 (2007).

12 Shi, Y. G. & Massague, J. Mechanisms of TGF-á signaling from cell membrane to the nucleus. *Cell* **113**, 685-700 (2003).

13 Schlessinger, J. Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. *Cell* **110**, 669-672 (2002).

14 Sorkina, T., Doolen, S., Galperin, E., Zahniser, N. R. & Sorkin, A. Oligomerization of dopamine transporters visualized in living cells by fluorescence resonance energy transfer microscopy. *J.Biol.Chem.* **278**, 28274-28283 (2003).

15 Makarova, A. *et al.* The LDL receptor-related protein can form homo-dimers in neuronal cells. *Neurosci.Lett.* **442**, 91-95 (2008).

16 Bouvier, M. Oligomerization of G-protein-coupled transmitter receptors. *Nat.Rev.Neurosci.* **2**, 274-286 (2001).

17 van Driel, I. R., Davis, C. G., Goldstein, J. L. & Brown, M. S. Self-Association of the Low Density Lipoprotein Receptor Mediated by the Cytoplasmic Domain. *J.Biol.Chem.* **262**, 16127-16134 (1987).

18 Coward, P. *et al.* Controlling signaling with a specifically designed Gi-coupled receptor *Proc.Natl.Acad.Sci.U.S.A.* **95**, 352-357 (1998).

19 Sartania, N., Appelbe, S., Pediani, J. D. & Milligan, G. Agonist occupancy of a single monomeric element is sufficient to cause internalization of the dimeric á2-adrenoceptor *Cell.Signal.* **19**, 1928-1938 (2007).

20 Ulbrich, M. H. & Isacoff, E. Y. Subunit counting in membrane-bound proteins. *Nat.Methods* **4**, 319-321 (2007).

21 Dorsch, S., Klotz, K. N., Engelhardt, S., Lohse, M. J. & Bunemann, M. Analysis of receptor oligomerization by FRAP microscopy. *Nat.Methods* **6**, 225-230 (2009).

22 Howarth, M. & Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat.Protocols* **3**, 534-545 (2008).

23 Howarth, M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat.Methods* **5**, 397-399 (2008).

24 Sege, R. D., Kozarsky, K., Nelson, D. L. & Krieger, M. Expression and Regulation of Human Low-Density Lipoprotein Receptors in Chinese Hamster Ovary Cells. *Nature* **307**, 742-745 (1984).

25 Basu, S. K., Goldstein, J. L., Anderson, R. G. W. & Brown, M. S. Monensin Interrupts the Recycling of Low Density Lipoprotein Receptors in Human Fibroblasts. *Cell* **24**, 493-502 (1981).

26 Heuser, J. E. & Anderson, R. G. W. Hypertonic Media Inhibit Receptor-Mediated Endocytosis by Blocking Clathrin-Coated Pit Formation. *J.Cell Biol.* **108**, 389-400 (1989).

27 Iacopetta, B. J., Rothenberger, S. & Kuhn, L. C. A Role for the Cytoplasmic Domain in Transferrin Receptor Sorting and Coated Pit Formation During Endocytosis. *Cell* **54**, 485-489 (1988).

28 Jeon, H. & Blacklow, S. C. Structure and physiologic function of the low-density lipoprotein receptor. *Ann.Rev.Biochem.* **74**, 535-562 (2005).

29 Brown, M. S., Anderson, R. G. W. & Goldstein, J. L. Recycling Receptors: the Round-Trip Itinerary of Migrant Membrane-Proteins. *Cell* **32**, 663-667 (1983).

30 Wu, S. C. & Wong, S. L. Development of an enzymatic method for site-specific incorporation of desthiobiotin to recombinant proteins in vitro. *Anal Biochem* **331**, 340-348 (2004).

31 Hirsch, J. D. *et al.* Easily reversible desthiobiotin binding to streptavidin, avidin, and other biotin-binding proteins: uses for protein labeling, detection, and isolation. *Anal Biochem* **308**, 343-357 (2002).

32 Howarth, M. & Ting, A. Y. Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. *Nat Protoc* **3**, 534-545 (2008).

33 Howarth, M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat Methods* **5**, 397-399 (2008).

34 Uttamapinant, C. *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A* **107**, 10914-10919 (2010).

35 Gama, L. & Breitwieser, G. E. Generation of epitope-tagged proteins by inverse PCR mutagenesis. *Biotechniques* **26**, 814-816 (1999).

Chapter 6 Single molecule imaging analysis of receptor internalization

Except for Section 6.3, work described in this chapter is unpublished. Section 6.3 is adapted from publication: J. D. Cohen, P. Zou, and A. Y. Ting. Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation. *ChemBioChem* 2012, 13, 888-894. Dr. Justin Cohen developed site-specific aryl hydrazine ligation technique.

6.1 Introduction

This chapter aims to resolve the detailed mechanism of receptor-mediated endocytosis at the single-molecule level, using recently developed site-specific protein labeling techniques in conjunction with fluorescence microscopy. Specifically, two models have been proposed in the literature regarding how receptors are targeted to the clathrin-coated pits (Figure 6-1) (*1*). It has been difficult to distinguish these models in experiments, and it is not known whether the mechanism is different for ligand-free receptors and ligand-bound ones (*2*). We used the low density lipoprotein receptor (LDLR) as a model system.

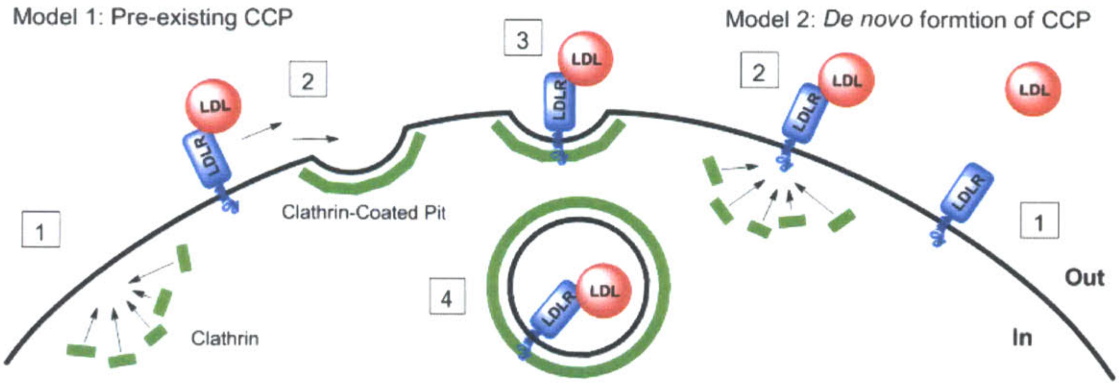


Figure 6-1. Two proposed models for receptor entrapment into the CCPs.
In the first model, CCP forms prior to ligand binding, and the entrapment is achieved by lateral diffusion of LDL receptor into the pit. In the second model, LDL binding induces the formation of CCP by recruiting clathrin and adaptor proteins.

This study is facilitated by novel site-specific protein labeling techniques utilizing the *E. coli* biotin ligase, BirA. In the previous chapter, we demonstrated that streptavidin labeled with organic fluorophores are useful for studying receptor trafficking at the ensemble level. Alternatively, it can be used to decorate quantum dots (QDs) for applications in single-molecule fluorescence imaging (*3,4*). Because of their intense brightness and resistance to photobleaching, QDs are ideal probes for single-molecule fluorescence imaging. Although streptavidin-functionalized QDs are commercially available, they are multivalent and have a large size (20 ~ 30 nm) due to their thick passivation layer. This has become a problem for their biological applications because these large particles may affect the function of targeted proteins and have limited access to crowded cellular environments. Therefore, it is crucial to develop novel coating methods to reduce QD size, while simultaneously maintaining high QD stability, solubility, and passivation.

Previously published small molecule QD coating ligands include mercaptoacetic acid (*5*), cysteine (*6*), peptides bearing cysteine residues (*7*), dihydrolipoic acid (DHLA) (*5*) and poly(ethylene glycol)-appended DHLA (*8*). Although QDs coated with these ligands are small in size, they often suffer from instability in the buffer during their storage. This problem calls for the design and synthesis of new coating. In this chapter, we have engineered quantum dots for single-molecule fluorescence imaging and using them to examine two proposed mechanisms of clathrin-coated pit initiation.

6.2 Engineering small QDs for studying receptor-mediated endocytosis

Establishing a model system

As shown in Figure 6-1, two models have been hypothesized to account for the entrapment of a receptor into the CCP:

1) "Pre-existing pit" model: upon ligand binding, the receptor is targeted to a pre-existing CCP on the cell surface;
2) "*De novo* pit formation" model: a ligand-bound receptor recruits clathrin and adaptor proteins, forming a new CCP beneath it.

The key to solving this problem lies in determining the temporal order of the CCP formation and the reduction in receptor mobility. This highly dynamic process is not amenable to traditional methods such as biochemistry and immunocytochemistry, because they often provide a static, ensemble-averaged picture of the cell; instead, single-molecule fluorescence microscopy allows us to examine the dynamic behavior of individual receptors and CCPs in live cells, thus providing the answer to the above question.

In order to specifically label LDL receptor in mammalian cells, the AP-tag was fused to the *N*-terminus of the LDL receptor, immediately after its leader sequence (AP-LDLR). Biotinylated LDL receptors on the cell surface could be imaged by targeting them with red-emitting streptavidin-derivatized QDs. The experiment also requires labeling CCPs. In previous research, this is achieved by fusing a fluorescent protein to one component in the endocytic machinery, such as the clathrin light chain or the σ2 subunit of the adaptor protein AP2 (*2*). The construct used in my experiment has σ2 fused with Venus (σ2-Venus, a gift from Prof. Xiaowei Zhuang at Harvard University). Because multiple copies of σ2-Venus are concentrated in a CCP, whose diameter (~100 nm) is smaller than the wavelength of the emitted light, each CCP appears as a strongly fluorescent, diffraction-limited spot when visualized under a microscope (Figure 6-2).

Figure 6-2 shows the images from a proof-of-principle experiment, in which I used a total internal reflection fluorescence (TIRF) microscope to image both the CCP and LDL receptor on the cell surface. Internalization events were observed in real-time – as a receptor moves into the cell, it leaves the evanescent field localized near the plasma membrane, and the QD attached to it becomes dark. In this experiment, the two proposed receptor entrapment models can be distinguished by the order of appearance and disappearance of Venus and QD signals. The preliminary data seem to favor the pre-existing CCP model.

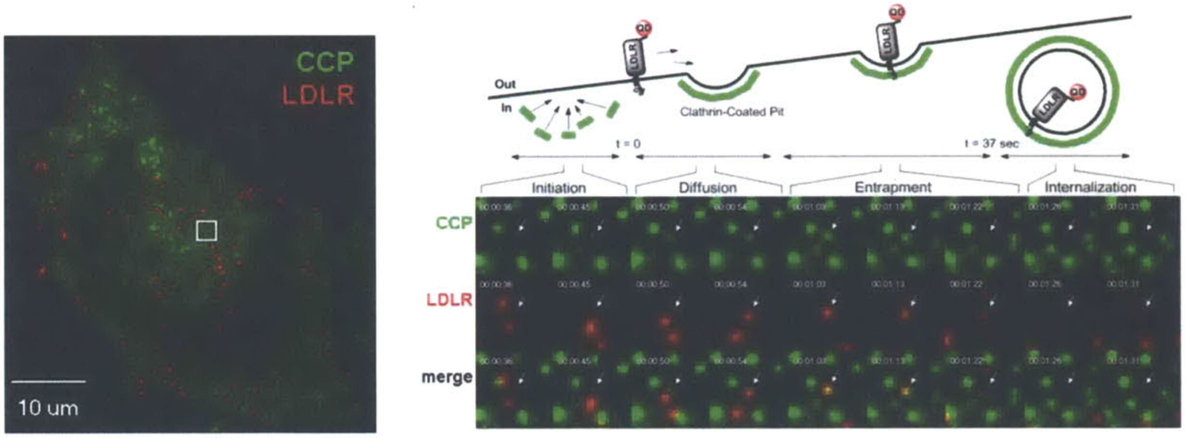


Figure 6-2. Single-molecule fluorescence imaging experiment with commercial QDs.
Left: TIRF image of the basal membrane of a Chinese hamster ovary (CHO-K1) cell. These cells were transiently transfected with AP-LDLR, BirA and σ2-Venus. CCPs are labeled with Venus (green) and the LDL receptors are specifically targeted with commercial QDs (red) via biotin ligase. Right: "pre-existing pit" model (top) and the tile view of the square region in (A) (bottom). Each column represents a different time point, and the arrows in each panel locate an internalization event, which supports the "pre-existing pit" model. Images are matched into the four phases of the model.

The above co-localization study was done with commercially available QDs which are known to be large in size and multivalent. Our next goal is to engineer small, monovalent QDs to further investigate this problem. In order to minimize the effect of protein overexpression, we have also developed a COS-7 cell line that stably expresses σ2-Venus.

Engineer small, monovalent QD reagents

QDs used in biological applications must fulfill the following requirements: small size, high targeting affinity, fast labeling kinetics, reduced tendency to crosslinking the target, complete passivation, and biocompatibility. Reduced-size QDs have been made by our group and others (*6*), but they have been plagued by problems of short shelf-life, poor water solubility, and non-specific binding to cells. Based on the properties of existing QD ligand (*6,9*), we designed a new coating structure with the following rationale:

1) Poly(ethylene glycol) (PEG) is commonly used to passivate macromolecules, inorganic particles and surfaces. It improves the solubility of QDs and reduces non-specific adsorption of QDs to cells. While longer PEG linkers usually provide a better passivation, shorter ones facilitate functionalization of QDs with proteins.
2) One end of the PEG segment is conjugated to dihydrolipoamine, whose thiol groups coordinate to the zinc sulfide layer on the QD surface.
3) The other end of the PEG terminates as a carboxyl group, whose negative charge at neutral pH both improves water solubility and facilitates electrophoretic purification of streptavidin-conjugated QDs. It also provides a functional handle for amide coupling.

The synthetic scheme of this ligand is shown in Figure 6-3. Essentially, it is achieved by coupling lipoamine to one end of poly(ethylene glycol) diacid (PEG diacid) through an amide linkage.

The quality of this ligand was subsequently determined by testing ligand-coated QDs at three levels: 1) biocompatibility, 2) targeting specificity after protein conjugation, and 3) potential for generating monovalent QDs.

Level I: biocompatibility. Starting with QDs in its growth solution (courtesy of Wen Liu and the Bawendi lab), we performed a cap exchange reaction to replace the hydrophobic trioctyl phosphine oxide (TOPO) on the QDs surface with my purified ligand. The resulting QDs coated with this new ligand were soluble in phosphate buffered saline (PBS) solution and have a hydrodynamic diameter of ~12 nm, as determined by dynamic light scattering. This size is significantly smaller than commercial QDs. To test whether these QDs were properly passivated, we incubated them with cells in the culturing medium and observed minimal stickiness.

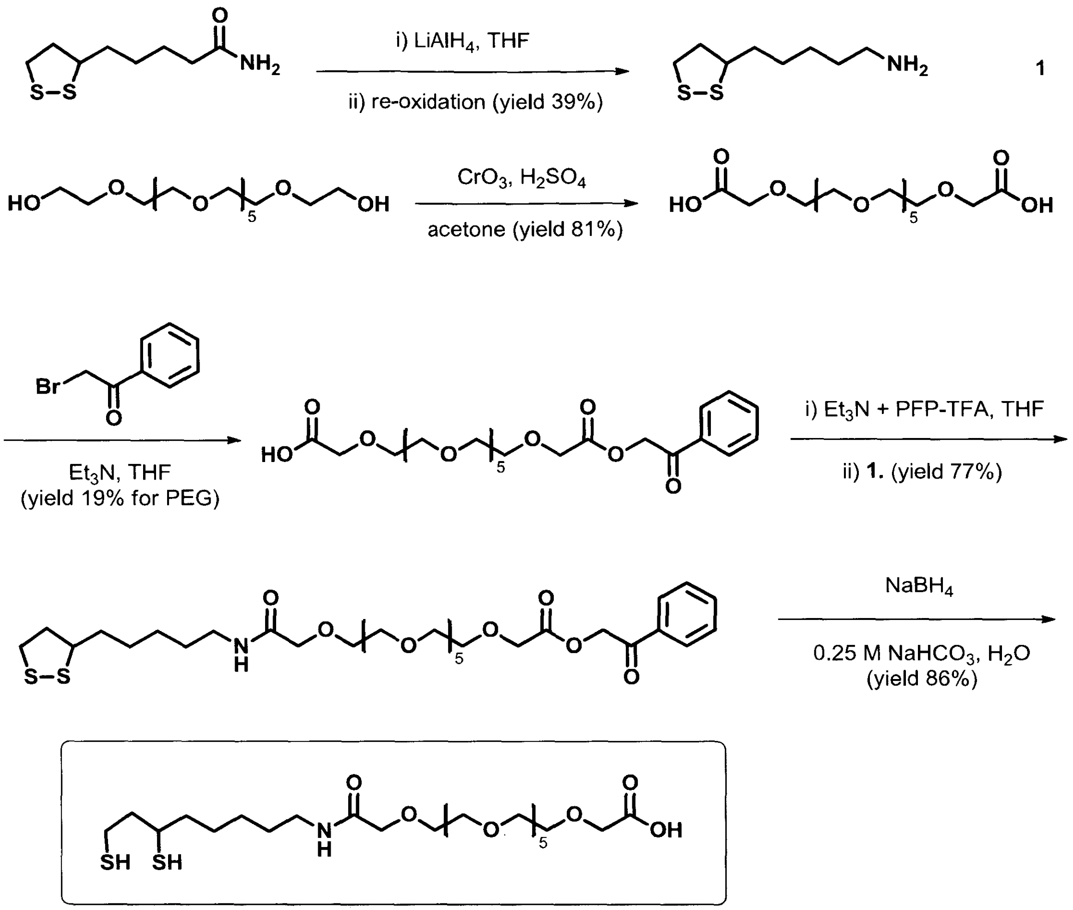

i) LiAlH$_4$, THF
ii) re-oxidation (yield 39%)
NH$_2$
S–S
1
HO
OH
5
CrO$_3$, H$_2$SO$_4$
acetone (yield 81%)
Br
Et$_3$N, THF
(yield 19% for PEG)
i) Et$_3$N + PFP-TFA, THF
ii) 1. (yield 77%)
NaBH$_4$
0.25 M NaHCO$_3$, H$_2$O
(yield 86%)
SH SH
N
H

Figure 6-3. The synthetic scheme of novel QD coating ligand.
Lipoamine was prepared by reducing the lipoamide with lithium aluminum hydride, and PEG diacid was prepared by oxidizing PEG with Jones reagent. In both cases, the starting material was commercially available, and the procedures were documented in the literature. The final product required conjugating PEG diacid and lipoamine in strictly one-to-one ratio, and this was facilitated by using a phenacyl group to protect one of the two carboxyl groups in PEG diacid. In this protecting step, a sub-stoichiometric amount of bromoacetophenone (a.k.a. phenacyl bromide) was reacted with PEG diacid, and the desired mono-protected product was purified from a mixture of unprotected and di-protected by-products. After conjugating lipoamine to phenacyl-protected PEG diacid, the protecting group was removed by reductive cleavage with sodium borohydride. In the same reduction reaction, the dithiolane ring opened to form two thiol groups, yielding the final product as viscous, colorless oil. The product was characterized by TLC, MS and NMR.

Level II: specific cellular targeting. Conjugating mSA to these reduced-size QDs was facilitated by the interaction between the His-tag of mSA and the zinc/cadmium sulfide coating of QDs (*10*). Here, shorter PEG segment in the ligand structure allows better accessibility of the

QD surface to the His-tag. The resulting mSA-QD conjugate were targeted to biotinylated LDL receptors on the cell surface. Again, minimal non-specific binding was observed (Figure 6-4).

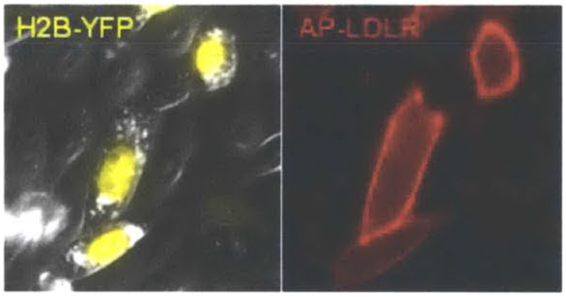


Figure 6-4. Specific targeting to biotinylated LDL receptor on the cell surface.
Confocal images of CHO-K1 cells transiently transfected with AP-LDLR, BirA and H2B-YFP as a transfection marker. Most transfected cells have YFP in the nucleus, and they are specifically labeled with mSA-QD.

Level III: control of monovalency. In order to make QDs truly useful for biological applications, it is desirable to use monovalent QDs to avoid crosslinking the target. To this end, QDs derivatized with only one copy of mSA should be separated from the mixture. Previously, this could be achieved by agarose gel electrophoresis (*9*). Due to the negative charge on the carboxyl group, QDs prepared with the novel ligand have high electrophoretic mobility, as shown in Figure 6-5. Conjugation to mSA significantly lowered this mobility, and QDs with different copy numbers of mSA were resolved into discrete bands on the agarose gel. Monovalent QD can be purified by isolating its band and extracting soluble content out of the gel, as previously described (*4*).

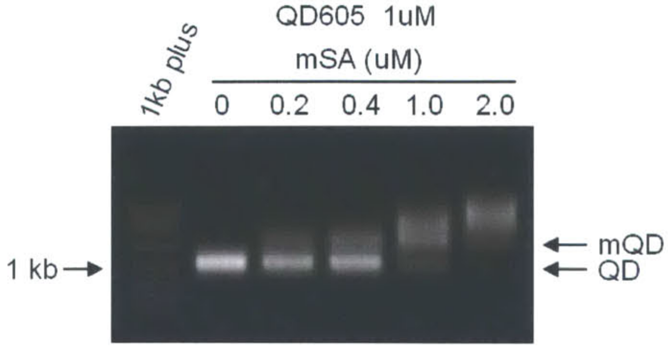


Figure 6-5. Separating mSA-conjugated QDs with new coating.
Agarose gel (1% in TAE buffer) electrophoresis of mSA-QD. Monovalent QD can be resolved from unconjugated QD and QDs with higher copy number of mSA.

The above three levels of analyses demonstrated the utility of this new QD coating ligand. However, the stability of these QDs decreases over the time – they begin to precipitate from the solution and become sticky to cells about 3 ~ 4 days after the preparation. This instability is thought to arise from the slow dissociation of dithiol ligand from the QD surface as well as the oxidation of thiols into disulfide groups, which has weak interaction with QDs. By comparing the stability of QDs coated with these negatively charged ligands (DHLA-PEG-COOH) and those with DHLA-PEG-OH (structure shown in Figure 6-3), we found that DHLA-PEG-OH coating significantly improves the stability of QDs, in spite of the structural similarity of these ligands. This difference in stability leads to the hypothesis that the dense negative charge created by DHLA-PEG-COOH on the QD surface tend to destabilize this coating layer, promoting ligand dissociation due to electrostatic interactions; neutral DHLA-PEG-OH ligand, on the other hand, does not have such problem.

6.3 Single-molecule imaging with phycoerythrin

For biotechnological applications requiring extreme fluorophore brightness, such as fluorescence activated cell sorting (FACS), phycoerythrin has been used as a much brighter alternative to organic dyes and a smaller and less expensive alternative to QDs. R-phycoerythrin (R-PE) is a 240 kDa protein of disk-shape, with a diameter of 11 nm and a thickness of 6 nm, containing 34 phycobilin-type chromophores. It is usually obtained by purification from red algae (*11*). With an extinction coeffient (ε) of 2.0×10^6 $M^{-1}cm^{-1}$ at 566 nm and quantum yield (QY) of 0.85, it is more than 25 times brighter than AlexaFluor568 (ε = 91,300 $M^{-1}cm^{-1}$ at 568 nm; QY = 0.69), an organic fluorophore with similar emission spectrum.

R-PE has rarely been explored as a reagent for single molecule imaging. In one example, Darrel Irvine *et al.* used PE to label major histocompatibility complex on the surface of antigen presenting cells in order to count their copy number (*12*). We wished to explore other methods of targeting PE to specific cell surface proteins, and to image them at the single molecule level. Since PE can only be practically added to cells at low micromolar concentrations, its targeting requires a method with an extremely high second order rate constant. With its extremely fast kinetics and cell compatibility, the bis-aryl hydrazone conjugation is ideal for this application.

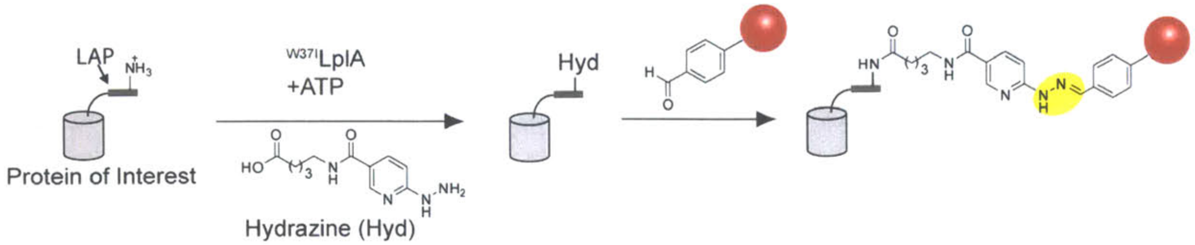


Figure 6-6. Scheme for using lipoic acid ligase (LplA) to conjugate a hydrazine functional group to LplA acceptor peptide (LAP) fusion proteins.
After ligation, the hydrazine moiety can be chemoselectively derivatized with aryl aldehyde-functionalized R-phycoerythrin (represented by red circle). The linkage is a hydrazone (highlighted in yellow) that is stabilized by conjugation to the two neighboring aromatic rings.

Based on the PRIME technique described in Chapter 4, we developed a method that site-specifically introduce aryl-hydrazine moiety to protein of interest. This method capitalizes on the W37I mutant of *E. coli* lipoate protein ligase, LplA, which catalyzes the covalent conjugation of an aryl hydrazine probe ("Hyd", as shown in Figure 6-6) with LplA acceptor peptide (LAP), with a kinetic constant k_{cat} of $0.021 \pm 0.003\ s^{-1}$ (*13*). Subsequent reaction with PE functionalized with aryl-aldehyde allows covalent conjugation of the fluorescent probe to the protein target (Figure 6-6). Bis-aryl hydrazone formation is one of the fastest bio-compatible ligation reactions, with a second order rate constant of $\sim 2000\ M^{-1}\ s^{-1}$ when aniline is used as a catalyst. In the absence of aniline, the reaction still proceeds at $\sim 1\ M^{-1}\ s^{-1}$ (*14*).

To first evaluate if we could conjugate phycoerythrin selectively, we transfected HEK293T cells with neurexin-1β, a cell surface protein, fused to LAP and a nuclear-YFP transfection marker. Cells expressing LAP- neurexin-1β was labeled with the Hyd probe using ^{W37I}LplA. After labeling, cells were washed and treated with 20 mM aniline and R-PE modified with 4-formylbenzamide (PE-Ald). After 45 min the cells were washed and imaged. As shown in Figure 6-7, clear labeling was observed in transfected cells. No labeling was seen in negative controls using wild-type LplA, with ATP omitted, or with an alanine mutation in LAP.

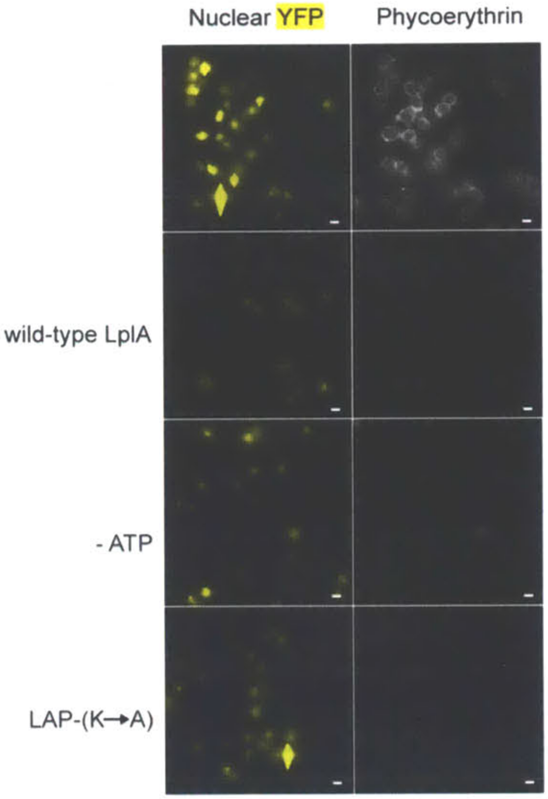


Figure 6-7. Protein labeling with phycoerythrin on the surface of living HEK293T cells. This figure was generated by Dr. Justin Cohen. Cells expressing LAP4.2-neurexin-1β were labeled with ^{W37I}LplA and Hyd substrate for 45 min, washed, then treated with 3 μM phycoerythrin-Ald in the presence of 10 mM aniline for 45 min, before confocal imaging. Negative controls are shown with wild-type LplA, ATP omitted, and an alanine mutation in LAP. Nuclear YFP was a co-transfection marker.

To perform single molecule imaging with PE, we next prepared COS-7 cells that express LAP-LDLR on the surface. We labeled the LDLR using our Hyd probe, followed by treatment with 20 mM aniline and PE-Ald. As shown in Figure 6-8, individual labeled LDLR molecules appeared as single diffraction-limited spots on the cell surface, imaged by total internal reflection fluorescence (TIRF) microscopy. To confirm that the labeled spots were indeed single receptors and not aggregates, we compared the intensity distribution of > 2900 spots on cells to individual PE molecules randomly distributed on glass slides, and observed similar distribution. The labeled receptors are also dynamic, as shown in time-lapse imaging experiments. The brightness of PE molecules offers high signal-to-background ratio that is unparalleled by organic fluorophores, and photobleaching is reduced because of the lower laser intensity required for illumination.

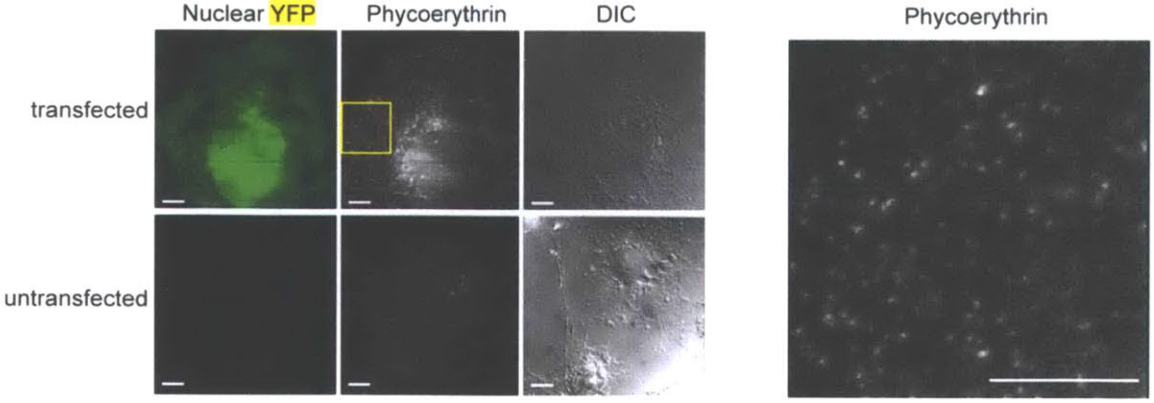

Nuclear YFP
Phycoerythrin
DIC
Phycoerythrin
transfected
untransfected

Figure 6-8. Single-molecule imaging with phycoerythrin on the surface of living COS7 cells.
Left: Cells expressing LAP-LDLR were labeled as in Figure 6-7, except that a lower concentration of phycoerythrin-Ald (0.3 μM) and 20 mM aniline was used and the cells were imaged in TIRF mode. Nuclear YFP is a transfection marker. Images of a transfected cell (top) and untransfected cell (bottom) are shown. Right: enlarged view of the boxed area on left showing individual phycoerythrin particles conjugated to LAP-LDLR on the COS7 cell surface. Scale bars, 10 μm.

6.4 Conclusion

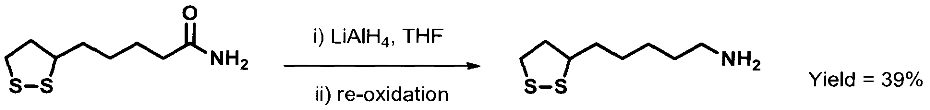
In summary, we have designed, synthesized and tested a novel QD coating ligand. Using this ligand, we have prepared mSA-conjugated QDs that can be specifically targeted to biotinylated receptors on the cell surface. Compared to previously published results, this new ligand fails to improve the stability of QDs after long-term storage; while the information learnt from this ligand provides the basis for our design of new ligands and new strategies of generating monovalent QDs. Divalent ligands such as dihydrolipoamine moiety has improved affinity toward the zinc sulfide shell of QDs, but the interaction is not strong enough to survive for longer than several days. Tetravalent ligands could potentially solve this problem. In general, thiol-based ligands tend to get oxidized by O_2 over the time, which is another factor for gradual dissociation from the shell. In future designs, it would be desirable to avoid using thiols as the chelating moiety. For instance, imidazole also shows strong affinity toward zinc sulfide shell, and it is resistant to oxidation.

As an alternative probe to QD, R-PE has also been employed as a probe for single-molecule fluorescence imaging. As a fluorescent protein, R-PE is naturally biocompatible with the cellular environment, and exhibited minimal stickiness. Our recently developed aryl-hydrazine ligation technique and the application of bis-aryl hydrozone formation reaction overcame the challenge of poor kinetics of covalently targeting R-PE to cell surface receptors.

6.5 Experimental methods

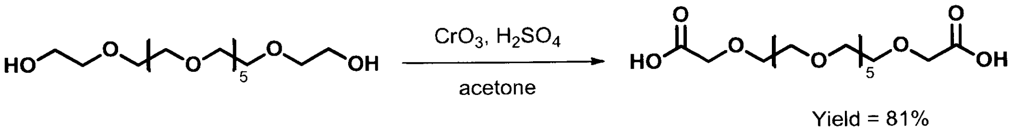
Quantum dot ligand synthesis

5-(1,2-dithiolan-3-yl)pentan-1-amine



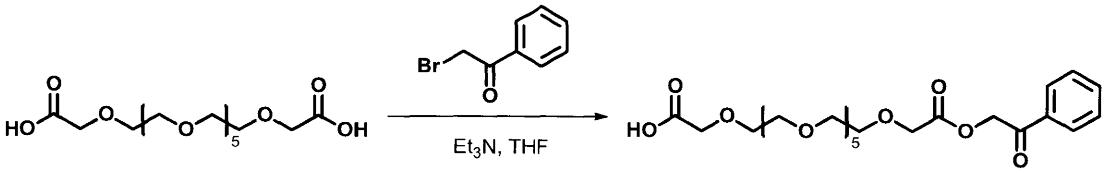
1.00 g of DL-α-lipoamide was dissolved into 100 mL of anhydrous tetrahydrofuran (THF). In another round bottom flask, 830 mg of $LiAlH_4$ power was added into 50 mL of anhydrous THF, and the mixture was stirred under nitrogen gas. DL-α-lipoamide solution was slowly added into this $LiAlH_4$ dispersion, and the reaction mixture was refluxed at 70 °C overnight. The reaction was worked-up by cooling down to 0 °C, followed by addition of 8 mL of distilled H_2O. The solution was stirred at 0 °C for another 30 min. THF solvent was removed under reduced pressure, and 50 mL of methanol (MeOH) was added to re-dissolve the reaction product. The insoluble was formed in MeOH solution and filtered away by running the solution through celite. The methanol solvent was removed from the supernatant on a rotevap. 100 mL of distilled H_2O was added and the pH was adjusted to 6.5 with 1 M HCl (aqueous solution). This solution was stirred at room temperature overnight. An equal volume of n-butanol was added into the solution to extract the product. The organic phase was washed with 100 mL of 1M NaOH, dried over anhydrous $MgSO_4$, and the solvent was removed under reduced pressure. Yield = 0.36 g (39%).

Polyethyleneglycol diacid



Jones reagent was prepared by dissolving 14.0 g of CrO_3 into 100 mL of distilled H_2O. The solution was chilled on ice, and then 12 mL of concentrated sulfuric acid was slowly added in small portions. The solution was warmed slowly to room temperature. 5.0 g (4.5 mL) of polyethyleneglycol-400 was dissolved into 500 mL of acetone. With stirring, 43 mL of Jones reagent was added into this solution. The solution was stirred at room temperature overnight. 0.5 g of activated charcoal and 1.0 g of sodium bicarbonate was added into the solution, and the mixture was stirred for another 2 hours. The mixture was filtered to obtain a clear, colorless solution. Solvent was removed from the supernatant on a rotevap. Yield = 4.35 g (81%).

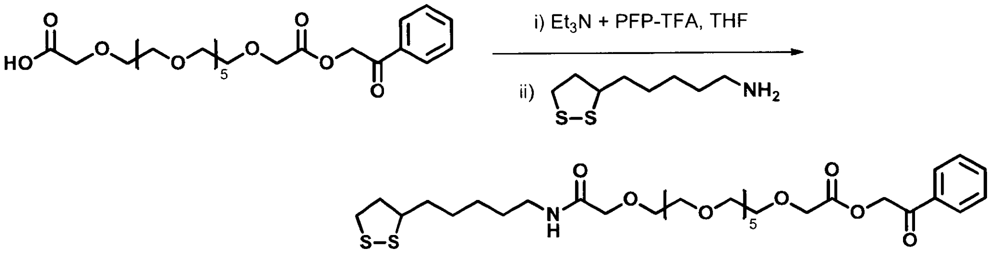
Phenacylcarboxyl PEG-acid



Yield = 19%

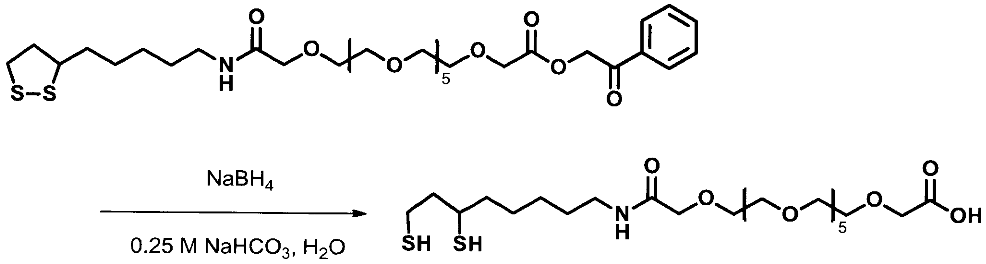
0.20 g of phenacyl bromide was dissolved into 5.0 mL of anhydrous THF, and 0.40 g of polyethyleneglycol diacid and 1.1 mL of triethylamine was added. The solution was stirred at room temperature overnight. 5 mL of distilled H_2O was added and the pH of the solution was adjusted to ~3 using 1M HCl. The aqueous phase was extracted twice with 20 mL of ethyl acetate (EtOAc), and the combined organic phase was dried over anhydrous $MgSO_4$. The solvent was removed under reduced pressure, and the product was purified on a silica column. A solution of EtOAc:MeOH:H_2O = 7:2:1 was used to elute the product. Yield = 0.10 g.

Phenacylcarboxyl PEG-lipoamide



0.10 g of phenacyl polyethyleneglycol diacid was dissolved into 1.0 mL of anhydrous THF, and into this solution were added 42 µL of anhydrous triethylamine (Et_3N) and 36 µL of pentafluorophenyltrifluoroacetate (PFP-TFA). The solution was stirred under nitrogen gas for 30 min. 40 mg of lipoamine was added into another 1.0 mL of anhydrous THF, and this solution was added drop-wise into the reaction mixture, which was stirred under nitrogen gas overnight. Solvent was removed *in vacuo*, the product was dissolved into 5 mL of chloroform ($CHCl_3$), and then washed with 5 mL of H_2O. The organic layer was dried over anhydrous $MgSO_4$. After filtering, solvent was removed again under vacuum. The product was purified with silica column. After sample loading, the column was washed with $CHCl_3$, and then the product was eluted with 2% MeOH/$CHCl_3$. Yield ~ 0.10 g (77%).

Polyethyleneglycoldiacyl monodihydrolipoamide



70 mg of phenacylcarboxyl PEG-lipoamide was dissolved into 1.0 mL of 0.25 M $NaHCO_3$ aqueous solution. 16 mg of $NaBH_4$ was slowly added into this solution (divided into four aliquots, with one aliquot added per 10 min). The reaction mixture was stirred on ice for 2 hours. 5 mL of MeOH was added to help solubilize the reaction mixture. Yellow color faded away in about 1 hour. The solution was stirred overnight. Solvent was removed under vacuum, and 5 mL of $CHCl_3$ was added to dissolve the residue. The organic phase was washed by 5 mL of 1M HCl and 5 mL brine. The aqueous phase was back-extracted with 5 mL $CHCl_3$. The combined organic phase was dried over anhydrous $MgSO_4$, and solvent was removed under vacuum. Yield = 60 mg (86%).

Mammalian cell culture and imaging

HEK and COS-7 cells were cultured in growth media consisting of Minimum Essential Medium (MEM, Cellgro) supplemented with 10% fetal bovine serum (FBS, PAA Laboratories). Cells were maintained at 37°C under 5% CO2. For imaging, HEK cells were grown on glass coverslips pre-treated with 50 μg/mL fibronectin (Millipore) to increase their adherence. COS-7 cells were grown in LabTek II chambered coverglass system 8 well plates.

Cells were imaged in Dulbecco's Phosphate Buffered Saline (DPBS) at room temperature. The images were collected on a Zeiss AxioObserver.Z1 microscope equipped with a Yokogawa spinning disk confocal head containing a Quad-band notch dichroic mirror (405/488/568/647 nm). YFP (491 nm laser excitation, 528/38 emission filter), AlexaFluor568 / Phycoerythrin (561 nm laser excitation, 617/73 emission filter) and DIC images were collected using Slidebook software (Intelligent Imaging Innovations). Images were acquired for 20 to 500 milliseconds using a Cascade II:512 camera. Fluorescence images in each experiment were normalized to the same intensity range.

TIRF images were acquired using 100x oil-immersion objective. YFP (491 nm laser excitation, 525/30 emission filter, 502 nm dichroic mirror), AlexaFluor568 / Phycoerythrin (561 nm laser excitation, 605/30 emission filter, 585 nm dichroic mirror) and DIC images were collected using Slidebook software (Intelligent Imaging Innovations). Digital images (16 bit) were obtained with a cooled EMCCD camera (QuantEM:512SC, Photometrics) with exposure times between 50 msec and 200 msec.

Cell surface labeling with R-PE

The amino acid sequence of LAP is GFEIDKVWHDFPA, in which the underscored lysine residue is covalently conjugated with the aryl hydrazine probe. HEK cells were transfected at ~70% confluence with 200 ng LAP-Neurexin-1β and 200 ng H2B-YFP co-transfection marker plasmid (per 48-well) using Lipofectamine 2000 (Invitrogen). 15 hours after transfection, the growth media was removed, and the cells were washed three times with DPBS with 0.5% casein. Casein was added to DPBS for all washing and labeling steps as a blocking agent and was required to reduce non-specific sticking of the probes. The cells were then labeled by applying 100 μM Hyd probe, 1 μM ^{W37I}LplA, 1 mM ATP, and 5 mM magnesium acetate in DPBS with 0.5% casein at room temperature for 45 minutes. Cells were then washed three times with DPBS with 0.5% casein and then treated with 10 mM aniline and 3 μM PE-Ald (4FB-R PE, Solulink) for 45 min at 4°C. Cells were washed an additional three times and then imaged. Similarly, COS-7 cells were transfected with 200 ng LAP-LDLR and 100 ng H2B-YFP co-transfection marker, only 20 μM Hyd probe was used in the initial labeling, and 0.3 μM PE-Ald with 20 mM aniline for 45 min was used for the fluorophore conjugation.

References

1 Rust, M. J., Lakadamyali, M., Zhang, F. & Zhuang, X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nat Struct Mol Biol* **11**, 567-573 (2004).

2 Ehrlich, M. *et al.* Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell* **118**, 591-605 (2004).

3 Howarth, M., Takao, K., Hayashi, Y. & Ting, A. Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* **102**, 7583-7588 (2005).

4 Howarth, M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat Methods* **5**, 397-399 (2008).

5 Mattoussi, H. *et al.* Self-assembly of CdSe-ZnS quantum dot bioconjugates using an engineered recombinant protein. *Journal of the American Chemical Society* **122**, 12142-12150 (2000).

6 Liu, W. H. *et al.* Compact cysteine-coated CdSe(ZnCdS) quantum dots for in vivo applications. *Journal of the American Chemical Society* **129**, 14530-+ (2007).

7 Pinaud, F., King, D., Moore, H. P. & Weiss, S. Bioactivation and cell targeting of semiconductor CdSe/ZnS nanocrystals with phytochelatin-related peptides. *Journal of the American Chemical Society* **126**, 6115-6123 (2004).

8 Uyeda, H. T., Medintz, I. L., Jaiswal, J. K., Simon, S. M. & Mattoussi, H. Synthesis of compact multidentate ligands to prepare stable hydrophilic quantum dot fluorophores. *Journal of the American Chemical Society* **127**, 3870-3878 (2005).

9 Howarth, M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat.Methods* **5**, 397-399 (2008).

10 Sapsford, K. E. *et al.* Kinetics of metal-affinity driven self-assembly between proteins or peptides and CdSe-ZnS quantum dots. *J Phys Chem C* **111**, 11528-11538 (2007).

11 Chang, W. R. *et al.* Crystal structure of R-phycoerythrin from Polysiphonia urceolata at 2.8 A resolution. *J Mol Biol* **262**, 721-731 (1996).

12 Irvine, D. J., Purbhoo, M. A., Krogsgaard, M. & Davis, M. M. Direct observation of ligand recognition by T cells. *Nature* **419**, 845-849 (2002).

13 Cohen, J. D., Zou, P. & Ting, A. Y. Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation. *Chembiochem* **13**, 888-894 (2012).

14 Dirksen, A. & Dawson, P. E. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug Chem* **19**, 2543-2548 (2008).

Curriculum Vitae

Peng Zou

Address	Department of Chemistry Massachusetts Institute of Technology 77 Massachusetts Avenue, Room 18-423 Cambridge, MA 02139
Phone	617-895-8977 (Cell) 617-258-0218 (Lab)
Email	zoup@mit.edu
Nationality	People's Republic of China
Born	July 18, 1985, Beijing, China

Education

2008 Summer	Physiology Course, Marine Biological Laboratory (Woods Hole, MA)
2007 – Present	Ph.D. student in Chemistry, Massachusetts Institute of Technology
2003 – 2007	B.S. *summa cum laude* in Chemistry, double major in Physics, Peking University, China

Research Experience

9/2007 – 12/2012, Massachusetts Institute of Technology

Graduate student with Prof. Alice Y. Ting.

Enzyme-based reporters for imaging specific proteins and mapping proteomes in living cells

5/2005 – 5/2007, Peking University

Undergraduate research with Prof. Bin Xia.

Study of the structure and chemical kinetics of severe acute respiratory syndrome coronavirus main protease

Honors and Awards

2011	Cold Spring Harbor Asia Fellowship (Cold Spring Harbor Laboratory)
2008	Lewis Paul Chapin Fellowship (MIT)
2008	Department of Chemistry Outstanding Teaching Award (MIT)
2005	President's Undergraduate Research Fellowship (Peking University)
2005	Wu Si Scholarship (Peking University)
2004	First Prize in Beijing College Physics Competition
2004	Xiang Lu Scholarship (Peking University)
2003	Freshman Scholarship (Peking University)
2003	Gold medal in Chinese Chemistry Olympiad
2003	Silver medal in Chinese Mathematics Olympiad

Publications

Cohen JD, Zou P, and Ting AY. Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation. *ChemBioChem* 2012, 13, 888-894.

Zou P, and Ting AY. Imaging LDL Receptor Oligomerization during Endocytosis Using a Co-internalization Assay. *ACS Chem. Biol.* 2011, 6, 308-313.

Prince RN, Schreiter ER, Zou P, Wiley HS, Ting AY, Lee RT, and Lauffenburger DA. The Heparin-binding domain of HB-EGF Mediates Localization to Sites of Cell-cell Contact and Prevents HB-EGF Proteolytic Release. *J. Cell Sci.* 2010, 123, 2308-2318.

Kang X, Zhong N, Zou P, Zhang S, Jin C, and Xia B. Foldon unfolding mediates the interconversion between M^{pro}-C monomer and 3D domain-swapped dimer. *Proc Natl Acad Sci U S A.* 2012, 109, 14900-14905.

Zhong N, Zhang S, Xue F, Kang X, Zou P, Chen J, Liang C, Rao Z, Jin C, Lou Z, and Xia B. C-terminal Domain of SARS-CoV Main Protease Can Form a 3D Domain-swapped Dimer. *Protein Sci.* 2009, 18, 839-844.

Zhong N, Zhang S, Zou P, Chen J, Kang X, Li Z, Liang C, Jin C, and Xia B. Without Its N-finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer Through Its C-terminal Domain. *J. Virology* 2008, 82, 4227-4234.