

**Enzyme-mediated labeling of proteins and protein-protein interactions
in vitro and in living cells**

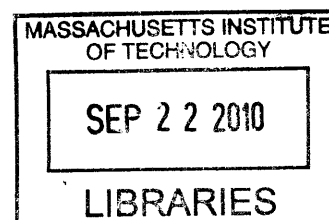
by
Sarah Ann Slavoff
B.S. Biochemistry (2005)
University of Maryland, College Park

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

September 2010

© Massachusetts Institute of Technology
All rights reserved



ARCHIVES

Signature of the author:

Department of Chemistry
July 13, 2010

Certified by:

Alice Y. Ting
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:

JoAnne Stubbe
Novartis Professor of Chemistry
Professor of Biology

Alice Y. Ting
Associate Professor of Chemistry
Thesis Supervisor

Catherine L. Drennan
Professor of Chemistry and Biology
Investigator and Professor, Howard Hughes Medical Institute

**Enzyme-mediated labeling of proteins and protein-protein interactions
in vitro and in living cells**

by
Sarah Ann Slavoff

Submitted to the Department of Chemistry
on July 13, 2010 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy

Abstract

The *E. coli* biotin ligase enzyme, BirA, has been previously used by the Ting research group for site-specific labeling of peptide-tagged cell surface proteins. We sought to expand the utility of biotin ligase-mediated labeling to functional group handles, including azides and alkynes, for bio-orthogonal chemistry. Since the BirA and its point mutants were unable to ligate these probes to an acceptor peptide, we screened biotin ligases from multiple species to identify more permissive enzymes. We determined that the *Pyrococcus horikoshii* biotin ligase utilizes an azide-bearing biotin analog and that the *Saccharomyces cerevisiae* biotin ligase can utilize an alkyne-functionalized biotin analog. We subsequently demonstrated that the azide-functionalized biotin analog can be derivatized with a phosphine probe *via* the Staudinger ligation.

We next turned to the goal of delivering quantum dots to the cytosol of living cells, which in the future may permit intracellular single-molecule imaging. We investigated viral methods of delivery, but found that our protocol caused quantum dots to be trapped in endocytic vesicles. We then validated previous reports that the pore-forming toxin streptolysin O be used to deliver quantum dots to the cytosol of living cells.

Lipoic acid ligase, or LplA, has been previously applied to site-specific protein labeling of peptide-tagged proteins using small molecule probes including lipoic acid and coumarin fluorophores. We utilized LplA and its substrate, the LAP peptide, to create sensors for protein-protein interactions. If LplA is fused to one protein and LAP is fused to another, only when the two proteins interact do LplA and LAP come into proximity, allowing probe ligation onto the peptide to occur as a readout of the interaction. We demonstrate that proximity-dependent coumarin ligation detects protein-protein interactions in living mammalian cells with extremely low background, a signal-to-background ratio of at least 5:1, and sufficiently fast kinetics to label interactions with a half-life of at least 1 minute. The reporter quantitatively responds to subpopulations of interacting proteins, allowing dissociation constants to be measured. Coumarin fluorescence accurately reports the subcellular localization of the interaction under study. Finally, we applied proximity-dependent coumarin ligation to imaging of the interaction of PSD-95 and neuroligin-1, two proteins involved in synaptic maturation, in neurons.

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

Acknowledgment

No acknowledgment can suffice to thank my advisor, Professor Alice Ting, for her mentorship over the last five years. She has been a constant source of ideas and support, from my first day in the lab, when she personally taught me how to transform bacteria and pour gels, to the time she spent helping me with my postdoctoral job search. I cannot imagine another advisor having such direct and positive impact on my scientific development from the most basic, tangible level of technical competence to (I hope) the highest level of intellectual growth.

I would also like to thank the members of my committee, Professor JoAnne Stubbe and Professor Catherine Drennan, for their helpful comments over the years and reading of my thesis. In particular I'd like to thank Professor Stubbe for her yearly encouragement and genuinely helpful critiques of my work.

I am deeply grateful to all past and present members of the Ting lab for their support and friendship. In particular, I'd like to thank Amar Thyagarajan not only for raucously entertaining and intellectually stimulating scientific and personal discussions, but also for his (sometimes) patient assistance with neuron labeling. I thank Irwin Chen for providing me with all of his interesting biotin analogs. Justin and Yoon-Aa, thank you for helping with my papers. My protein-protein interactions work would never have happened without advice from Marta, Katie and Tao. I thank everyone already mentioned as well as Mark Howarth, Yi, Sujiet, Hemanta, Takashi, Chi-Wang, Jackie Chan, Peng, Dan, Scott, Ken Hamill, Jen, Steph, Ken Loh, Jeff, Sam, Xin, Bill, and Phil for forming such a great network of supportive, smart, caring, capable people, who can always be counted on to help, advise, share, and just talk whenever needed. I affectionately wish you all the best in your future endeavors.

I would not be the person I am today without the steadfast friendship of Christine Tinberg. Her integrity, determination, intelligence, and tremendously caring nature inform every facet of her personal and professional life, and combine to make her one of the most capable and rigorous scientists – and the single greatest friend - I've ever known. She has always provided the emotional and intellectual support I needed to make it through life and graduate school, and a place to sleep during intermittent bouts of homelessness. Good luck in Seattle. I'll visit!

To my brother Steve and my Mom-Mom and Pop-Pop Slavoff and Pop-Pop Olson: I love you. Thank you for being there for me. The memory of my late Mom-Mom Olson will always inspire me to be a more caring person. I also want to thank my friends from MIT Chemistry, Chia-Hung Wu, Erik Dill, and Meredith Hartley, and my college roommates Tiffany and Shoshana, for their friendship and support over the years.

More than anyone else, I have to thank my husband, Julius Kusuma. He's been my rock through two years of long-distance visits, three weddings, two cats, and the final push to my Ph.D. For that, I beseech the Madonna del Ghisallo for his canonization.

I dedicate this thesis to Barbara Slavoff: my new beginning comes a little later than yours, Mom, but we're both stronger now, and I'm so proud of you for discovering your independence (and just a touch of fierceness).

Table of Contents

Title Page.....	1
Signature Page.....	2
Abstract.....	3
Acknowledgment.....	4
Table of Contents.....	6
List of Figures.....	10
List of Tables.....	12
List of Abbreviations.....	13
Chapter 1: Introduction: Current methodologies for protein labeling, intracellular nanoparticle delivery, and protein-protein interaction detection.....	15
Part I: Introduction to site-specific protein labeling in the cellular context.....	16
Unnatural amino acid mutagenesis.....	17
Protein-tagging methods.....	18
Peptide-tagging methods.....	20
Part II: Biotin ligase-mediated protein labeling and desired extensions to new small-molecule probes and intracellular quantum dot labeling.....	26
Biotin ligase.....	26
Physiological role and biochemical properties of biotin ligase.....	26
Application of biotin ligase to protein labeling.....	27
Protein labeling with small-molecule probes in the biological context: Introduction to bio-orthogonal chemistry.....	29
Intracellular labeling with quantum dots: Introduction to the delivery problem.....	31
Part III: Lipoic acid ligase-mediated protein labeling and introduction to protein-protein interaction detection.....	38
Lipoic acid ligase.....	38
Physiological role and biochemical properties of lipoic acid ligase.....	38
Application of lipoic acid ligase to protein labeling.....	39
Introduction to protein-protein interaction detection.....	40
Biochemical methods.....	42
Co-immunoprecipitation and tandem affinity purification.....	42

Chemical cross-linking and photo-cross-linking.....	44
Yeast genetic methods.....	46
Yeast two-hybrid.....	46
Split ubiquitin.....	48
Methods applicable to fixed mammalian cells.....	49
In situ proximity ligation.....	49
Proximity biotinylation.....	50
Methods applicable to live mammalian cells.....	50
Co-localization imaging.....	51
Resonance energy transfer.....	51
Protein complementation assays.....	55
Conclusion.....	62
References.....	66
Chapter 2: Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species.....	75
Introduction.....	76
<i>E. coli</i> BirA has low tolerance for biotin analog ligation to the AP peptide.....	77
Properties of biotin ligase enzymes across evolutionary space.....	78
Specificity of biotin ligase enzymes toward protein substrates.....	78
Monofunctional and bifunctional biotin ligases.....	78
Previously reported biotin ligase enzymes from other species: similarities and differences in biochemical properties.....	79
Comments on biotin ligase crystal structures.....	85
Study design.....	87
Results and discussion.....	88
Panel of biotin analogs.....	88
Panel of enzymes.....	91
Substrate selection and assay design.....	95
Probe screening by HPLC; product identification and assessment of ligation specificity by gel-shift assay and LC-MS.....	99
Secondary derivatization of DTB-Az by the Staudinger ligation.....	105
Kinetic analysis.....	107
Attempt to improve ligation kinetics: assaying the <i>P. horikoshii</i> biotin carboxyl carrier protein.....	109
Conclusion.....	110
Methods.....	112
References.....	121

Chapter 3: Delivering quantum dots into the cytosol of living cells.....	125
Introduction.....	126
Part I: Investigation of Influenza Virosomes for Quantum Dot Delivery.....	128
Introduction.....	128
Results, discussion and conclusions.....	131
Part II: Streptolysin O-mediated delivery of quantum dots into the cytosol of living cells.....	146
Introduction.....	146
Results and discussion.....	148
Delivering cell-impermeant cargo into the cytosol of living cells with streptolysin O.....	148
Targeting QDs to specifically biotinylated intracellular proteins.....	153
Conclusion.....	158
Methods.....	159
References.....	164
 Chapter 4: Imaging intracellular protein-protein interactions with proximity lipoylation and proximity coumarin ligation.....	 166
Introduction.....	167
Part I: Design and validation of proximity lipoylation and proximity coumarin ligation protein-protein interaction reporters.....	172
Introduction.....	172
Kinetic advantage of lipoic acid ligation by LplA.....	172
Live-cell labeling of PPIs with coumarin.....	173
Model system: Rapamycin-dependent interaction of FRB and FKBP.....	176
Results and Discussion.....	177
Proximity lipoylation to detect PPIs <i>in vitro</i> and in cells.....	177
Construct design and <i>in vitro</i> proximity lipoylation.....	177
Proximity lipoylation in living cells with immunoblotting detection.....	181
Proximity lipoylation in living cells with immunofluorescence detection.....	183
Proximity coumarin ligation for live-cell PPI imaging.....	185
Method demonstration and optimization: Identification of the optimal enzyme and peptide, demonstration of labeling specificity.....	185
Generality of proximity coumarin ligation: investigating different fusion geometries and applying the method in other cell lines.....	195
Kinetic analysis of proximity coumarin ligation.....	198
Rapamycin dose-response for proximity coumarin ligation.....	199
Wedge method to determine labeling yield and sensitivity of proximity coumarin ligation.....	203

Comparing proximity coumarin ligation to bimolecular fluorescence complementation.....	207
Part II: Application of proximity coumarin ligation to study the interaction of PSD-95 and neuroligin-1 in neurons.....	215
Introduction.....	215
Results and discussion.....	218
Specific labeling of the interaction of PSD-95 and neuroligin-1 in HEK cells.....	218
Coumarin proximity labeling of the interaction of PSD-95 and neuroligin-1 in neurons.....	223
Part III: Conclusion.....	227
Methods.....	231
References.....	246
Curriculum vitae.....	249

List of Figures

Chapter 1: Introduction: Current methodologies for protein labeling, intracellular nanoparticle delivery, and protein-protein interaction detection

Figure 1-1. Proposed mechanism of the Staudinger ligation.....30

Chapter 2: Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species

Figure 2-1. *E. coli* biotin ligase (BirA)-mediated ligation of biotin and ketone biotin to the acceptor peptide (AP).....76

Figure 2-2. Structures of biotin analogs used in this study.....88

Figure 2-3. Synthetic routes to (A) desthiobiotin azide (DTB-Az) and (B) *cis*-propargyl biotin (PB).....89

Figure 2-4. ClustalW alignment of a portion of the catalytic domain of biotin ligase enzymes utilized in this study.....93

Figure 2-5. SDS-PAGE characterization of purified biotin ligases.....95

Figure 2-6. The AP peptide is not recognized by biotin ligases from other species.....96

Figure 2-7. Biotinylation time courses of all enzymes under identical conditions.....99

Figure 2-8. Native gel shift assay showing preferential ligation of *cis*-PB over *trans*-PB by yeast biotin ligase.....102

Figure 2-9. HPLC detection of (A) DTB-Az and (B) *cis*-PB ligation to p67 acceptor protein by PhBL and yBL, respectively.....103

Figure 2-10. ESI-MS analysis of p67 conjugates to (A) DTB-Az and (B) *cis*-propargyl biotin.....105

Figure 2-11. Functionalization of ligated DTB-Az by Staudinger ligation.....107

Figure 2-12. Kinetic analysis of DTB-Az and *cis*-PB ligation reactions.....109

Figure 2-13. Gel-shift assay for *P. horikoshii* biotin ligase-catalyzed DTB-Az ligation to *P. horikoshii* BCCP (PhBCCP).....110

Chapter 3: Delivering quantum dots into the cytosol of living cells

Figure 3-1. Proposed mechanism of virosome-mediated quantum dot (QD) delivery.....130

Figure 3-2. Schematic of virosome preparation protocol.....132

Figure 3-3. Imaging virosomes reconstituted in the presence of QDs.....134

Figure 3-4. Cell-surface acidification assay for virosome fusogenicity.....137

Figure 3-5. Internalization of QD-containing virosomes into HEK cells.....140

Figure 3-6. Comparison of internalized virosome localization to a marker of late endosomes reveals endosomal trapping.....142

Figure 3-7. Proposed mechanism of streptolysin O (SLO)-mediated QD delivery.....148

Figure 3-8. Delivering fluorescein into cell lines with streptolysin O.....151

Figure 3-9. Delivering proteins into HEK cells with streptolysin O.....152

Figure 3-10. Delivering QDs into cell lines with streptolysin O.....153

Figure 3-11. SLO-mediated targeting of QDs to EphA3-AP and vimentin-AP.....156

Figure 3-12. SLO-mediated targeting of QDs to actin and specificity controls.....157

Chapter 4: Imaging intracellular protein-protein interactions with proximity lipoylation and proximity coumarin ligation

Figure 4-1. Schematic of proximity ligation.....	168
Figure 4-2. Structures of probes utilized in protein-protein interaction (PPI) reporter design.....	174
Figure 4-3. Ternary complex of FRB, FKBP, and rapamycin.....	177
Figure 4-4. Domain structures of proximity lipoylation constructs.....	178
Figure 4-5. Purification of FRB-LpIA and FKBP-LAP.....	179
Figure 4-6. <i>In vitro</i> proximity lipoylation analyzed by immunoblotting.....	180
Figure 4-7. Intracellular proximity lipoylation analyzed by immunoblotting.....	183
Figure 4-8. Immunofluorescence imaging of proximity lipoylation in the nucleus of COS-7 cells.....	184
Figure 4-9. Determination of optimal enzyme and peptide for proximity coumarin labeling: Representative images.....	188
Figure 4-10. Determination of optimal enzyme and peptide for proximity coumarin labeling: Single-cell plots of coumarin intensity against anti-c-myc intensity.....	189
Figure 4-11. Imaging proximity coumarin ligation in the cytoplasm of living HEK cells.....	191
Figure 4-12. Imaging proximity coumarin ligation in the nucleus of living HEK cells.....	193
Figure 4-13. Investigation of FRB-LpIA(W37V) and FKBP-LAP1 expression and localization during proximity coumarin ligation by immunofluorescence.....	194
Figure 4-14. Domain structures of proximity coumarin ligation constructs.....	195
Figure 4-15. Investigation of geometric sensitivity of proximity coumarin ligation.....	196
Figure 4-16. Imaging proximity coumarin ligation in the cytoplasm of living COS-7 and HeLa cells.....	198
Figure 4-17. Michaelis-Menten curve for coumarin ligation to LAP1 by LpIA(W37V).....	199
Figure 4-18. Rapamycin dose-response: representative images.....	202
Figure 4-19. Rapamycin dose-response: quantitation.....	203
Figure 4-20. Determination of proximity coumarin ligation labeling yield: representative images.....	206
Figure 4-21. Determination of proximity coumarin ligation labeling yield.....	207
Figure 4-22. Principle of yellow fluorescent protein (YFP) bimolecular fluorescence complementation (BiFC).....	208
Figure 4-23. Domain structures of proximity coumarin ligation and YFP BiFC constructs.....	209
Figure 4-24. Comparison of proximity coumarin ligation to BiFC: single-cell quantitation.....	212
Figure 4-25. Comparison of proximity coumarin ligation to BiFC: representative images.....	213
Figure 4-26. Diagrammatic representation of the interaction of neuroligin-1 and PSD-95 in the post-synaptic density.....	217
Figure 4-27. Domain structures of neuroligin-1 and PSD-95 constructs.....	219
Figure 4-28. Imaging the interaction of PSD-95 and neuroligin-1 in living HEK cells.....	223
Figure 4-29. Imaging the interaction of PSD-95 and neuroligin-1 in rat hippocampal neurons.....	226

List of Tables

Chapter 1: Introduction: Current methodologies for protein labeling, intracellular nanoparticle delivery, and protein-protein interaction detection

Table 1-1. Comparison of previously reported site-specific protein labeling methods.....25

Table 1-2. Comparison of previously reported quantum dot (QD) delivery methods.....37

Table 1-3. Comparison of current protein-protein interaction (PPI) detection methods.....65

Chapter 2: Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species

Table 2-1. Comparison of some biochemical properties of previously reported biotin ligase enzymes.....84

Table 2-2. Results of screening biotin ligase enzymes against probes.....100

Table 2-3. Cloning primers and template sources for biotin ligase enzymes.....114

Chapter 3: Delivering quantum dots into the cytosol of living cells

Chapter 4: Imaging intracellular protein-protein interactions with proximity lipoylation and proximity coumarin ligation

Table 4-1. Previously characterized kinetic parameters for LplA enzymes, LplA acceptor peptides (LAPs), and small molecule probes.....176

List of Abbreviations

AGT	O ⁶ -alkylguanine-DNA alkyltransferase
AM	Acetoxymethyl ester
AMP	Adenosine monophosphate
AP	<i>E. coli</i> biotin ligase acceptor peptide
ATP	Adenosine triphosphate
BirA	<i>E. coli</i> biotin ligase
BCA	Bicinchoninic acid
BCCP	Biotin carboxyl carrier protein
BCCP-87	C-terminal 87-amino acid fragment of <i>E. coli</i> BCCP
BiFC	Bimolecular fluorescence complementation
bio-5'-AMP	biotinyl-5'-AMP
BRET	Bioluminescence resonant energy transfer
BSA	Bovine serum albumin
C12E8	octaethylene glycol monododecyl ether
cDNA	Complementary DNA
CFP	Cyan fluorescent protein
co-IP	Co-immunoprecipitation
CoA	Coenzyme A
DHFR	Dihydrofolate reductase
DIC	Differential interference contrast
DIV	Days <i>in vitro</i>
DMEM	Dulbecco's modification of Eagle's medium
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's modification of phosphate buffered saline
DPBS-Mg	Dulbecco's modification of phosphate buffered saline with 1 mM MgCl ₂
DTT	Dithiothreitol
EBFP	Enhanced blue fluorescent protein
EDTA	Ethylenediamine tetraacetic acid
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionization mass spectrometry
EYFP	Enhanced yellow fluorescent protein
FBS	Fetal bovine serum
FITC-PE	Fluorescein isothiocyanate-phosphatidylethanolamine
FKBP	FK506 binding protein
FLIM	Fluorescence lifetime imaging
FRB	FKBP-rapamycin binding domain of mTor
FRET	Fluorescence resonant energy transfer
GFP	Green fluorescent protein
HA	Hemagglutinin
HCS	Holocarboxylase synthetase
HPLC	High-performance liquid chromatography
IPTG	isopropylthiogalactoside
LAP	<i>E. coli</i> lipolic acid ligase acceptor peptide

LAP1	First-generation <i>E. coli</i> lipoic acid ligase acceptor peptide
LAP2	Second-generation <i>coli</i> lipoic acid ligase acceptor peptide
LpIA	<i>E. coli</i> lipoic acid ligase
mRNA	Messenger RNA
MCCase	Methylcrotonyl-CoA carboxylase
MEM	Modified Eagle's medium
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
NBD	Nitrobenzoxadiazole
NLS	Nuclear localization sequence
OD	Optical density
P-LISA	Proximity ligation <i>in situ</i> assay
p-NPP	<i>para</i> -nitrophenylphosphonate
p67	C-terminal 67-amino acid fragment of human pyruvate carboxylase
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCA	Protein complementation assay
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
PhBCCP	<i>P. horikoshii</i> biotin carboxyl carrier protein
PhBL	<i>P. horikoshii</i> biotin ligase
PMSF	phenyl-methyl-sulfonyl fluoride
PPI	Protein-protein interaction
QD	Quantum dot
RNA	Ribonucleic acid
ROI	Region of interest
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	Small interfering RNA
SLO	Streptolysin O
SVMT	Sodium-dependent multivitamin transporter
TAP	Tandem affinity purification
TBS-T	Tris-buffered saline with 0.05% Tween-20
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
tRNA	Transfer RNA
yBL	Yeast (<i>Saccharomyces cerevisiae</i>) biotin ligase
YFP	Yellow fluorescent protein

Chapter 1: Introduction: Current methodologies for protein labeling, intracellular nanoparticle delivery, and protein-protein interaction detection

Part I: Introduction to site-specific protein labeling in the cellular context

The discovery of the green fluorescent protein (GFP) revolutionized the field of protein imaging. By fusing proteins of interest to GFP, or engineered fluorescent protein variants with emission spectra spanning the visual spectrum,[1] the localization, trafficking, and interactions of proteins can be visualized in living cells by fluorescence microscopy.[2] However, protein labeling by genetic fusion to fluorescent proteins suffers several fundamental limitations. First, fluorescent proteins are generally restricted to ensemble fluorescence imaging, because it is difficult to detect fluorescent proteins at the single-molecule level unless technically challenging experimental techniques, such as immobilization and stroboscopic excitation, are employed.[3] Although photo-activatable fluorescent protein variants have recently been used for single-molecule imaging,[4] organic fluorophores are generally brighter than fluorescent proteins and so are better suited to this purpose. In addition, labeling proteins with expanded biophysical functionalities would extend our ability to interrogate protein function beyond a simple fluorescent readout. Second, fluorescent proteins are large, and therefore can interfere with the function, interactions, and localization of the proteins to which they are fused.[5-7] The field of protein labeling, therefore, has focused on introduction of chemical probes to proteins, including small organic fluorophores, small biophysical probes such as photo-activatable cross-linkers, and inorganic nanoparticles applicable to new imaging modalities. The difficulty with this approach is that, while genetic fusion to GFP is both technically easy and perfectly specific, chemical probes are difficult to target to specific proteins in the cellular context.

Traditional methods for *in vitro* labeling with small molecules take advantage of nucleophiles present in proteins. For example, maleimides and haloalkanes react with the thiol group of cysteine residues, and N-hydroxysuccinimidyl esters react with amines present in lysine sidechains and the protein N-terminus. Of course, these reactions are not amenable to labeling a specific protein in the cellular context because they are not selective; instead they label all similarly reactive functional groups, which are present in most proteins.[8] While proteins can be labeled *ex vivo* and subsequently introduced into cells, for example by microinjection, this is a cumbersome and technically challenging technique. In order to bring chemical labeling of proteins into cells, new, selective targeting methods have therefore been devised.

In the most general sense, labeling specificity is achieved in the cellular context through genetic targeting. The protein of interest can be genetically fused to either a protein or peptide sequence that possesses reporter activity, or that has unique reactivity toward a chemical probe. Alternatively, unique functionality can be co-translationally incorporated into the protein of interest through unique mRNA recruiting sequences. By encoding unique reactivity or functionality at the genetic level, perfect specificity of chemical targeting can be ensured.

We will provide a brief review of the mechanisms by which site-specific small-molecule protein labeling on or in living cells has been accomplished through genetic fusions to proteins, peptides, or through co-translational reporter incorporation in order to motivate the need for the new labeling methods presented in this work.

Unnatural amino acid mutagenesis

Unnatural amino acid mutagenesis labels proteins by amber codon suppression, allowing co-translational insertion of an unnatural amino acid at a specific site in a protein. This method has been utilized for insertion of amino acids bearing reactive groups for chemoselective derivatization with secondary probes, photo-activatable cross-linkers, and fluorophores, among other functionalities, in bacterial, yeast, and mammalian cells.[9-12] Among all the methods we will discuss, unnatural amino acid incorporation has the greatest potential for non-perturbative protein labeling, as there is no additional sequence appended to the protein; it also provides perfect labeling specificity. However, the method suffers from competing termination at amber codons, which can produce truncated protein that can, in some instances, produce dominant negative effects. Furthermore, for each unnatural amino acid structure one desires to incorporate, a new aminoacyl-tRNA synthetase must be designed.

Protein-tagging methods

Protein tagging can be accomplished with genetic fusions to proteins that bind small molecules through non-covalent, though high-affinity, interactions. One example is dihydrofolate reductase, or DHFR. Two DHFR inhibitors, methotrexate and trimethoprim, bind to the enzyme with picomolar affinity.[13] Fluorophore conjugates of both of these molecules have been utilized for cellular labeling of DHFR fusion proteins; however, in the case of methotrexate labeling, DHFR-deficient cell lines must be utilized due to background from probe binding to endogenous DHFR.[14, 15] Another example of a protein used for non-covalent labeling is the F36V point mutant of the FK506 binding protein (FKBP), which binds with picomolar affinity to a synthetic ligand, SLF'. [16] Because SLF' binds very poorly to endogenous mammalian FKBP, this

method has been used for highly selective labeling in mammalian cells by conjugating dyes to SLF'. [17] Importantly, all of these small-molecule probes are membrane-permeable, permitting labeling of intracellular proteins., the non-covalent nature of these binding interactions mean that the probe dissociates over time, leading to signal decay; in the case of DHFR labeling with methotrexate derivatives, signal is lost in only one hour. [8]

Enzymes that can covalently self-label with “suicide substrates” solve the problem of probe dissociation. Two widely utilized, commercially available such methods have begun to be applied to solving biological problems. The first, developed by the Johnsson lab, utilizes the DNA repair enzyme *O*⁶-alkylguanine-DNA alkyltransferase (AGT). The natural reaction of AGT is to repair *O*⁶-alkylated guanine residues by covalent transfer of the alkyl group to an active-site cysteine. [18] AGT can react irreversibly with fluorescent derivatives of *O*⁶-benzylguanine, thereby labeling any protein to which AGT is genetically fused. [18] This technology is referred to as the SNAP-tag. An orthogonal mutant AGT enzyme that specifically reacts with *O*⁶-propargylguanine derivatives, referred to as the CLIP-tag, has been developed, facilitating two-color labeling of intracellular proteins. [19]

Another enzyme-based labeling method utilizes a mutant bacterial haloalkane dehalogenase called HaloTag. In the natural dehalogenase reaction, an active site aspartate forms an ester bond with the alkane substrate, which is subsequently cleaved through hydrolysis mediated by an active-site histidine that serves as a general base. [20] In HaloTag, the histidine has been mutated to prevent hydrolysis, affording an enzyme that reacts irreversibly with

haloalkanes.[20] Covalent labeling of HaloTag fusions with fluorophore-derivatized haloalkanes has been demonstrated for intracellular proteins.[20]

Cutinase, a fungal esterase, reacts irreversibly with *para*-nitrophenylphosphonate (p-NPP).[21] By synthesizing a p-NPP analog with a terminal thiol, derivatization of the cutinase-pNPP adduct with maleimide-containing secondary probes is possible.[21] Of course, highly negatively charged p-NPP is membrane-impermeant, and thiol nucleophiles are found throughout cellular proteins, so this method is restricted to labeling cell-surface proteins.

Intracellular proteins have also been labeled using native chemical ligation. By expressing proteins of interest as fusions to inteins, subsequent intein self-splicing generates a protein with an N-terminal cysteine; this cysteine then reacts selectively with membrane-permeable thioester-functionalized probes.[22] While recently reported selections to improve the rate of intracellular intein splicing may solve previous problems of slow kinetics,[23] competition with endogenous cysteine continues to limit the utility of the method.

Unfortunately, all protein-based tagging methods suffer from the same size problem as GFP, potentially affecting the localization or function of the protein of interest. Therefore alternative labeling methods targeted to short peptides have been developed.

Peptide-tagging methods

We begin with peptides that selectively bind small-molecule probes with high affinity. Perhaps the most important and widely used peptide-tagging method is FIAsH, developed by the Tsien laboratory. In this method, a tetracysteine-containing hairpin peptide is fused to the protein of

interest, which binds with low-picomolar affinity to biarsenical dyes, which fluoresce upon binding to the peptide.[24, 25] While the original method was developed for labeling with fluorescein, additional colors have been developed as biarsenical dyes,[26] and the resorufin-containing ReAsH probe can be utilized for photoconversion of diaminobenzidine for electron microscopy,[27] allowing multicolor and multimodal imaging. Because these probes are membrane-permeant,[28] FIAsh can label intracellular proteins, and has already been utilized for multiple cell biological studies (unlike many of the other methods described in this section).[29-32] However, the method suffers some problems. First, the biarsenical dyes have weak affinity for mono- and dithiols present in cells, leading to low-level background; therefore, only highly over-expressed or multimeric proteins can be imaged.[33] Arsenic toxicity remains a concern. Background and toxicity are reduced by treating the cells with dithiols, which leads to long labeling times. Finally, because the tetracysteine tag must be in its reduced form to react with arsenic, FIAsh labeling is restricted from oxidizing cellular compartments such as the cell surface and secretory pathway.

Metal ion-binding peptides have also been pressed into use for protein labeling. Hexahistidine or decahistidine peptides can bind nickel-nitriloacetic acid complexes conjugated to fluorophores,[34] but low-micromolar affinity of the peptides for nickel complexes leads to reporter dissociation and rapid signal degradation. An improvement on the method utilizes a fluorescein dizinc compound to improve affinity,[35] but can only be used for labeling cell-surface proteins. A similar method utilizes polyaspartate peptides to bind dinuclear zinc complexes for cell-surface protein labeling.[36] This method is also subject to probe

dissociation unless the target protein contains or is engineered to contain a surface cysteine that can react with a chloroacetamide moiety in the probe.[37]

Peptides that specifically bind small molecules and proteins have also been developed. A peptide that specifically binds the organic fluorophore Texas Red with picomolar affinity has been reported and utilized for cell-surface protein labeling, but accumulation of Texas Red in mitochondria creates high background that limits the utility of the method.[38] Peptides that bind the proteins streptavidin and bungarotoxin have been used for cell-surface protein labeling,[39] but the large size of these proteins prevents their intracellular delivery and could potentially interfere with the function of the labeled protein.

Tremendous improvements in peptide-directed labeling have been made by introducing the additional selectivity of enzyme-mediated peptide labeling. In these approaches, enzymes that normally catalyze covalent post-translational modifications are hijacked to specifically attach functionalized small-molecule probes to a peptide tag, retaining the specificity of enzyme-substrate interactions while maintaining the small tag size of peptide labeling approaches. The first example of this utilizes the phosphopantetheinyltransferase Sfp, which normally catalyzes attachment of coenzyme A (CoA) to a specific serine residue of acyl carrier proteins and peptidyl carrier proteins involved in the biosynthesis of polyketides and nonribosomal polypeptides.[40] The principle advantage of Sfp is that it accepts CoA derivatives bearing many substituents, including biotin and thiols, on the terminal thiol as substrates.[41] Sfp can site-specifically ligate these CoA derivatives to a specific lysine residue of short engineered peptides in the cellular context.[42] Because CoA is highly charged, the probes used with this method

cannot cross the cell membrane, meaning it is restricted to the cell surface. Even so, the phosphopantetheinyltransferase labeling methodology has already contributed significantly to biological study of the mobility of cell surface receptors.[43, 44]

Formylglycine generating enzyme has recently been put to use in site-specific labeling of cell surface proteins. This enzyme normally catalyzes co-translational formation of formylglycine in the active site of sulfatases, a modification which is necessary for their catalytic activity.[45] The “aldehyde tag” peptide recognized by the enzyme is exceptionally short at only 6 amino acid residues.[45] Because mammalian formylglycine generating enzyme is localized to the endoplasmic reticulum, this method has been recently used to install aldehydes into peptide-tagged cell-surface proteins as they transit the secretory pathway.[46] Aldehydes are otherwise absent from the cell surface, so the aldehyde-labeled protein can then be chemoselectively labeled with with hydrazine- and hydroxylamine-functionalized fluorophores in living cells.[46] Unfortunately, because aldehydes are prevalent in intracellular metabolites, the method will be generally restricted to the cell surface or purified proteins.

Protein farnesyltransferase has been appropriated for attaching azide- and alkyne-modified isoprenoid probe onto proteins bearing a 4-amino acid farnesylation motif.[47] These probes can be selectively derivatized in a secondary chemical step to install fluorophores. However, the farnesylation motif employed is not orthogonal to endogenous farnesylated proteins, so this will not be a general tool for site-specific labeling in the intracellular context, as all farnesylated proteins will be labeled.

Guinea pig transglutaminase normally catalyzes protein cross-linking events by catalyzing formation of an amide bond between lysine and glutamine sidechains.[48] Most interestingly for protein labeling applications, transglutaminase is permissive for the structures of amines it can utilize in this amidation reaction, including biotin- and fluorescein-cadaverine.[48] Furthermore, a short glutamine-containing peptide called the Q-tag has been reported as a substrate of transglutaminase.[48] These properties have been co-opted for labeling of Q-tagged cell surface proteins with fluorophores.[48] However, glutamine residues are not an orthogonal reactive group on the cell surface, so the enzyme modifies endogenous cell-surface glutamines as well as the Q-tag. This lack of specificity limits the utility of the method.

We provide a summary and comparison of all of these methods for site-specific protein labeling in Table 1-1. The Ting research group has reported improved peptide-directed site-specific protein labeling methodologies based on the enzymes biotin ligase and lipoic acid ligase.

Generally, the theme of this thesis is expanding the utility of the biotin ligase and lipoic acid ligase labeling methods by extending them to new reactions and biological problems. Part II of this introduction will discuss the enzymology of biotin ligase, how the enzyme has been utilized by the Ting Lab for protein labeling, and our efforts to use biotin ligase to incorporate new small molecule probes as well as to extend its utility to the intracellular context. Part III will introduce lipoic acid ligase, discuss how this enzyme has been used for labeling, and explain the new biological problem to which we apply the enzyme: detection of protein-protein interactions.

Labeling method	Tag size	Specificity for target protein	Intracellular labeling in living cells
Unnatural amino acid mutagenesis	One amino acid	Perfect	Yes
Protein-directed methods			
Fluorescent proteins	Large	Perfect	Yes
Dihydrofolate reductase	Large	High in DHFR-deficient cell lines	Yes
FK506 binding protein	Large	High	Yes
<i>O</i> ⁶ -alkylguanine-DNA alkyltransferase (SNAP/CLIP tags)	Large	High	Yes
HaloTag	Large	High	Yes
Cutinase	Large	High for enzymatic step; moderate for chemical step because of competition with endogenous cysteine	No
Native chemical ligation	Large	Low; competes with endogenous cysteine	Limited
Peptide-directed methods			
FLAsH	Small	Moderate; competes with endogenous cysteine	Yes
Poly(histidine)	Small	High on cell surface	No
Poly(aspartate)	Small	High on cell surface	No
Texas Red binding peptide	Small	High	Limited
Streptavidin binding peptide	Small	High	No
Bungarotoxin binding peptide	Small	High	No
Phosphopantetheinyl-transferase	Small	High	No
Formylglycine generating enzyme	Small	High for enzymatic step; high for chemical step only on cell surface	No
Protein farnesyltransferase	Small	Low for enzymatic step; high for chemical step	No
Transglutaminase	Small	Low; competes with endogenous glutamines	No
Biotin ligase	Small	High for enzymatic step; high for ketone biotin derivatization only on cell surface	No
Lipoic acid ligase	Small	High	Yes

Table 1-1. Comparison of previously reported site-specific protein labeling methods. Note that the biotin ligase method is described with respect to ketone biotin labeling.

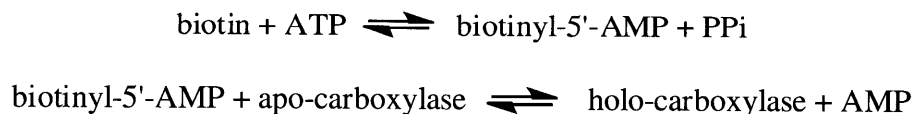
Part II: Biotin ligase-mediated protein labeling and desired extensions to new small-molecule probes and intracellular quantum dot labeling

Biotin ligase

Physiological role and biochemical properties of biotin ligase

Biotin is an essential cofactor responsible for carboxyl group transfer reactions involved in several catabolic pathways, fatty acid biosynthesis, and gluconeogenesis.[49] It must be covalently attached to a specific lysine residue of the carboxylase proteins that require it; this attachment is catalyzed by the enzyme biotin ligase.[50] Because biotin is essential for metabolism, all organisms express one or two biotin ligase enzymes.[51]

The best-characterized biotin ligase by far is the *E. coli* enzyme, BirA. BirA catalyzes amide bond formation between a lysine residue and biotin in a two-step reaction as shown in the following equations:[52]



In the first step, biotin is activated as the adenylate intermediate and pyrophosphate (PPi) is released; in the second step, the lysine residue serves as a nucleophile to attack the adenylate, forming the amide product and releasing adenosine monophosphate (AMP).[53] (Apo-carboxylase refers to the substrate protein prior to post-translational modification; holo-carboxylase refers to the substrate protein with biotin attached.)

Biotinylation is an extremely rare and specific post-translational modification. *E. coli* BirA catalyzes biotinylation of exactly one protein, the biotin carboxyl carrier protein (BCCP) of acetyl CoA carboxylase ligase; yeast contains up to five biotinylated proteins, and mammals and plants have four.[51] (While the biotinylation of histones has been sporadically investigated over the years,[54] this modification has been demonstrated not to occur *in vivo*. Detection of histone biotinylation is an artefact of non-specific modification due to high protein concentrations and long reaction times *in vitro*,[55] as well as non-specific binding of avidin and certain antibodies to native, non-biotinylated histones.[56]) Furthermore, exactly one lysine residue on the biotinylated protein is modified by biotin ligase; it is found at the center of a consensus sequence that is typically close to the C terminus of the carboxylase.[51] That these enzymes are able to specifically modify one to five proteins in a cell, while never transferring biotin to any other lysine residue of any other protein, testifies to their exquisite specificity for their protein substrates. The kinetics of biotin transfer to BCCP by *E. coli* BirA have been previously characterized; the k_{cat} is 0.16 s^{-1} , and the K_m values for biotin, ATP, and BCCP are $0.49\text{ }\mu\text{M}$, 0.3 mM , and $4.39\text{ }\mu\text{M}$, respectively.[57]

A 15-amino acid peptide substrate for BirA, the acceptor peptide, or AP, has previously been reported.[58] The kinetics of AP biotinylation by BirA are identical to biotinylation of BCCP in the steady state, with k_{cat}/K_m of $10,000\text{ M}^{-1}\text{ s}^{-1}$ for the AP, and $11,900\text{ M}^{-1}\text{ s}^{-1}$ for BCCP.[58] Interestingly, while biotin ligase enzymes can generally biotinylate BCCPs from other species, the AP is specifically recognized by BirA and not by other enzymes.[59, 60]

Application of biotin ligase to protein labeling

The biochemical properties of BirA are ideally suited for application to a peptide-based labeling methodology. BirA has exquisitely high specificity for labeling the AP in the context of mammalian cells, and mammalian biotin ligase enzymes do not recognize the AP.[59, 61, 62] Therefore, BirA-mediated protein labeling achieves its site-specific small-molecule conjugation by harnessing the specificity of an enzyme for its substrate, while retaining the advantages of small tag size obtained with peptide-directed labeling technologies.

Biotinylation of the AP tag can be detected with fluorescently labeled streptavidin or with streptavidin-coated quantum dots (QDs).[62-64] QDs are semiconductor nanoparticles with ideal photophysical properties for single-molecule imaging; they are exceptionally bright, photostable, and have very narrow emission spectra.[62] BirA-mediated targeting of streptavidin and QDs has been applied to ensemble and single-molecule imaging of AP-tagged cell surface proteins.[62-64] However, streptavidin and QDs are both currently restricted to cell-surface applications, because both are too large to diffuse across the cell membrane.

A ketone isostere of biotin has also been developed and utilized for labeling of cell-surface proteins.[61, 65] Ketone biotin can be ligated to the AP by BirA and, because ketones are absent from the cell surface, subsequently selectively reacted with hydrazine- and hydroxylamine-functionalized secondary probes under physiological conditions.[61] This methodology has been used to install fluorophores and a benzophenone photo-activatable cross-linker on cell-surface proteins.[61] However, due to the abundance of ketones and aldehydes in intracellular metabolites, the secondary conjugation step cannot be applied to intracellular proteins, restricting ketone biotin labeling to the cell surface.

While these labeling technologies are undeniably powerful, we have identified two improvements that can be made in BirA-mediated protein labeling: the toolbox of small-molecule probes that can be attached to proteins must be expanded, and quantum dot labeling must be extended to intracellular proteins, through the development of methods to deliver QDs into the cytosol of living cells. The next two sections will provide background information on each of these problems.

Protein labeling with small-molecule probes in the biological context: Introduction to bio-orthogonal chemistry

In order to expand the utility of BirA for small-molecule labeling, we focused on designing probes with reactive “functional handles” compatible with intracellular bio-orthogonal derivatization. Bio-orthogonal reactions are, in general, chemoselective reactions that can occur in the biological milieu. For a reaction to be bio-orthogonal, the functional groups involved in the bio-orthogonal reaction must not be present in biological molecules, the reaction must not perturb native biomolecules, and it must proceed rapidly under physiological conditions.[66] Ketone biotin and formylglycine, in their reactivity toward hydrazines and hydroxylamines, are excellent examples of such a system, but as discussed, this chemistry is only bio-orthogonal on the cell surface.

In particular, we wished to introduce probes bearing azide and alkyne functional groups. Azides and alkynes are absent both from the cytoplasm and the cell surface. Alkynes can undergo copper-catalyzed [3+2] cycloaddition with azides (termed “Click chemistry”) to form stable triazole adducts under physiological conditions.[67, 68] Azides are particularly useful for bio-

orthogonal conjugation in cells because they can be derivatized with additional, non-toxic secondary reactions aside from Click chemistry. The Staudinger ligation of azides and triarylphosphines introduces an electrophilic methyl ester in the phosphine reagent to trap the reactive aza-ylide intermediate, generating a stable amide bond in the product (Figure 1-1).[69, 70] Additionally, a copper-free version of the [3+2] azide-alkyne cycloaddition has been developed, in which the azide reacts rapidly with a strained cyclo-octyne.[71] We hoped that by extending BirA-mediated labeling to include azide- and alkyne-functionalized probes, we would be able to access these more selective secondary conjugation reactions.

Chapter 2 of this thesis describes our efforts to expand the small-molecule substrate scope of BirA to new probe classes, including probes functionalized for bio-orthogonal chemistry, by investigating the substrate specificity of biotin ligase enzymes from many species. A further discussion of previously characterized biotin ligase enzymology that is directly relevant to the study design will also be presented in Chapter 2.

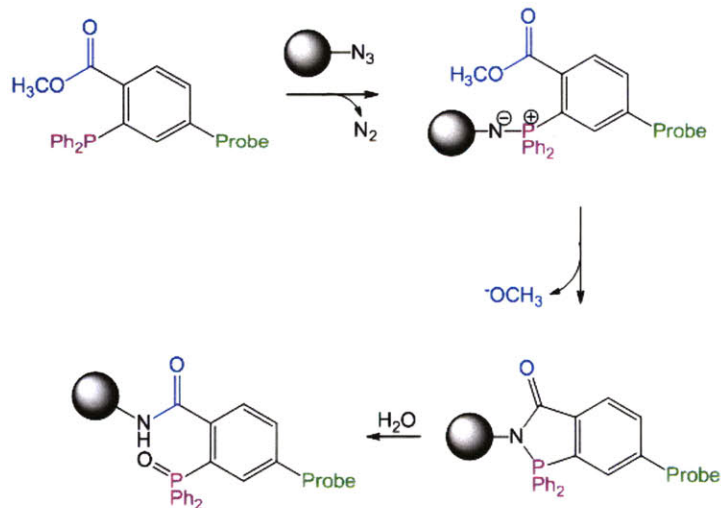


Figure 1-1. Proposed mechanism of the Staudinger ligation. A protein, represented by the gray circle, displaying an azide functional group can react with a triarylphosphine-bearing probe, proceeding through an aza-ylide intermediate that reacts with an intramolecular electrophile, shown in blue, to produce an amide linkage in the product. Any compatible probe structure can be introduced.

Adapted from reference 18 with permission, copyright 2005, American Chemical Society.

Intracellular labeling with quantum dots: Introduction to the delivery problem

As mentioned above, the use of QDs for single-molecule imaging is currently restricted to the cell surface because the size (approximately 20 nm)[72] of QDs prevents them from crossing the cell membrane by diffusion. We sought to extend BirA-mediated QD targeting to intracellular proteins by devising a method to deliver QDs into the cytosol of living cells (and, once delivered, we anticipate that delivered streptavidin-coated QDs should be targetable to intracellular proteins biotinylated by BirA, because biotinylation of the AP in the cellular context has previously been shown to be specific[59, 61, 62]). More specifically, we sought to develop methods that are high-throughput (that is, deliver QDs into many cells at once), non-toxic, simple to use, and accessible to the non-expert.

QDs can be readily internalized by endocytosis; however, these endocytosed QDs remain trapped in vesicles, which they cannot readily escape,[73, 74] precluding their targeting to proteins of interest inside the cell. An effective method would transport QDs across the plasma membrane, or out of endosomes, into the cytosol, leaving them freely diffusible.

Many other methods have been investigated for the delivery of QDs into the cytoplasm of living cells. We provide a brief review of the field here, along with a discussion of the problems each method faces, to motivate the need for new methods. We discuss only methods that have been applied for QD delivery in living cells, as cell fixation can introduce localization artifacts (especially for charged molecules) and precludes rigorous conclusions from being drawn about delivery efficacy; we also focus on methods that are applicable in commonly used mammalian

cell lines, rather than methods that are unique to specialized cell types and therefore have no hope of generality.

Microinjection has been established to afford targetable intracellular QDs.[75] For example, QDs have been targeted to an AP tag on the cytoplasmic tail of the receptor EphA3 after intracellular biotinylation by BirA; conformational changes in individual QD-labeled EphA3 molecules were then observed in the cellular context.[76] While microinjection is inarguably effective, it is technically difficult, low-throughput, and requires specialized instrumentation. Atomic force microscopy-based nanoinjectors are similarly efficacious, but are even more inaccessible and low-throughput.[77] Encapsulating QDs in nanospheres consisting of Poly(D,L-lactide-*co*-glycolide) has also been reported to afford cytosolic, diffusible QDs that are targetable *via* antibody conjugation to intracellular proteins.[78] However, this preparation is technically challenging, so until these encapsulated QDs become widely available, the method is not accessible to the non-chemist. A simple, non-toxic, inexpensive method that can deliver QDs into the cytosol of many cells would make intracellular QD targeting a general method and extend its accessibility to the non-expert.

One of the earliest methods investigated for QD delivery was targeting to recycling endosomes *via* QD conjugation to receptor ligands. Transferrin-conjugated QDs have been demonstrated to exhibit punctuate, endosomal localization inside living HeLa cells.[79] Direct conjugation of folate to QDs as well as QD encapsulation in folate-displaying lipid micelles has also been reported to result in punctuate and likely endosomal localization.[80, 81] Therefore, while

conjugation to receptor ligands does promote QD uptake relative to the rate of free QD internalization, it does not generally permit internalized QDs to freely enter the cell cytosol.

Cell penetrating peptides, or CPPs, are small, positively charged peptides or proteins that can cross cell membranes and have received extensive attention in the fields of gene and drug delivery.[82-84] TAT peptide has been extensively used for delivery of nucleic acids, but TAT-conjugated QDs remain encapsulated in endosomes to the extent that escape into the cytosol is undetectable.[74, 85] In fact, Tat peptide only efficiently delivers QDs into cells after their plasma membrane has been permeabilized with the glycoside detergent digitonin.[86] A similar endosomal trapping problem has been reported for Pep-1-conjugated quantum dots in some studies,[87, 88] though another reported demonstrates nuclear targeting of Pep-1-delivered QDs.[89] Insect neuropeptide has been reported to deliver QDs into the cytosol and nucleus of living cells, but no assay to demonstrate free diffusibility was performed, so the utility of this method for intracellular protein targeting remains to be confirmed.[90] The palmitoylated peptide Palm-1 has been conjugated to QDs, and 48 hours after delivery, some percentage of QDs were demonstrated to escape from endosomes into the cytosol based on co-localization analysis with an endosomal marker.[91] Despite this long delivery time, these cytosolic QDs may prove useful for protein labeling. It is clear, though, that with peptide-based delivery methods, at best some percentage of QDs escape from endocytic vesicles, with the rest remaining trapped.

Cationic polymers such as polyethyleneimine (PEI) and poly-L-lysine have similarly been extensively used for transfection. The mechanism of PEI delivery has been the subject of some

study. The association of the PEI with negatively charged cargo such as DNA promotes its association with and uptake by cells. Protonatable groups on the polymer then act as a “proton sponge,” “soaking up” protons as the endosome acidifies, promoting additional proton and chloride counter-ion uptake and eventual osmotic rupture.[92, 93] While PEI has been reported to afford endosomal escape of delivered QDs, the concentrations of PEI required for this effect were highly toxic; non-toxic concentrations and PEI formulations resulted in endosomal trapping.[94] Derivatized poly-L-lysine has been demonstrated to promote QD uptake, but the resultant intracellular QD pattern is punctuate, a likely indicator of endosomal trapping.[95]

Cationic liposomes are effective commercially available nucleic acid transfection reagents. Commercial lipofectamine has been put to the purpose of delivering QDs, and found to be efficacious in affording endosomal escape based on co-delivery of siRNA[96] and also comparison to endosomal dye localization.[75] However, the delivered QDs were observed to form huge aggregates inside the cell of several hundred nanometers in size, rather than to freely diffuse.[75] The commercial transfection reagent FuGene has been reported to deliver targetable QDs into cells, but no experimental data to confirm the reported localization was presented, and aberrant kinesin motility was observed, so this result must be confirmed.[97]

Carbon nanotubes have been extensively studied for drug delivery applications because they associate with the cell membrane and internalize into clathrin-coated vesicles with high efficiency.[98, 99] However, QD-nanotube conjugates remain trapped in endosomes and are eventually trafficked to lysosomes,[100].

Electroporation is a classical method of transfection. While electroporation delivers QDs into cells, it results in formation of very large QD aggregates rather than freely diffusible QD delivery.[75]

Osmotic lysis is a delivery method in which cells are allowed to take up cargo *via* pinocytosis in buffer made hypertonic with high concentrations of sucrose and polyethyleneglycol. After the cargo has been loaded into vesicles, the cells are exchanged into hypotonic buffer, causing osmotic shock and vesicle rupture.[101] This method has been utilized for loading of QD-conjugated kinesin and myosin V into cells; subsequently, the motor protein-QD conjugates were observed to freely diffuse as well as to “walk” on microtubules or actin, respectively, with the expected velocity and step size.[102, 103] This suggests that osmotic lysis can afford free cytosolic QDs. However, the method is not without toxicity; subjecting cells to osmotic lysis of pinocytic vesicles once does not cause gross changes in cell morphology, after multiple rounds of lysis, up to 40% of cells die.[101]

Biological toxins including shiga toxin, diphtheria toxin, anthrax toxin, and cholera toxin, which either afford delivery from endocytic vesicles or escape from the endoplasmic reticulum after undergoing retrograde transport, have been used to deliver many types of cargo.[104] Cholera toxin has been applied to QD delivery, but resulted in QDs being trapped in endosomes that eventually accumulated in the perinuclear region, the typical late endosome localization.[105] Shiga toxin and ricin have similarly been shown to result in endosomal trapping of QDs.[106] Streptolysin O is a particularly interesting biological toxin that forms pores in biological membranes through which membrane-impermeant cargo can passively diffuse.[107]

Streptolysin O treatment has previously been shown to afford cytosolic delivery of monodisperse, diffusible QDs into the cytosol of living cells, but limited imaging data was presented in the literature and no targeting to proteins was attempted.[108] While this method has potential for cytosolic QD delivery, it has issues of toxicity. We will discuss the mechanism, benefits, and drawbacks of streptolysin O in much greater detail in Chapter 3 of this thesis.

In conclusion, most methods applied to the delivery of QDs suffer from technical inaccessibility, endosomal trapping, QD aggregation, or toxicity. We provide a summary comparison of these methods in Table 1-2. Therefore, the application of BirA-mediated QD targeting to intracellular proteins awaits the development of a general, easy-to-use method for delivering monodisperse, diffusible QDs into the cytosol of living cells. Chapter 3 of this thesis describes our efforts to apply viral mechanisms of cell entry and streptolysin O-mediated cell permeabilization to the QD delivery problem.

Delivery method	Subcellular localization of QDs after delivery	Evidence of aggregation	Targeting to intracellular proteins	Toxicity
Direct injection methods				
Microinjection	Cytosolic	No	Yes	Mild
Nanoinjector	Cytosolic	No	No	Low
Receptor ligands				
Transferrin	Endosomal	No	No	Low
Folate	Endosomal	No	No	Low
Cell-penetrating peptides				
Tat	Endosomal	No	No	Low
Pep-1	Endosomal	No	No	Low
Insect neuropeptide	Cytoplasmic and nuclear	No	No	Low
Palm-1	Cytosolic and endosomal	No	No	Low
Cationic polymers				
Polyethyleneimine	Cytosolic	No	No	High
Poly-L-lysine	Endosomal	No	No	Low
Cationic liposomes				
Lipofectamine	Cytosolic	Yes	No	Low
FuGene	Cytosolic	No	Yes	Low
Biological toxins				
Cholera toxin	Endosomal	No	No	Low
Shiga toxin	Endosomal	No	No	Low
Ricin	Endosomal	No	No	Low
Streptolysin O	Cytosolic	No	No	Moderate
Other delivery methods				
Carbon nanotubes	Endosomal/lysosomal	No	No	Low
Electroporation	Cytosolic	Yes	No	Low
Osmotic lysis of pinocytic vesicles	Cytosolic	No	No; QD-protein conjugate prepared <i>ex vivo</i> and delivered intact	Mild
Poly(D,L-lactide-co-glycolide) nanosphere encapsulation	Cytosolic	No	Yes	Low

Table 1-2. Comparison of previously reported quantum dot (QD) delivery methods.

Part III: Lipoic acid ligase-mediated protein labeling and introduction to protein-protein interaction detection

Since the work introduced in Part II of this chapter, a superior ligase enzyme for small-molecule protein labeling has been introduced: *E. coli* lipoic acid ligase, or LplA. We therefore abandoned our BirA engineering efforts and designed future applications to utilize LplA-mediated labeling. In this section we will first describe the biochemical properties of LplA and how it has been applied to site-specific protein labeling, then detail the problem to which we will apply LplA: detection of protein-protein interactions.

Lipoic acid ligase

Physiological role and biochemical properties of lipoic acid ligase

Lipoic acid, like biotin, is a cofactor that must be covalently attached to the enzymes that require it, such as the glycine cleavage system, alpha-ketoglutarate dehydrogenase, and pyruvate dehydrogenase.[109] Free lipoic acid can be conjugated to specific lysine residues of its cognate proteins through the ATP-dependent action of lipoic acid ligase, or LplA, in *E. coli*,[110] in exact analogy to BCCP biotinylation by BirA. (A parallel lipoylation pathway exists, catalyzed by the enzymes LipB and LipA, in which octanoic acid produced by the fatty acid biosynthetic pathway is transferred from acyl carrier proteins to lipoyl acceptor domains, then enzymatically converted to lipoic acid *via* sulfur insertion.[109]) The kinetic properties of the enzyme have been investigated; its K_m for lipoic acid and ATP are 1.7 μM and 3 μM , respectively.[110] Its k_{cat} for lipoic acid ligation is 0.22 s^{-1} . [111] Interestingly, LplA also has

previously been reported to catalyze attachment of octanoic acid and selenolipoic acid to lipoyl acceptor proteins.[110]

Two peptide substrates for LplA have previously been reported by the Ting lab. The original peptide substrate, called LAP1, was rationally designed by comparing consensus sequences of naturally occurring lipoylated proteins.[112] While the k_{cat} for ligation of a small-molecule substrate to this peptide by LplA, at 0.05 s^{-1} , was found to be only about 2-fold slower than the k_{cat} of ligation of this substrate to a lipoylated domain of *E. coli* pyruvate dehydrogenase, at 0.11 s^{-1} . However, the affinity of LplA for LAP1 is so poor that initial attempts to measure the K_m were unsuccessful, except to estimate that it must be greater than $200 \mu\text{M}$. [112] A high-affinity peptide substrate, the LAP2, was therefore obtained by yeast-display evolution.[111] LplA catalyzes lipoic acid attachment to the LAP2 with a k_{cat} of 0.22 s^{-1} and a tremendously decreased K_m of $13 \mu\text{M}$. [111]

Application of lipoic acid ligase to protein labeling

Similarly to BirA, the ability of LplA to selectively ligate small molecules to peptide substrates had great promise for developing a peptide-directed labeling strategy for recombinant proteins in the cellular context. LplA, however, has an important advantage: many small molecule probes can be accommodated by this enzyme.

The capacity of wild-type LplA to attach octanoic acid and selenolipoic acid to lipoyl acceptor proteins suggested that this enzyme may exhibit plasticity toward small-molecule substrates while retaining specificity for its protein or peptide substrate. Fernández-Suárez *et al.* found that this is indeed the case, and that wild-type LplA can ligate a series of functionalized probes,

including an alkyl azide, to its LAP1 acceptor peptide.[112] The alkyl azide was used to label LAP1-tagged cell-surface proteins with cyclo-octyne-bearing probes and fluorophores.[112]

Perhaps the most interesting feature of LpIA is the capacity for mutants of the enzyme to accommodate interesting probe structures as substrates. Specifically, mutations at W37, which lies deep in the binding pocket and interacts with the dithiolane ring of lipoic acid,[113-115] increase the size of the binding pocket and permit large probes to bind. First, a mutant of LpIA have been demonstrated to ligate a probe bearing an aryl azide photo-activatable cross-linker site-specifically to the LAP1 peptide, which was subsequently utilized for cross-linking interacting proteins.[116]

Most excitingly for the field of protein imaging, mutations at W37 have been shown to accommodate a bright blue fluorescent 7-hydroxycoumarin probe and to ligate this probe to LAP2-tagged proteins.[117] Despite its blue emission, the fluorescence spectrum of 7-hydroxycoumarin is compatible with imaging of cultured cells (excitation 387–405 nm; emission 448 nm).[118] Furthermore, despite its small size, the anionic form of 7-hydroxycoumarin is a brighter fluorophore (with a ϵ of $36,700 \text{ M}^{-1} \text{ cm}^{-1}$ and quantum yield of 0.7)[118] than the enhanced blue fluorescent protein (EBFP), which is commonly used for protein imaging in cells (with a ϵ of $31,500 \text{ M}^{-1} \text{ cm}^{-1}$ and quantum yield of 0.2).[119] Therefore, LAP2-conjugated coumarin represents an advantage over EBFP for protein imaging, not only because of its improved brightness, but because of the small total size of this tag. LpIA-mediated coumarin labeling has been put to use in site-specific protein labeling and fluorescence imaging both on the surface and in the cytoplasm of living cells.[117]

Importance of protein-protein interactions in cell biology

While the importance of site-specific protein labeling cannot be overstated, as we still have much to learn about the localization, trafficking, and dynamic behavior of many proteins inside living cells, we can also apply our enzyme-mediated labeling technologies to more complex problems. In particular, we are interested in studying the occurrence and dynamics of protein-protein interactions inside living mammalian cells.

Proteins rarely function individually in cellular processes.[120] The recent characterization of several interactomes, or sets of all protein-protein interactions (PPIs) occurring in a cell, demonstrates not only that many proteins undergo PPIs, but that those interactions are fundamental to the biological function of these proteins.[121-125] Some proteins exist in stable complexes inside cells, such as RNA polymerase and the nuclear pore complex.[126] Other proteins form transient associations, which form the basis of signaling pathways inside cells. Protein-protein interactions convey information from the extracellular environment into the cell interior, beginning with ligand-receptor interactions at the cell surface, then propagating the signal into the cell *via* short-lived associations of protein partners in the signaling pathway, including kinases and phosphatases.[127] It is therefore clear that the functions of many proteins are governed by their interactions with other proteins, and where and when those interactions occur inside the cell.

Reductionist methods for identifying PPIs rely on characterization of interactions between purified proteins, protein domains, or peptides *in vitro*. While such methods provide information about dissociation constants and the effects of inhibitors, they cannot report on

whether the interaction occurs in cells, where such processes are regulated by spatially and temporally regulated, and where other biomolecules may compete for binding. Therefore we focus on methods that detect interactions within the cellular milieu, be it cell lysates, fixed cells, or living cells.

In this section we will discuss current methods for the study of PPIs in the cellular context, beginning with biochemical methods used in cell lysates, then moving to methods used to identify PPIs in the heterologous context of yeast cells, then to methods used to detect interactions in fixed mammalian cells, and finally to methods for use in living mammalian cells. We will discuss the strengths and weaknesses of these methods as well as examples of their applications in order to motivate the need for improved methods.

Biochemical methods

Co-immunoprecipitation and tandem affinity purification

The most commonly used method to identify PPIs in cells is co-immunoprecipitation, or co-IP. In this experimental paradigm, a protein is precipitated from a cell lysate using resin conjugated to an antibody specific to the protein of interest. Any proteins that interact with the target protein will remain complexed during the precipitation step, and can be eluted from the beads and subsequently identified.[128] In its modern incarnation, co-IP is combined with mass spectrometry (MS)-based proteomics to identify interacting proteins.[129, 130] The assignment can then be confirmed by immunoblotting against the co-precipitated protein with an antibody specific to its suspected identity.

An advantage of co-IP is that if a specific and high-affinity antibody against a protein of interest exists, the method can be used to pull down and interrogate binders of endogenous proteins. Barring the existence of such an antibody, the protein of interest can be fused either to an epitope tag for which a high-affinity antibody exists, or to a protein such as glutathione-S-transferase that can bind a resin-conjugated small molecule to facilitate specific, high-affinity pull-downs.

Co-IP is, however, subject to false positives. First, the antibody utilized for precipitation of the protein of interest may bind to and precipitate other proteins in off-target events.[126] Second, cell lysis can generate false positive results for proteins that would normally not be co-compartmentalized or co-localized in an intact cell, but that can interact once the cell is lysed.[126] Therefore, an independent method must be used to demonstrate that any detected interaction is physiologically relevant. Finally, proteins co-precipitated may not directly interact with the protein of interest, but may be part of a larger multi-protein complex.[126]

Additionally, co-IP suffers from false negatives. In particular, due to the dilution effect of cell lysis and washing steps, as well as the lengthy purification protocols, transient and/or low-affinity interactions may not be detected.[131] Since protein concentrations inside cells can vary from 0.3pM to 1 mM, the dissociation constants of intracellular PPIs must occur over that range; but even the most sensitive purification-based methods applicable to lysates are limited to detecting interactions with dissociation constants of less than or approximately equal to 10 μ M.[126]

Tandem affinity purification, or TAP-tagging, has been developed more recently to reduce the false positives that plague co-IP. The noise reduction arises from two sequential purification steps.[132] Generally, two affinity tags are fused to the protein of interest; the tags are separated by a tobacco etch virus (TEV) protease recognition sequence. The protein and its associated interaction partners are first purified by binding to an affinity resin specific to the first part of the TAP tag; elution is performed by adding TEV protease. The complex is then purified on a resin specific for the second part of the TAP tag.[132] Unfortunately, tandem affinity purification is prone to low sensitivity due to the multiple dilution and washing steps involved, necessitating the use of very large amounts of starting material to achieve sufficient signal for analysis.[133] It therefore suffers the same limited ability to detect transient and low-affinity interactions that characterizes co-IP. Despite these limitations, TAP-tagging has recently been used to construct an interactome network of *Saccharomyces cerevisiae*. [123]

Chemical cross-linking and photo-cross-linking

In order to solve the false-negative problem of affinity purification methods due to dissociation events, covalent cross-linking methods have been developed. Classical methods for protein cross-linking utilize a reagent such as formaldehyde or a di-*N*-hydroxysuccinimide ester with a linker between the reactive groups, which can react with two proteins in close proximity to one another in biological samples (i.e., intact cells or cell lysates).[131, 132] The covalent complexes are then affinity purified from the sample and subjected to immunoblotting or MS-based analysis. Because cross-linking reagents react non-specifically with protein functional groups in

a proximity-dependent manner, cross-linking approaches are amenable to analysis of interactions of a protein of interest with endogenous proteins.

The utility and sensitivity of the chemical cross-linking approach have been demonstrated by an elegant method for covalently trapping kinases that interact with a particular phosphorylation substrate.[134, 135] A cysteine is engineered into the phosphorylated protein of interest near the phosphorylation site; all kinases have a lysine in the active site that is mechanistically required for phosphate group transfer. A cross-linking probe was designed to bind specifically in the kinase active site, either utilizing an adenine moiety or a small-molecule kinase inhibitor, bearing an *o*-phthaldialdehyde moiety. The cross-linker can then react with the kinase lysine residue and the substrate cysteine to create a cross-linked product. This approach is amenable to kinase cross-linking in cell lysates, demonstrating that cross-linking strategies can be applied to the some of the most labile of biologically relevant PPIs, kinase-substrate interactions. (Of course, the labeling chemistry is still relatively slow, requiring that the cross-linking probe be designed around a very high-affinity kinase inhibitor in order to achieve cross-linking in cell lysates, so the sensitivity and generality of this method remain to be demonstrated.) Thus cross-linking approaches can be utilized to capture transient interactions that would never be detected by co-IP or affinity purification.

Additional sensitivity in PPI detection can be achieved through the use of photo-activatable cross-linkers. While chemical cross-linking reagents are limited to reaction with electrophilic and nucleophilic substituents of amino acids, photo-activatable cross-linkers generate highly reactive intermediates that can react with essentially any functional group, including C-H

bonds.[136] Because photo-cross-linking probes can be site-specifically incorporated into proteins, they offer the additional advantage (over chemical cross-linking methods) of generating relatively few cross-linked products, simplifying analysis. Unnatural amino acid mutagenesis was the first method reported for site-specific incorporation of photo-cross-linkers into proteins, including aryl azides and benzophenones.[11] A method to enzymatically attach an aryl azide to a small peptide tag using the *E. coli* lipoic acid ligase enzyme has been recently reported by our research group.[116] Both of these methods have been demonstrated to specifically cross-link known interaction partners in cell lysates. Of course, aryl azides (and the smallest photo-activatable cross-linkers, diazirines) are inefficient cross-linkers because their photo-activation involves dissociation of nitrogen, producing nitrene or carbene intermediates, respectively; which can react non-productively with solvent, producing very low cross-linking yields[112, 116]. Benzophenones are much more efficient photo-cross-linkers because the reactive ketyl diradical intermediate can relax back to the ground state and be re-activated through multiple cycles of excitation, producing greater cross-linking yields.[137] Because these site-specific targeting methods are relatively new, application of photo-cross-linking reagents in new PPI discovery has not yet been extensively demonstrated.

Yeast genetic methods

Yeast two-hybrid

The yeast two-hybrid is a powerful method for detecting PPIs in yeast cells. In this assay, the transcription factor GAL4 is separated into two domains, a DNA-binding domain and an activator domain. If interacting proteins are each fused to one of the domains, the interaction

causes the transcription factor to reassemble, translocate to the nucleus, and activate transcription of a reporter gene.[138, 139] The yeast two-hybrid can be used either to interrogate a suspected interaction between two proteins of interest, or to screen libraries of prey proteins for interaction with a bait protein of interest. It is in library screening that the yeast two-hybrid method has found its most powerful application.

The yeast two-hybrid method suffers from a high false positive rate. False positives can arise from several sources. First, false positives can arise from off-target interactions of a particular prey. If the prey protein binds to the promoter directly, interacts with another transcription factor or protein involved in transcription, or if it interacts with the activator domain of the bait fusion construct, reporter transcription can be activated.[140] Second, false positives can arise because the proteins are being expressed in a heterologous system; if they are mislocalized or overexpressed in yeast, they may undergo interactions which would never occur in the mammalian cell.

Finally, the information that can be obtained about a PPI is limited in the yeast two-hybrid system. The method cannot report on the localization of the interaction, nor can it provide temporal information about interactions, because the readout depends on transcription and translation of a reporter gene, which requires several hours' lag time. Furthermore, because the reassembled transcription factor must be able to translocate to the nucleus, this method is inapplicable to proteins that cannot undergo this translocation, in particular membrane proteins.[141]

Despite these limitations, the yeast two-hybrid remains one of the most commonly used methods for new PPI discovery, as evidenced by its use in delineating large-scale interactome maps for *Drosophila melanogaster*,^[125] *Saccharomyces cerevisiae*,^[142] *Caenorhabditis elegans*,^[121] as well as partial interactome studies of human proteins.^[143, 144]

Split ubiquitin

A genetic reporter of PPIs that promises to improve on the two-hybrid method is split ubiquitin. In this system, ubiquitin is split into two halves that are not recognized by ubiquitin binding proteins and are not cleaved when expressed individually. When the two ubiquitin halves are fused to interacting proteins, the forced proximity (that is, the high local concentration) induced by the interaction causes the fragments of ubiquitin to associate and fold, forming an active, though non-covalently associated, ubiquitin molecule joining the interacting proteins. The folded ubiquitin is recognized by ubiquitin binding proteins and cleaved.^[145] If a reporter protein is incorporated into the construct after the cleavage site, the ubiquitin cleavage event will liberate it from the fusion construct, which can be used to read out the interaction event. The reporter in this case is an engineered transcription factor that, upon liberation, translocates to the nucleus and specifically activates transcription of an introduced transgene, such as beta-galactosidase. The subsequent colorimetric response of the β -galactosidase reports on the interaction.^[146] Because the transcription factor is freed from the interacting proteins after cleavage, the proteins of interest may reside anywhere in the cell, as long as the ubiquitin and transcription factor are localized in the cytosol. In fact, split ubiquitin has been demonstrated to detect interactions between membrane proteins, a significant advance over the yeast two-

hybrid.[146] While this method is powerful for detecting PPIs, it is performed in the heterologous context of yeast cells, and so suffers the same problems of context that the yeast two-hybrid does. Furthermore, because the readout is genetic, the method does not provide spatial or temporal information about the interaction under study.

Methods applicable to fixed mammalian cells

In situ proximity ligation

The proximity ligation *in situ* assay (P-LISA) is a method for detecting endogenous PPIs in fixed cells or tissue sections.[147] The general principle is the coupling of two DNA oligonucleotides to antibodies against the two target proteins; when the antibodies are in close proximity, the attached oligonucleotides template association of circularizable DNA “connectors”, which can be enzymatically ligated *in situ* to produce a circular DNA primer. The circular DNA can then prime rolling-circle amplification of one of the antibody-conjugated oligonucleotides. The amplified product remains attached to the antibody, and is then detected with DNA probes by fluorescence *in situ* hybridization. This method provides both sensitivity and specificity of detection; specificity arises because only when the DNA-conjugated antibodies are in close proximity is the circular primer formed, and sensitivity comes from the amplification of the rolling circle amplification product, which can produce up to several thousand repeats of the target sequence.[148]

The P-LISA method has been utilized to detect heterodimers of the transcription factors *c-Myc* and *Max* in cell lines and in clinical tissue sections.[147] In some cell lines, individual spots can be observed, which may or may not be signal from single *c-Myc/Max* interaction events.

The major advantages of the P-LISA method are (1) the ability to detect interactions of endogenous proteins, and therefore applicability to clinical histological samples, as long as a suitable antibody is available; and (2) the tremendous signal amplification arising from the DNA replication step. However, the major disadvantage is that the method requires fixation and therefore cannot be applied in living cells.

Proximity biotinylation

A method to detect PPIs in lysates and fixed mammalian cells has recently been reported by the Ting lab.[149] Proximity biotinylation adapts the BirA-mediated biotinylation of the AP peptide (see Part II of this chapter as well as Chapter 4 for more details) to create a PPI sensor. In this design, BirA is fused to one protein partner, and an AP peptide to the other. When the two proteins interact, BirA and the peptide substrate are brought into proximity, biotinylation occurs, and the signal can be read out with streptavidin staining. In order to decrease background from interaction-independent AP biotinylation, the peptide was engineered to decrease its affinity for BirA, but retain a sufficiently high k_{cat} , by deleting three amino acids from the C-terminus; the resulting peptide is called AP(-3). Proximity biotinylation utilizing BirA and AP(-3) generates, at best, a signal-to-background ratio of 28:1 in the presence of a PPI. However, the method suffers from a limited dynamic range from slow background biotinylation of AP(-3) and impaired k_{cat} of AP(-3) relative to the original AP peptide (0.53 min^{-1} instead of 12 min^{-1})[149]. Furthermore the method is limited to fixed-cell detection because of the requirement for streptavidin staining.

Methods applicable to live mammalian cells

Co-localization imaging

A common method utilized in the study of interacting proteins in both fixed and live cells is co-localization imaging. In this method, each protein is detected with a unique fluorophore; in fixed cells, this can be accomplished with fluorophore-conjugated antibodies against the endogenous proteins, while in live cells, fluorescent protein fusions can be utilized. Upon imaging, coincidence of the two colors is taken as confirmation of co-localization of the two proteins. Live-cell co-localization analysis has been applied, for example, to studying co-trafficking of synaptic proteins in neurons[150] and to the discovery of the purinosome, a presumptive complex of the proteins involved in *de novo* purine biosynthesis.[151]

However, the resolution of co-localization imaging in most conventional light-based microscopes is limited by the diffraction of light. Using the shortest near-UV excitation light that is non-toxic in cellular imaging, approximately 400 nm, the maximum theoretical achievable resolution in the focal plane is 150 nm.[152] In practice, the real achievable resolution is 200-300 nm.[153] This resolution limit is larger than some subcellular structures,[153] and certainly larger than individual proteins. Therefore coincidence of fluorescence reports only that the proteins are within the same diffraction-limited spot, not that they are directly interacting. While several new techniques have been developed for sub-diffraction imaging, they are technically challenging, require specialized microscopy instrumentation, and are currently restricted to use in fixed samples.[153]

Resonance energy transfer

As an improvement to co-localization imaging, the direct interaction of two fluorophore-tagged proteins can be detected by resonance energy transfer. The principle of resonant energy transfer is the nonradiative transfer of energy from a donor fluorophore to an acceptor fluorophore. The efficiency of energy transfer depends on the distance between the donor and acceptor, decaying as a function of the sixth power of this distance and therefore only occurring at very short donor-acceptor separation. The Förster distance is defined as the separation between the donor and acceptor giving 50% energy transfer efficiency, and for most organic fluorophores is approximately 1-10 nm.[141] This distance is within the range of the distances involved in biological protein-protein interactions.[154]

The most commonly used method to detect PPIs inside living cells, in real time and with spatial resolution, is fluorescence resonant energy transfer, or FRET. Several biologically relevant fluorophore pairs have been reported for FRET. The first pair used was the blue fluorescent protein (BFP) and the green fluorescent protein (GFP), but cell autofluorescence interferes significantly with BFP emission, generating noise.[155] The most commonly used pair is the cyan fluorescent protein, CFP, and yellow fluorescent protein, YFP.[156] Several additional pairs have been reported, including YFP/dsRED[154] and CFP/fluorescein (targeted using the FIAsh methodology).[157] At its simplest, the interrogation of PPIs can be accomplished by fusion of a donor and acceptor fluorophore to individual proteins, followed by quantitation of intermolecular FRET. The advantage of this arrangement is that the interaction can be observed associating and dissociating in real time due to the rapid FRET response, while obtaining spatial information about the interaction. However, there are severe limitations in terms of signal-to-noise attainable in intermolecular FRET. First, every reported donor-acceptor pair suffers from

intrinsic channel cross-talk.[156] Donor excitation wavelengths often directly excite the acceptor to some extent, producing low background acceptor fluorescence. Furthermore, the broad emission spectra of fluorescent proteins mean that the tail of donor emission extends into the acceptor emission channel, again producing low background signal from donor emission. This cross-talk fundamentally limits the dynamic range of intensity-based FRET measurements, since the background will never be zero. A secondary consequence of these bleed-through effects is that intermolecular FRET is most sensitive if the donor and acceptor (and thus, both interaction partners) are present in a 1:1 stoichiometry.[141] Additional copies of either donor or acceptor contribute background that may overwhelm specific FRET signal. The final limitation is the requirement for not only spatial proximity but proper orientation of the donor relative to acceptor.[156] This means that FRET is exquisitely sensitive to fusion geometry, and even in the case where proteins interact, FRET may not occur if the orientations of and distance between the donor and acceptor are not adequate.

Several developments have attempted to address the limitations of intermolecular FRET. First, FRET-FLIM, or fluorescence lifetime imaging, is a method of FRET quantitation that addresses the limitations of spectral cross-talk by eliminating the measurement of acceptor emission entirely.[155] The excited state lifetime of the donor is measured in the presence and absence of the acceptor; if FRET occurs, the lifetime of the excited state is expected to decrease.[158] This method retains spatial information and has a better dynamic range than intensity-based FRET because only the donor emission is measured, but it is technically challenging. A second method to address the stoichiometry problem is the construction of fusion proteins containing both interaction partners for intramolecular FRET. The most important examples of this PPI

reporter class are the cameleon calcium sensors.[159] These reporters are generally fusions of calmodulin to M13 peptide, a peptide derived from myosin light-chain kinase that binds Ca^{2+} , which interact upon calcium binding. CFP and YFP are fused at each end of the construct, and when calcium is present, calmodulin and M13 interact, bringing CFP and YFP into proximity and increasing FRET.[156] However, as a general design principle, presenting both interaction partners on the same molecule may promote their interaction simply because of their increased local concentration, producing false positives.[141]

Bioluminescent resonance energy transfer, or BRET, relies on energy transfer between the bioluminescent product of a luciferase enzyme with a fluorescent protein acceptor. Similar to FRET, the donor-acceptor distances producing BRET are in the range of 1-10 nm.[160] Because the donor is bioluminescent, donor excitation light is eliminated, removing background due to autofluorescence and direct acceptor excitation.[160] The dynamic range and sensitivity of BRET are therefore superior to FRET. The most common donor-acceptor pair utilized in BRET studies is *Renilla* luciferase, which produces a blue bioluminescent product from the small-molecule substrate coelenterazine, and YFP.[160] Recently, extension of BRET to far-red imaging *in vivo* has been reported using an indocyanine dye as the acceptor;[161] to make this method generally applicable, chemical methods for specific indocyanine targeting to proteins of interest must be developed.

In conclusion, FRET has found widespread application for both intermolecular and intramolecular PPI detection, though it suffers from a limited dynamic range and complicated

technical corrections for spectral overlap; BRET may begin to address these spectral limitations in the near future.

Protein complementation assays

Protein complementation assays (PCAs) are a class of PPI detection methods that make use of reporter protein reconstitution to report on PPIs. In general, a reporter protein (either an enzyme or a fluorescent protein) is split in an internal loop to create two inactive fragments. This can either be done by rational design, where known or predicted disordered regions are chosen for dissection, or by screening libraries of truncation mutants for complementation activity. The fragments are fused to each member of an interacting pair of proteins. Upon interaction of the proteins of interest, the fragments of the reporter protein associate, fold, and reconstitute the reporter functionality, be it enzymatic activity or fluorescence. The fragments must also be designed such that in the absence of a PPI, they do not spontaneously reassemble to generate reporter activity. This assay design has been applied to several methods of PPI detection in living cells.

PCAs in general suffer several limitations. One drawback is that reporter reconstitution depends on accessibility of the fragments in the interaction complex; the fusions to the proteins of interest must be constructed such that the fragments of the PCA reporter are geometrically accessible for folding and re-assembly, and failure to do so may result in false negative results. Another drawback is that reporter reconstitution is generally irreversible, meaning that the interacting proteins are trapped in complex, a non-physiological condition in which the

proteins' normal trafficking is perturbed. Finally, some suffer from background due to spontaneous reporter re-assembly, especially at high expression levels.[162]

Many proteins have been pressed into use as PCA reporters, and we will discuss each in turn.

1. Dihydrofolate reductase

The first reported PCA utilized dihydrofolate reductase (DHFR) as the reporter. This enzyme can be reassembled from inactive fragments to regenerate DHFR enzymatic activity when fused to interacting proteins.[163] The DHFR activity can then be read out either by a survival assay or by imaging using a fluorescent probe.[164] The survival assay works by expressing the DHFR fragment fusions to interaction partners in a cell line that lacks endogenous DHFR; if the proteins interact, the introduced DHFR activity allows survival in nucleotide-free media. This assay, while sensitive, provides no spatial or temporal information about the PPI under investigation. The imaging-based method utilizes fluorescein-conjugated methotrexate, an inhibitor that strongly binds DHFR, as a probe for active enzyme. Unfortunately, while this method in theory has the potential to provide spatial information about interactions, methotrexate is toxic to cells because it inhibits nucleotide biosynthesis.

Despite these limitations, the DHFR survival assay has recently been used to great effect in elucidating a complete interaction network in *Saccharomyces cerevisiae*. [124] Reconstitution of a methotrexate-resistant mutant of DHFR was utilized to interrogate interaction of the products of over 4000 yeast ORFs, and over 2700 interactions were identified. This impressive application demonstrates the power of PCAs for identifying new PPIs, even if additional biological information about the interaction cannot be acquired with the method.

2. β -galactosidase

Split β -galactosidase was developed next.[165] By splitting β -galactosidase into two fragments with no enzymatic activity and also low ability to spontaneously reassemble when coexpressed, low background is achieved; enzymatic activity is strongly induced in the context of fusion to interacting proteins. The enzyme can utilize chemiluminescent and colorimetric substrates, but substrate treatment must be performed in cell lysate or fixed cells, therefore eliminating any temporal or spatial information that can be obtained about the enzyme.

3. β -lactamase

The ampicillin resistance gene product, β -lactamase, has been utilized as a PCA sensor. The enzyme catalyzes hydrolysis of cephalosporins. A fluorogenic substrate for β -lactamase, CCF2-AM, has previously been developed for use in live cells.[166] A colorimetric β -lactamase substrate, nitrocefin, can be hydrolyzed by the enzyme to produce a red product, but nitrocefin is not membrane-permeable, so can only be used in cell lysates.[167]

In order to create a PCA reporter, the enzyme was dissected at a site shown to be amenable to linker insertion and circular permutation.[166] Reconstitution of the reporter and enzymatic activity were detected in the presence of both of these interactions both in vitro, using the nitrocefin substrate, and in living cells by microscopy using the CCF2-AM substrate, when fused to a homodimerizing leucine zipper.

The β -lactamase reporter therefore provides a colorimetric or fluorogenic response to PPI and reporter reconstitution. However, several important limitations exist. First, the background is

low but nonzero for this method in the absence of interactions, potentially producing false positive results. Second, the CCF2-AM substrate for use in living cells is diffusible, limiting the spatial information that can be obtained about the interaction.

4. Luciferase

PCAs have been reported using firefly,[168] *Renilla*,[169] and *Gaussia*[170] luciferases. The firefly luciferase utilizes molecular oxygen and ATP to oxidize a small-molecule substrate called luciferin, producing yellow bioluminescence; the *Renilla* and *Gaussia* luciferases act on the coelenterate luciferin (coelenterazine) substrate, oxidizing it with molecular oxygen to produce blue bioluminescence without any additional cofactor requirement.[171] The first report of a *Renilla* luciferase PCA was in 2003.[172] However, significant background was observed when the *Renilla* luciferase fragments were coexpressed in cells, either alone or when fused to non-interacting proteins. An improved *Renilla* luciferase PCA was reported in 2007.[169] In this design, which splits the protein in a different location, background is low (though non-zero) and bioluminescence recovery in the event of a PPI is high. Counter-intuitively, the formation of the *Renilla* luciferase reporter from its fragments is demonstrably reversible, meaning that the folded reporter can again dissociate into its fragments, permitting the interacting proteins to dissociate.

The *Gaussia* luciferase PCA was reported in 2006.[170] Two fragments of the enzyme were identified that could fold in less than 1 minute to generate 10% of the enzymatic activity of an equivalent amount of intact *Gaussia* luciferase. This reporter also demonstrates reversibility.

Firefly luciferase has been utilized for real-time PPI detection.[168] This reporter folds within 1 minute and produces 18-fold greater bioluminescence in the presence of an interaction than in the absence of one. The firefly luciferase fragments produce background bioluminescence when coexpressed in mammalian cells in the absence of any fusion, though the induced bioluminescence in the context of a fusion to an interacting pair is much higher. This reporter was used to image the phosphorylation-dependent association of Cdc25C with 14-3-3 ϵ and the constitutive homodimerization of the transcription factor STAT1 in cultured mammalian cells, and the rapamycin-dependent association of FKBP with FRB, the domain of the mammalian target of rapamycin (mTOR) that interacts with FKBP and rapamycin, in living mice.

In conclusion, split luciferase assays provide sensitive bioluminescent detection of PPIs both in living cells and in living animals, with multiple colors and orthogonal substrates available. The association of these reporter fragments is reversible, thus not trapping the complexes.

However, these assays suffer in general from background signal in the absence of a PPI that could be interpreted as false positives. Furthermore, the luciferins are diffusible, limiting the spatial information that can be obtained.

5. Fluorescent proteins

The bimolecular fluorescence complementation (BiFC) assay has found extensive application in the detection and imaging of PPIs with high sensitivity and spatial resolution on the surface of and inside living cells. The principle of BiFC rests on the fact that GFP and related fluorescent proteins contain a buried fluorophore within an eleven-stranded beta-barrel.[173] The stepwise formation of mature fluorescent GFP from a nascent polypeptide is as follows: first the protein

folds into its tertiary structure, then the fluorophore autocatalytically cyclizes, followed by oxidation of the fluorophore by molecular oxygen. If the barrel structure is not properly folded or denatured, the chromophore is non-fluorescent. In order to turn GFP into a PPI sensor, the protein is split into two pieces which do not fluoresce individually because the barrel structure is incomplete. Each piece is fused to one member of an interacting pair of proteins. When the proteins interact, the fragments of GFP associate and fold; subsequent to protein folding, the fluorophore forms and becomes oxidized, and fluorescence is reconstituted.[174]

The first and still most widely applied fluorescence complementation method was split GFP. GFP is split in the loop between beta-strands 10 and 11, leaving strands 1-10 forming the larger N-terminal fragment, and the strand 11 peptide forming the small C-terminal fragment of GFP. Split GFP has been reported to detect PPIs with affinities as weak as 1 mM.[175] The first report of split GFP reconstitution was by Ghosh *et al.*, who demonstrated that two non-fluorescent fragments of GFP could reconstitute fluorescence when fused to artificial leucine zippers and either refolded from aggregates *in vitro* or expressed inside *E. coli*.[176] These leucine zipper fusions were subsequently demonstrated to recombine in mammalian cells and allowed visualization of the subcellular localization of the interaction when directed to various organelles.[177] Ozawa *et al.* also reported reconstitution of split GFP by a slightly different method, wherein each fragment of GFP was fused to a split intein, as well as calmodulin and M13. Upon calcium-induced interaction, the intein was reconstituted and covalently recombined the fragments of GFP, yielding fluorescence.[178] Since these initial reports, split GFP has been used to image protein-protein interactions across the synaptic junction[179] and the trafficking and co-compartmentalization of bacterial pathogen proteins with various

endogenous mammalian proteins during infection;[180] it has also been used to screen randomized protein libraries for the design of new peptide binding activity,[181] among many other applications.

YFP BiFC has been developed specifically for use in visualizing PPIs inside living cells. The first report of YFP BiFC came from the Kerppola lab, who described fragments of enhanced YFP split at residue 155 to generate a large N-terminal and small C-terminal fragment, which could produce fluorescence complementation in living mammalian cells when fused to leucine zipper domains of the transcription factors Fos and Jun and accurately reported the subcellular (nuclear) localization of the interaction.[174]

Multicolor BiFC has recently been reported. Cyan fluorescent protein (CFP),[182] several red fluorescent proteins (mCherry[183] and mRFP1[184]), and a far-red fluorescent protein, mLumin,[185] have been demonstrated to support fluorescence complementation upon PPI-mediated reconstitution. Because the small C-terminal fragment of CFP and YFP used for BiFC are nearly identical, a fusion of the C-terminal CFP peptide to one interacting partner can complement either the N-terminal fragment of CFP or YFP on another interacting partner, facilitating the interrogation of competing interactions with a single protein.

BiFC displays excellent sensitivity, even to transient interactions, accurately reports subcellular localization of PPIs, works in living cells, and is available in many colors. As such it is well suited to microscopy as well as FACS-based applications. However, there are several limitations to the technique that speak to the need for new methods to image PPIs in living cells. First, fluorescence recovery for split YFP after a PPI is induced is detectable after 10 minutes, but only

quantifiable after about 1 hour and maximal 8 hours after the interaction is induced;[186] the *in vitro* half-time for fluorescence maturation is about an hour.[127] This slow maturation limits the application of the method to inducible or dynamic PPIs. Second, reconstitution of the fluorescent protein reporter from its constituent fragments is irreversible, trapping interacting proteins in complex.[175] Third, both the large and small fragments of fluorescent proteins used as reporters are generally insoluble and tend to cause aggregation of the protein to which they are fused.[186, 187] Well-folded fragments of split superfolder GFP were selected from a library generated by DNA shuffling; these fragments produce fluorescence only ten minutes after complementation, but they spontaneously reassemble, precluding their use as a reporter of PPIs.[187] Reporter misfolding presents several serious problems in detecting inducible PPIs, which we will discuss in detail in Chapter 4. Finally, the method has been reported to be prone to false positive results that arise from association, folding, and subsequent fluorescence in the absence of a PPI, especially when the constructs are expressed at very high levels (though this is dependent on the fluorescent protein employed in the reporter design, occurring, for example, for Venus but not EYFP).[162, 186]

Conclusion

From the sheer number of PPI detection methods discussed in this chapter, it should be clear that the study of known PPIs and the discovery of new PPIs are of tremendous interest to cell biologists and protein biochemists. Despite the repeated application of high-throughput co-purification and two-hybrid approaches to interactome mapping, very little overlap between detected interactions has been observed between similar studies. This is due to the high false

positive and negative rates of current methods; it has been estimated that up to 60% of the interactions detected in these studies are false positives that have not been verified by independent methods.[132] Therefore, the yeast and human interactomes are estimated to be only 50% and 10% complete, respectively,[132] and new methods to discover and characterize these remaining PPIs are needed.

The pertinent features of each PPI detection method described in this section are summarized in Table 1-3. While biochemical methods have proven historically powerful in identifying and characterizing PPIs, and indeed are still utilized to confirm any new PPI that has been putatively identified by other methods, they cannot provide information about whether an interaction occurs in intact cells, and certainly cannot provide spatial or temporal information about the interaction. Yeast genetic methods are amenable to high-throughput screening but notoriously subject to false positives. Histological methods cannot provide dynamic information about PPIs. FRET suffers from limited dynamic range and BRET cannot assess subcellular localization. Protein complementation assays suffer either high background, lack of spatial resolution, or aggregation problems. While each of these methods is powerful in its own right, no method yet satisfies all the requirements of an ideal PPI detection method. It is clear, then, that new methods are required to improve on the sensitivity and specificity of existing methods, as well as to provide improved PPI imaging modalities to gain spatial and temporal information about an interaction without significantly perturbing it. Finally, we note that even new methods for PPI detection will not likely address all limitations of current methods, and that any new interaction must be interrogated by multiple methods, because they provide different and synergistic information about PPIs.

While proximity biotinylation represented an advance in PPI detection in mammalian cells relative to PCAs because of its good sensitivity, low background, and because it does not trap interacting proteins, its dynamic range is limited and it can only be used for PPI detection in fixed cells. In Chapter 4 of this thesis, we will describe our efforts to improve on this method by utilizing LplA and the LAP peptide to create new sensors for visualizing PPIs.

Detection method	Sensitive to stoichiometry	Provides temporal information	Provides spatial information	Traps interaction	Works in living mammalian cells	Works for membrane proteins
Biochemical methods						
Co-immunoprecipitation and TAP tagging	No	No	No	No	No	Yes
Chemical and photo-cross-linking	No	No	No	No	No	Yes
Yeast genetic methods						
Yeast two-hybrid	No	No	No	No	No	No
Split ubiquitin	No	No	No	No	No	Yes
Fixed mammalian cell methods						
Proximity ligation <i>in situ</i> (P-LISA)	No	No	Yes	No	No	Yes
Proximity biotinylation	No	No	Yes	No	No	Yes
Resonant energy transfer (live mammalian cell methods)						
Fluorescence resonant energy transfer (FRET)	Yes	Yes	Yes	No	Yes	Yes
Bioluminescence resonant energy transfer (BRET)	No	Yes	Yes	No	Yes	Yes
Protein complementation assays (live mammalian cell methods)						
Split dihydrofolate reductase	No	No	No	Yes	Yes	Yes
Split β -lactamase	No	Yes	No	Yes	Yes	Yes
Split β -galactosidase	No	No	No	Yes	Yes	Yes
Split luciferase	No	Yes	No	No	Yes	Yes
BiFC	No	Moderate	Yes	Yes	Yes	Yes

Table 1-3. Comparison of current protein-protein interaction (PPI) detection methods. Note that the dihydrofolate reductase (DHFR) protein complementation assay (PCA) is described in terms of the survival assay.

References

1. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins*. Nat Methods, 2005. **2**(12): p. 905-9.
2. Lippincott-Schwartz, J. and G.H. Patterson, *Development and use of fluorescent protein markers in living cells*. Science, 2003. **300**(5616): p. 87-91.
3. Xie, X.S., et al., *Single-molecule approach to molecular biology in living bacterial cells*. Annu Rev Biophys, 2008. **37**: p. 417-44.
4. Betzig, E., et al., *Imaging intracellular fluorescent proteins at nanometer resolution*. Science, 2006. **313**(5793): p. 1642-5.
5. Dresbach, T., et al., *Synaptic targeting of neuroligin is independent of neurexin and SAP90/PSD95 binding*. Mol Cell Neurosci, 2004. **27**(3): p. 227-35.
6. Lisenbee, C.S., S.K. Karnik, and R.N. Trelease, *Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER*. Traffic, 2003. **4**(7): p. 491-501.
7. Marguet, D., et al., *Lateral diffusion of GFP-tagged H2Ld molecules and of GFP-TAP1 reports on the assembly and retention of these molecules in the endoplasmic reticulum*. Immunity, 1999. **11**(2): p. 231-40.
8. Chen, I. and A.Y. Ting, *Site-specific labeling of proteins with small molecules in live cells*. Curr Opin Biotechnol, 2005. **16**(1): p. 35-40.
9. Chin, J.W., et al., *An expanded eukaryotic genetic code*. Science, 2003. **301**(5635): p. 964-7.
10. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells*. Nat Methods, 2007. **4**(3): p. 239-44.
11. Wang, L., J. Xie, and P.G. Schultz, *Expanding the genetic code*. Annu Rev Biophys Biomol Struct, 2006. **35**: p. 225-49.
12. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo*. Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10092-7.
13. Sasso, S.P., et al., *Thermodynamic study of dihydrofolate reductase inhibitor selectivity*. Biochim Biophys Acta, 1994. **1207**(1): p. 74-9.
14. Miller, L.W., et al., *In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag*. Nat Methods, 2005. **2**(4): p. 255-7.
15. Miller, L.W., et al., *Methotrexate conjugates: a molecular in vivo protein tag*. Angew Chem Int Ed Engl, 2004. **43**(13): p. 1672-5.
16. Clackson, T., et al., *Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity*. Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10437-42.
17. Marks, K.M., P.D. Braun, and G.P. Nolan, *A general approach for chemical labeling and rapid, spatially controlled protein inactivation*. Proc Natl Acad Sci U S A, 2004. **101**(27): p. 9982-7.
18. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo*. Nat Biotechnol, 2003. **21**(1): p. 86-9.
19. Heinis, C., et al., *Evolving the substrate specificity of O6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging*. ACS Chem Biol, 2006. **1**(9): p. 575-84.
20. Los, G.V., et al., *HaloTag: a novel protein labeling technology for cell imaging and protein analysis*. ACS Chem Biol, 2008. **3**(6): p. 373-82.
21. Bonasio, R., et al., *Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots*. Proc Natl Acad Sci U S A, 2007. **104**(37): p. 14753-8.
22. Chattopadhyaya, S., et al., *Site-specific covalent labeling of proteins inside live cells using small molecule probes*. Bioorg Med Chem, 2009. **17**(3): p. 981-9.

23. Lockless, S.W. and T.W. Muir, *Traceless protein splicing utilizing evolved split inteins*. Proc Natl Acad Sci U S A, 2009. **106**(27): p. 10999-1004.
24. Griffin, B.A., S.R. Adams, and R.Y. Tsien, *Specific covalent labeling of recombinant protein molecules inside live cells*. Science, 1998. **281**(5374): p. 269-72.
25. Martin, B.R., et al., *Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity*. Nat Biotechnol, 2005. **23**(10): p. 1308-14.
26. Adams, S.R., et al., *New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications*. J Am Chem Soc, 2002. **124**(21): p. 6063-76.
27. Gaietta, G., et al., *Multicolor and electron microscopic imaging of connexin trafficking*. Science, 2002. **296**(5567): p. 503-7.
28. Adams, S.R. and R.Y. Tsien, *Preparation of the membrane-permeant biarsenicals FIAsh-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins*. Nat Protoc, 2008. **3**(9): p. 1527-34.
29. Gaietta, G.M., et al., *Golgi twins in late mitosis revealed by genetically encoded tags for live cell imaging and correlated electron microscopy*. Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17777-82.
30. Tour, O., et al., *Calcium Green FIAsh as a genetically targeted small-molecule calcium indicator*. Nat Chem Biol, 2007. **3**(7): p. 423-31.
31. Roberti, M.J., et al., *Fluorescence imaging of amyloid formation in living cells by a functional, tetracysteine-tagged alpha-synuclein*. Nat Methods, 2007. **4**(4): p. 345-51.
32. Tour, O., et al., *Genetically targeted chromophore-assisted light inactivation*. Nat Biotechnol, 2003. **21**(12): p. 1505-8.
33. Stroffekova, K., C. Proenza, and K.G. Beam, *The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins*. Pflugers Arch, 2001. **442**(6): p. 859-66.
34. Guignet, E.G., R. Hovius, and H. Vogel, *Reversible site-selective labeling of membrane proteins in live cells*. Nat Biotechnol, 2004. **22**(4): p. 440-4.
35. Hauser, C.T. and R.Y. Tsien, *A hexahistidine-Zn²⁺-dye label reveals STIM1 surface exposure*. Proc Natl Acad Sci U S A, 2007. **104**(10): p. 3693-7.
36. Ojida, A., et al., *Oligo-Asp tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins*. J Am Chem Soc, 2006. **128**(32): p. 10452-9.
37. Nonaka, H., et al., *Non-enzymatic covalent protein labeling using a reactive tag*. J Am Chem Soc, 2007. **129**(51): p. 15777-9.
38. Marks, K.M., M. Rosinov, and G.P. Nolan, *In vivo targeting of organic calcium sensors via genetically selected peptides*. Chem Biol, 2004. **11**(3): p. 347-56.
39. McCann, C.M., et al., *Peptide tags for labeling membrane proteins in live cells with multiple fluorophores*. Biotechniques, 2005. **38**(6): p. 945-52.
40. Lambalot, R.H., et al., *A new enzyme superfamily - the phosphopantetheinyl transferases*. Chem Biol, 1996. **3**(11): p. 923-36.
41. Yin, J., et al., *Labeling proteins with small molecules by site-specific posttranslational modification*. J Am Chem Soc, 2004. **126**(25): p. 7754-5.
42. Yin, J., et al., *Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase*. Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15815-20.
43. Kropf, M., et al., *Subunit-specific surface mobility of differentially labeled AMPA receptor subunits*. Eur J Cell Biol, 2008. **87**(10): p. 763-78.
44. Yin, J., et al., *Single-cell FRET imaging of transferrin receptor trafficking dynamics by Sfp-catalyzed, site-specific protein labeling*. Chem Biol, 2005. **12**(9): p. 999-1006.

45. Carrico, I.S., B.L. Carlson, and C.R. Bertozzi, *Introducing genetically encoded aldehydes into proteins*. *Nature Chemical Biology*, 2007. **3**(6): p. 321-322.
46. Wu, P., et al., *Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag*. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. **106**(9): p. 3000-3005.
47. Duckworth, B.P., et al., *Selective labeling of proteins by using protein farnesyltransferase*. *Chembiochem*, 2007. **8**(1): p. 98-105.
48. Lin, C.W. and A.Y. Ting, *Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells*. *Journal of the American Chemical Society*, 2006. **128**(14): p. 4542-4543.
49. Chapman-Smith, A. and J.E. Cronan, Jr., *Molecular biology of biotin attachment to proteins*. *J.Nutr.*, 1999. **129**(2S Suppl): p. 477S-484S.
50. Beckett, D., *The Escherichia coli biotin regulatory system: a transcriptional switch*. *J.Nutr.Biochem.*, 2005. **16**(7): p. 411-415.
51. Cronan, J.E., Jr., *Biotinylation of proteins in vivo. A post-translational modification to label, purify, and study proteins*. *J.Biol.Chem.*, 1990. **265**(18): p. 10327-10333.
52. Xu, Y. and D. Beckett, *Kinetics of biotinyl-5'-adenylate synthesis catalyzed by the Escherichia coli repressor of biotin biosynthesis and the stability of the enzyme-product complex*. *Biochemistry*, 1994. **33**(23): p. 7354-7360.
53. Xu, Y. and D. Beckett, *Biotinyl-5'-adenylate synthesis catalyzed by Escherichia coli repressor of biotin biosynthesis*. *Methods Enzymol.*, 1997. **279**: p. 405-421.
54. Stanley, J.S., J.B. Griffin, and J. Zemleni, *Biotinylation of histones in human cells. Effects of cell proliferation*. *Eur.J.Biochem.*, 2001. **268**(20): p. 5424-5429.
55. Healy, S., et al., *Nonenzymatic biotinylation of histone H2A*. *Protein Sci.*, 2009. **18**(2): p. 314-328.
56. Healy, S., et al., *Biotin is not a natural histone modification*. *Biochim.Biophys.Acta*, 2009. **1789**(11-12): p. 719-733.
57. Chapman-Smith, A., et al., *Molecular recognition in a post-translational modification of exceptional specificity. Mutants of the biotinylated domain of acetyl-CoA carboxylase defective in recognition by biotin protein ligase*. *J.Biol.Chem.*, 1999. **274**(3): p. 1449-1457.
58. Beckett, D., E. Kovaleva, and P.J. Schatz, *A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation*. *Protein Sci.*, 1999. **8**(4): p. 921-929.
59. de Boer, E., et al., *Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice*. *Proc.Natl.Acad.Sci.U.S.A.*, 2003. **100**(13): p. 7480-7485.
60. Chen, I., Y.A. Choi, and A.Y. Ting, *Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins*. *J.Am.Chem.Soc.*, 2007. **129**(20): p. 6619-6625.
61. Chen, I., et al., *Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase*. *Nat.Methods*, 2005. **2**(2): p. 99-104.
62. Howarth, M., et al., *Targeting quantum dots to surface proteins in living cells with biotin ligase*. *Proc.Natl.Acad.Sci.U.S.A.*, 2005. **102**(21): p. 7583-7588.
63. Howarth, M., et al., *A monovalent streptavidin with a single femtomolar biotin binding site*. *Nat.Methods*, 2006. **3**(4): p. 267-273.
64. Howarth, M., et al., *Monovalent, reduced-size quantum dots for imaging receptors on living cells*. *Nat Methods*, 2008. **5**(5): p. 397-9.
65. McNeill, E., I. Chen, and A.Y. Ting, *Synthesis of a ketone analogue of biotin via the intramolecular Pauson-Khand reaction*. *Org.Lett.*, 2006. **8**(20): p. 4593-4595.

66. Sletten, E.M. and C.R. Bertozzi, *Bioorthogonal chemistry: fishing for selectivity in a sea of functionality*. *Angew.Chem.Int.Ed Engl.*, 2009. **48**(38): p. 6974-6998.
67. Rostovtsev, V.V., et al., *A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes*. *Angew.Chem.Int.Ed Engl.*, 2002. **41**(14): p. 2596-2599.
68. Wang, Q., et al., *Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition*. *J.Am.Chem.Soc.*, 2003. **125**(11): p. 3192-3193.
69. Saxon, E. and C.R. Bertozzi, *Cell surface engineering by a modified Staudinger reaction*. *Science*, 2000. **287**(5460): p. 2007-10.
70. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*. *Proc Natl Acad Sci U S A*, 2002. **99**(1): p. 19-24.
71. Agard, N.J., et al., *A comparative study of bioorthogonal reactions with azides*. *ACS Chem.Biol.*, 2006. **1**(10): p. 644-648.
72. Liu, W., et al., *Compact biocompatible quantum dots functionalized for cellular imaging*. *J Am Chem Soc*, 2008. **130**(4): p. 1274-84.
73. Jaiswal, J.K., et al., *Long-term multiple color imaging of live cells using quantum dot bioconjugates*. *Nat Biotechnol*, 2003. **21**(1): p. 47-51.
74. Ruan, G., et al., *Imaging and tracking of tat peptide-conjugated quantum dots in living cells: new insights into nanoparticle uptake, intracellular transport, and vesicle shedding*. *J Am Chem Soc*, 2007. **129**(47): p. 14759-66.
75. Derfus, A.M., W.C.W. Chan, and S.N. Bhatia, *Intracellular delivery of quantum dots for live cell labeling and organelle tracking*. *Advanced Materials*, 2004. **16**(12): p. 961-+.
76. Janes, P.W., et al., *Cytoplasmic Relaxation of Active Eph Controls Ephrin Shedding by ADAM10*. *Plos Biology*, 2009. **7**(10): p. -.
77. Chen, X., et al., *A cell nanoinjector based on carbon nanotubes*. *Proc Natl Acad Sci U S A*, 2007. **104**(20): p. 8218-22.
78. Kim, B.Y., et al., *Biodegradable quantum dot nanocomposites enable live cell labeling and imaging of cytoplasmic targets*. *Nano Lett*, 2008. **8**(11): p. 3887-92.
79. Chan, W.C. and S. Nie, *Quantum dot bioconjugates for ultrasensitive nonisotopic detection*. *Science*, 1998. **281**(5385): p. 2016-8.
80. Schroeder, J.E., et al., *Folate-mediated tumor cell uptake of quantum dots entrapped in lipid nanoparticles*. *J Control Release*, 2007. **124**(1-2): p. 28-34.
81. Bharali, D.J., et al., *Folate-receptor-mediated delivery of InP quantum dots for bioimaging using confocal and two-photon microscopy*. *J Am Chem Soc*, 2005. **127**(32): p. 11364-71.
82. Murriel, C.L. and S.F. Dowdy, *Influence of protein transduction domains on intracellular delivery of macromolecules*. *Expert Opin Drug Deliv*, 2006. **3**(6): p. 739-46.
83. Cronican, J.J., et al., *Potent Delivery of Functional Proteins into Mammalian Cells in Vitro and in Vivo Using a Supercharged Protein*. *ACS Chem Biol*, 2010.
84. McNaughton, B.R., et al., *Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins*. *Proc Natl Acad Sci U S A*, 2009. **106**(15): p. 6111-6.
85. Delehanty, J.B., et al., *Self-assembled quantum dot-peptide bioconjugates for selective intracellular delivery*. *Bioconjug Chem*, 2006. **17**(4): p. 920-7.
86. Nitin, N., et al., *Tat peptide is capable of importing large nanoparticles across nuclear membrane in digitonin permeabilized cells*. *Ann Biomed Eng*, 2009. **37**(10): p. 2018-27.
87. Mattheakis, L.C., et al., *Optical coding of mammalian cells using semiconductor quantum dots*. *Anal Biochem*, 2004. **327**(2): p. 200-8.

88. Chang, J.C., H.L. Su, and S.H. Hsu, *The use of peptide-delivery to protect human adipose-derived adult stem cells from damage caused by the internalization of quantum dots*. *Biomaterials*, 2008. **29**(7): p. 925-36.
89. Rozenzhak, S.M., et al., *Cellular internalization and targeting of semiconductor quantum dots*. *Chem Commun (Camb)*, 2005(17): p. 2217-9.
90. Biju, V., et al., *Quantum dot-insect neuropeptide conjugates for fluorescence imaging, transfection, and nucleus targeting of living cells*. *Langmuir*, 2007. **23**(20): p. 10254-61.
91. Delehanty, J.B., et al., *Delivering quantum dot-peptide bioconjugates to the cellular cytosol: escaping from the endolysosomal system*. *Integr Biol (Camb)*, 2010. **2**(5-6): p. 265-77.
92. Akinc, A., et al., *Exploring polyethylenimine-mediated DNA transfection and the proton sponge hypothesis*. *J Gene Med*, 2005. **7**(5): p. 657-63.
93. Sonawane, N.D., F.C. Szoka, Jr., and A.S. Verkman, *Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes*. *J Biol Chem*, 2003. **278**(45): p. 44826-31.
94. Duan, H. and S. Nie, *Cell-penetrating quantum dots based on multivalent and endosome-disrupting surface coatings*. *J Am Chem Soc*, 2007. **129**(11): p. 3333-8.
95. Mok, H., J.W. Park, and T.G. Park, *Enhanced intracellular delivery of quantum dot and adenovirus nanoparticles triggered by acidic pH via surface charge reversal*. *Bioconjug Chem*, 2008. **19**(4): p. 797-801.
96. Derfus, A.M., et al., *Targeted quantum dot conjugates for siRNA delivery*. *Bioconjug Chem*, 2007. **18**(5): p. 1391-6.
97. Yoo, J., et al., *Intracellular imaging of targeted proteins labeled with quantum dots*. *Exp Cell Res*, 2008. **314**(19): p. 3563-9.
98. Kam, N.W. and H. Dai, *Carbon nanotubes as intracellular protein transporters: generality and biological functionality*. *J Am Chem Soc*, 2005. **127**(16): p. 6021-6.
99. Dhar, S., et al., *Targeted single-wall carbon nanotube-mediated Pt(IV) prodrug delivery using folate as a homing device*. *J Am Chem Soc*, 2008. **130**(34): p. 11467-76.
100. Bottini, M., et al., *Full-length single-walled carbon nanotubes decorated with streptavidin-conjugated quantum dots as multivalent intracellular fluorescent nanoprobe*. *Biomacromolecules*, 2006. **7**(8): p. 2259-63.
101. Okada, C.Y. and M. Rechsteiner, *Introduction of macromolecules into cultured mammalian cells by osmotic lysis of pinocytotic vesicles*. *Cell*, 1982. **29**(1): p. 33-41.
102. Courty, S., et al., *Tracking individual kinesin motors in living cells using single quantum-dot imaging*. *Nano Letters*, 2006. **6**(7): p. 1491-1495.
103. Pierobon, P., et al., *Velocity, processivity, and individual steps of single myosin V molecules in live cells*. *Biophys J*, 2009. **96**(10): p. 4268-75.
104. Sandvig, K. and B. van Deurs, *Delivery into cells: lessons learned from plant and bacterial toxins*. *Gene Ther*, 2005. **12**(11): p. 865-72.
105. Chakraborty, S.K., et al., *Cholera toxin B conjugated quantum dots for live cell labeling*. *Nano Letters*, 2007. **7**(9): p. 2618-26.
106. Tekle, C., et al., *Cellular trafficking of quantum dot-ligand bioconjugates and their induction of changes in normal routing of unconjugated ligands*. *Nano Lett*, 2008. **8**(7): p. 1858-65.
107. Walev, I., et al., *Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O*. *Proc Natl Acad Sci U S A*, 2001. **98**(6): p. 3185-90.
108. Xing, Y., et al., *Molecular profiling of single cancer cells and clinical tissue specimens with semiconductor quantum dots*. *Int J Nanomedicine*, 2006. **1**(4): p. 473-81.

109. Morris, T.W., K.E. Reed, and J.E. Cronan, Jr., *Lipoic acid metabolism in Escherichia coli: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein*. J.Bacteriol., 1995. **177**(1): p. 1-10.
110. Green, D.E., et al., *Purification and properties of the lipoate protein ligase of Escherichia coli*. Biochem.J., 1995. **309 (Pt 3)**: p. 853-862.
111. Puthenveetil, S., et al., *Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase*. J.Am.Chem.Soc., 2009. **131**(45): p. 16430-16438.
112. Fernandez-Suarez, M., et al., *Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes*. Nat.Biotechnol., 2007. **25**(12): p. 1483-1487.
113. Kim, D.J., et al., *Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains*. J.Biol.Chem., 2005. **280**(45): p. 38081-38089.
114. McManus, E., B.F. Luisi, and R.N. Perham, *Structure of a putative lipoate protein ligase from Thermoplasma acidophilum and the mechanism of target selection for post-translational modification*. J.Mol.Biol., 2006. **356**(3): p. 625-637.
115. Fujiwara, K., et al., *Crystal structure of lipoate-protein ligase A from Escherichia coli. Determination of the lipoic acid-binding site*. J.Biol.Chem., 2005. **280**(39): p. 33645-33651.
116. Baruah, H., et al., *An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking*. Angew.Chem.Int.Ed Engl., 2008. **47**(37): p. 7018-7021.
117. Uttamapinant, C., et al., *A fluorophore ligase for site-specific protein labeling inside living cells*. Proc Natl Acad Sci U S A, 2010. **107**(24): p. 10914-9.
118. Sun, W.C., K.R. Gee, and R.P. Haugland, *Synthesis of novel fluorinated coumarins: excellent UV-light excitable fluorescent dyes*. Bioorg Med Chem Lett, 1998. **8**(22): p. 3107-10.
119. Yang, T.T., et al., *Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein*. J Biol Chem, 1998. **273**(14): p. 8212-6.
120. Morell, M., S. Ventura, and F.X. Aviles, *Protein complementation assays: Approaches for the in vivo analysis of protein interactions*. Febs Letters, 2009. **583**(11): p. 1684-1691.
121. Li, S., et al., *A map of the interactome network of the metazoan C. elegans*. Science, 2004. **303**(5657): p. 540-543.
122. Butland, G., et al., *Interaction network containing conserved and essential protein complexes in Escherichia coli*. Nature, 2005. **433**(7025): p. 531-537.
123. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae*. Nature, 2006. **440**(7084): p. 637-643.
124. Tarassov, K., et al., *An in vivo map of the yeast protein interactome*. Science, 2008. **320**(5882): p. 1465-1470.
125. Giot, L., et al., *A protein interaction map of Drosophila melanogaster*. Science, 2003. **302**(5651): p. 1727-1736.
126. Phizicky, E.M. and S. Fields, *Protein-protein interactions: methods for detection and analysis*. Microbiol.Rev., 1995. **59**(1): p. 94-123.
127. Kerppola, T.K., *Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells*. Annu.Rev.Biophys., 2008. **37**: p. 465-487.
128. Elion, E.A., *Detection of protein-protein interactions by coprecipitation*. Curr Protoc Mol Biol, 2006. **Chapter 20**: p. Unit20 5.
129. Downard, K.M., *Ions of the interactome: the role of MS in the study of protein interactions in proteomics and structural biology*. Proteomics, 2006. **6**(20): p. 5374-84.
130. Chang, I.F., *Mass spectrometry-based proteomic analysis of the epitope-tag affinity purified protein complexes in eukaryotes*. Proteomics, 2006. **6**(23): p. 6158-66.

131. Miernyk, J.A. and J.J. Thelen, *Biochemical approaches for discovering protein-protein interactions*. Plant Journal, 2008. **53**(4): p. 597-609.
132. Berggard, T., S. Linse, and P. James, *Methods for the detection and analysis of protein-protein interactions*. Proteomics, 2007. **7**(16): p. 2833-2842.
133. Collins, M.O. and J.S. Choudhary, *Mapping multiprotein complexes by affinity purification and mass spectrometry*. Current Opinion in Biotechnology, 2008. **19**(4): p. 324-330.
134. Maly, D.J., J.A. Allen, and K.M. Shokat, *A mechanism-based cross-linker for the identification of kinase-substrate pairs*. J Am Chem Soc, 2004. **126**(30): p. 9160-1.
135. Statsuk, A.V., et al., *Tuning a three-component reaction for trapping kinase substrate complexes*. J Am Chem Soc, 2008. **130**(51): p. 17568-74.
136. Brunner, J., *New photolabeling and crosslinking methods*. Annual Review of Biochemistry, 1993. **62**: p. 483-514.
137. Dorman, G. and G.D. Prestwich, *Benzophenone photophores in biochemistry*. Biochemistry, 1994. **33**(19): p. 5661-73.
138. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions*. Nature, 1989. **340**(6230): p. 245-6.
139. Chien, C.T., et al., *The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest*. Proc Natl Acad Sci U S A, 1991. **88**(21): p. 9578-82.
140. Luban, J. and S.P. Goff, *The yeast two-hybrid system for studying protein-protein interactions*. Curr Opin Biotechnol, 1995. **6**(1): p. 59-64.
141. Villalobos, V., S. Naik, and D. Piwnicka-Worms, *Current state of imaging protein-protein interactions in vivo with genetically encoded reporters*. Annual Review of Biomedical Engineering, 2007. **9**: p. 321-349.
142. Yu, H., et al., *High-quality binary protein interaction map of the yeast interactome network*. Science, 2008. **322**(5898): p. 104-110.
143. Stelzl, U., et al., *A human protein-protein interaction network: A resource for annotating the proteome*. Cell, 2005. **122**(6): p. 957-968.
144. Rual, J.F., et al., *Towards a proteome-scale map of the human protein-protein interaction network*. Nature, 2005. **437**(7062): p. 1173-1178.
145. Johnsson, N. and A. Varshavsky, *Split Ubiquitin As A Sensor of Protein Interactions In-Vivo*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(22): p. 10340-10344.
146. Stagljar, I., et al., *A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(9): p. 5187-5192.
147. Soderberg, O., et al., *Direct observation of individual endogenous protein complexes in situ by proximity ligation*. Nature Methods, 2006. **3**(12): p. 995-1000.
148. Baner, J., et al., *Signal amplification of padlock probes by rolling circle replication*. Nucleic Acids Res, 1998. **26**(22): p. 5073-8.
149. Fernandez-Suarez, M., T.S. Chen, and A.Y. Ting, *Protein-protein interaction detection in vitro and in cells by proximity biotinylation*. J.Am.Chem.Soc., 2008. **130**(29): p. 9251-9253.
150. Gerrow, K., et al., *A preformed complex of postsynaptic proteins is involved in excitatory synapse development*. Neuron, 2006. **49**(4): p. 547-562.
151. An, S., et al., *Reversible compartmentalization of de novo purine biosynthetic complexes in living cells*. Science, 2008. **320**(5872): p. 103-6.
152. Hell, S.W., M. Dyba, and S. Jakobs, *Concepts for nanoscale resolution in fluorescence microscopy*. Current Opinion in Neurobiology, 2004. **14**(5): p. 599-609.

153. Huang, B., M. Bates, and X.W. Zhuang, *Super-Resolution Fluorescence Microscopy*. Annual Review of Biochemistry, 2009. **78**: p. 993-1016.
154. Ciruela, F., *Fluorescence-based methods in the study of protein-protein interactions in living cells*. Curr.Opin.Biotechnol., 2008. **19**(4): p. 338-343.
155. Truong, K. and M. Ikura, *The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo*. Curr.Opin.Struct.Biol., 2001. **11**(5): p. 573-578.
156. Miyawaki, A., *Visualization of the spatial and temporal dynamics of intracellular signaling*. Developmental Cell, 2003. **4**(3): p. 295-305.
157. Hoffmann, C., et al., *A FIAsh-based FRET approach to determine G protein-coupled receptor activation in living cells*. Nat Methods, 2005. **2**(3): p. 171-6.
158. Wallrabe, H. and A. Periasamy, *Imaging protein molecules using FRET and FLIM microscopy*. Curr Opin Biotechnol, 2005. **16**(1): p. 19-27.
159. Miyawaki, A., et al., *Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin*. Nature, 1997. **388**(6645): p. 882-7.
160. Pfleger, K.D. and K.A. Eidne, *Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET)*. Nat Methods, 2006. **3**(3): p. 165-74.
161. Wu, C., et al., *In vivo far-red luminescence imaging of a biomarker based on BRET from Cypridina bioluminescence to an organic dye*. Proc.Natl.Acad.Sci.U.S.A, 2009. **106**(37): p. 15599-15603.
162. Kerppola, T.K., *Visualization of molecular interactions by fluorescence complementation*. Nature Reviews Molecular Cell Biology, 2006. **7**(6): p. 449-456.
163. Pelletier, J.N., F.X. Campbell-Valois, and S.W. Michnick, *Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(21): p. 12141-12146.
164. Remy, I. and S.W. Michnick, *Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(10): p. 5394-5399.
165. Rossi, F., C.A. Charlton, and H.M. Blau, *Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation*. Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8405-10.
166. Zlokarnik, G., et al., *Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter*. Science, 1998. **279**(5347): p. 84-88.
167. Remy, I., G. Ghaddar, and S.W. Michnick, *Using the beta-lactamase protein-fragment complementation assay to probe dynamic protein-protein interactions*. Nat.Protoc., 2007. **2**(9): p. 2302-2306.
168. Luker, K.E., et al., *Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals*. Proc.Natl.Acad.Sci.U.S.A, 2004. **101**(33): p. 12288-12293.
169. Stefan, E., et al., *Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo*. Proc.Natl.Acad.Sci.U.S.A, 2007. **104**(43): p. 16916-16921.
170. Remy, I. and S.W. Michnick, *A highly sensitive protein-protein interaction assay based on Gaussia luciferase*. Nat.Methods, 2006. **3**(12): p. 977-979.
171. Prescher, J.A. and C.H. Contag, *Guided by the light: visualizing biomolecular processes in living animals with bioluminescence*. Current Opinion in Chemical Biology, 2010. **14**(1): p. 80-89.

172. Paulmurugan, R. and S.S. Gambhir, *Monitoring protein-protein interactions using split synthetic renilla luciferase protein-fragment-assisted complementation*. *Anal Chem*, 2003. **75**(7): p. 1584-9.
173. Tsien, R.Y., *The green fluorescent protein*. *Annu.Rev.Biochem.*, 1998. **67**: p. 509-544.
174. Hu, C.D., Y. Chinenov, and T.K. Kerppola, *Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation*. *Mol.Cell*, 2002. **9**(4): p. 789-798.
175. Magliery, T.J., et al., *Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism*. *J.Am.Chem.Soc.*, 2005. **127**(1): p. 146-157.
176. Ghosh, I., A.D. Hamilton, and L. Regan, *Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein*. *Journal of the American Chemical Society*, 2000. **122**(23): p. 5658-5659.
177. Zhang, S.F., C. Ma, and M. Chalfie, *Combinatorial marking of cells and organelles with reconstituted fluorescent proteins*. *Cell*, 2004. **119**(1): p. 137-+.
178. Ozawa, T., et al., *A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing*. *Analytical Chemistry*, 2000. **72**(21): p. 5151-5157.
179. Feinberg, E.H., et al., *GFP Reconstitution Across Synaptic Partners (GRASP) defines cell contacts and synapses in living nervous systems*. *Neuron*, 2008. **57**(3): p. 353-363.
180. Van Engelenburg, S.B. and A.E. Palmer, *Imaging type-III secretion reveals dynamics and spatial segregation of Salmonella effectors*. *Nat.Methods*, 2010. **7**(4): p. 325-330.
181. Jackrel, M.E., et al., *Screening Libraries To Identify Proteins with Desired Binding Activities Using a Split-GFP Reassembly Assay*. *ACS Chem Biol*, 2010.
182. Hu, C.D. and T.K. Kerppola, *Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis*. *Nat.Biotechnol.*, 2003. **21**(5): p. 539-545.
183. Fan, J.Y., et al., *Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein-protein interactions in living cells*. *Biochem.Biophys.Res.Commun.*, 2008. **367**(1): p. 47-53.
184. Jach, G., et al., *An improved mRFP1 adds red to bimolecular fluorescence complementation*. *Nat.Methods*, 2006. **3**(8): p. 597-600.
185. Chu, J., et al., *A novel far-red bimolecular fluorescence complementation system that allows for efficient visualization of protein interactions under physiological conditions*. *Biosens.Bioelectron.*, 2009. **25**(1): p. 234-239.
186. Robida, A.M. and T.K. Kerppola, *Bimolecular fluorescence complementation analysis of inducible protein interactions: effects of factors affecting protein folding on fluorescent protein fragment association*. *J.Mol.Biol.*, 2009. **394**(3): p. 391-409.
187. Cabantous, S. and G.S. Waldo, *In vivo and in vitro protein solubility assays using split GFP*. *Nat.Methods*, 2006. **3**(10): p. 845-854.

Chapter 2: Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species

Reproduced in part with permission from *Journal of the American Chemical Society*
S. A. Slavoff, I. Chen, Y.-A. Choi, and A. Y. Ting. *Journal of the American Chemical Society* **2008**
130, 1160-1162.

Destihobiotin azide and propargyl biotin were synthesized by Dr. Irwin Chen. Iodouracil valeric acid and thiouracil valeric acid were synthesized by Dr. Yoon-Aa Choi.

Introduction.

E. coli biotin ligase, BirA, has been previously utilized by the Ting research group for labeling cell-surface proteins with biotin, followed by detection with streptavidin-conjugated quantum dots for single-molecule imaging, and with ketone biotin, which can be derivatized in chemoselective fashion with hydroxylamine probes.[1, 2] The enzyme selectively ligates biotin and ketone biotin to its 15-amino acid acceptor peptide, the AP, which can be genetically fused to proteins of interest (Figure 2-1). These reactions, much like the natural reaction catalyzed by BirA, attachment of biotin to the biotin carboxyl carrier protein (BCCP) of acetyl CoA carboxylase, proceed in two Mg^{2+} -dependent steps, where the small molecule acid is first activated using ATP as the adenylate ester, followed by attack of the activated intermediate by a specific lysine residue of the peptide of protein substrate to form the amide bond in the product.

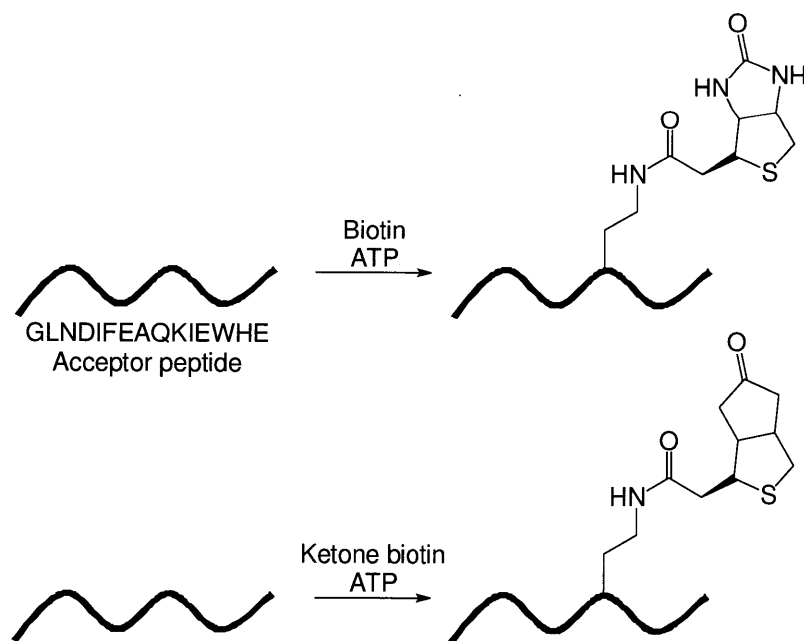


Figure 2-1. *E. coli* biotin ligase (BirA)-mediated ligation of biotin and ketone biotin to the acceptor peptide (AP). Top, biotin ligation; bottom, ketone biotin ligation. The sequence of the acceptor peptide is provided; the unique lysine is the point of attachment. Both reactions require ATP for generation of the reactive adenylate intermediate.

While these two labeling schemes have proved useful for the study of cell surface proteins, they cannot currently be applied to intracellular proteins. The ketone derivatization will be inefficient inside the cell, where many ketones and aldehydes are present in metabolites; streptavidin and quantum dots are too large to cross the cell membrane. We therefore sought new labeling chemistries to expand the utility of biotin ligase-mediated protein labeling.

E. coli BirA has low tolerance for biotin analog ligation to the AP peptide

In this project we sought to incorporate a wider variety of functional groups, such as azides and alkynes, which can be derivatized with probes using bio-orthogonal chemistry, as well as probes containing fluorophores and photo-crosslinkers for one-step labeling, using BirA. However, the efforts of many members of the Ting research group failed to identify BirA activity toward a wide variety of probes.[3, 4] Furthermore, efforts to mutagenize the BirA active site, as well as selections using randomized BirA libraries and *in vitro* compartmentalization, failed to produce an enzyme that could ligate any new probes from a library of biotin analogs onto the AP.[3, 4] This strong preference of BirA for biotin, as well as our difficulty in engineering BirA to be more permissive for its small-molecule substrate, eventually led us to a different approach: we decided to screen biotin ligase enzymes from other species to see if we could identify an enzyme with increased small-molecule permissivity. Such differential substrate specificity among homologous enzymes from different species has been previously observed.[5, 6] Designing our study required advance knowledge of the general properties of biotin ligase enzymes, including their biochemical properties, structures, and substrate specificity, which we will review here.

Properties of biotin ligase enzymes across evolutionary space

I. Specificity of biotin ligase enzymes toward protein substrates

Despite exhibiting extremely high specificity toward the proteins they modify (discussed in detail in Chapter 1), biotin ligase homologs have the fascinating property of biotinylating carboxylase domains from other species.[7] This is a result of the extremely high sequence conservation among both biotinyl domains and biotin ligase catalytic domains across evolution.[7] The functional equivalence of biotin ligases from multiple species in biotinylating each other's BCCPs has been demonstrated by complementation assays, in which a temperature-sensitive mutation of *E. coli* BirA is rescued by biotin ligases from other species, as well as direct detection of biotinylation of the *E. coli* BCCP by other biotin ligase enzymes (vide infra for more details).

II. Monofunctional and bifunctional biotin ligases

In some species, the biotin ligase enzyme has a second function, transcriptional regulation of biotin biosynthesis. The best-studied example is the *E. coli* system. *E. coli* BirA, in addition to its large and strongly conserved catalytic domain, contains a small winged helix-turn-helix N-terminal domain that can bind the biotin operator, a specific DNA sequence upstream of the biotin biosynthetic operon.[8] Because of the ordered stepwise mechanism of BirA, discussed in detail in Part II of Chapter 1, free biotin is constantly converted to the reaction intermediate biotinyl-5'-AMP, or bio-5'-AMP. If excess free biotin is present but all BCCP present in the cell is biotinylated, the bio-5'-AMP intermediate remains bound with high affinity in the enzyme active site. The intermediate-bound enzyme then dimerizes, and the dimer binds the biotin operator

to repress transcription of the biotin biosynthetic operon.[9] Because bio-5'-AMP-bound BirA homodimerizes along the same interface at which apo-BCCP (that is, unbiotinylated BCCP) binds during catalysis, the kinetic partitioning between catalytic turnover and homodimerization, followed by DNA binding, is controlled by competing protein-protein interactions, and therefore depends on the concentration of apo-BCCP in the cell.[10] Therefore BirA acts as a sensor of biotin status and creates negative feedback to prevent excess biotin biosynthesis when the vitamin levels are adequate.

Not all biotin ligases are bifunctional: two classes of enzymes exist, monofunctional enzymes that catalyze biotin transfer and bifunctional enzymes that also have a DNA-binding domain. A computational analysis of all sequenced bacterial and archaeal genomes revealed that orthologs of bifunctional biotin ligase (that is, BirA orthologs with predicted DNA-binding domains) are common among archaea as well as several groups of bacteria, such as the Bacillus/Clostridium group, and the gamma-proteobacteria .[11] Several species were also found to encode two biotin ligase paralogs, one monofunctional and one bifunctional.

Therefore it is clear that there is tremendous variation across evolution in the regulatory function of biotin ligases. While DNA-binding capacity does not directly affect the catalytic function of biotin ligases, it exemplifies the structural and functional diversity of these enzymes across species.

III. Previously reported biotin ligase enzymes from other species: similarities and differences in biochemical properties

While *E. coli* BirA is by far the best-characterized biotin ligase, many homologs have been reported. All biotin ligase enzymes contain a strongly conserved catalytic domain (vide infra for an alignment of biotin ligase catalytic domains relevant to this project). However, they vary widely in other biochemical parameters, including substrate specificity (toward both protein and small molecule substrates), domain structure and DNA binding properties, oligomerization properties, and known rate constants. We briefly summarize the previously reported biotin ligase homologs here, and provide a summary of their relevant biochemical properties in Table 2-1.

Several bacterial biotin ligases have been previously reported. The previously cloned, though not purified, biotin ligase from *Bacillus subtilis*, a widely distributed bacterium that can be found in many soils and aquatic environments,[12] has 27% sequence identity with the *E. coli* BirA, and is also bifunctional.[13, 14] The monomeric, monofunctional biotin ligases from *Mycobacterium tuberculosis*[15], the bacterium that causes tuberculosis, and *Aquifex aeolicus*,[16, 17] a hyperthermophilic bacterium, have been cloned and purified. The biotin ligase enzymes from *Propionibacterium shermanii*,[18] the bacterium utilized in fermenting Swiss cheese,[19] and *Bacillus stearothermophilus* [20] a thermophilic bacterium, have been purified from the organisms and biochemically characterized, though their DNA-binding capacity has not been directly investigated. Interestingly, while most biotin ligase enzymes require Mg^{2+} to support catalysis, the *P. shermanii* enzyme preferentially utilizes Zn^{2+} . [18] *Leuconostoc mesenteroides* is a bacterium used in industrial dextran production;[21] it is also used in starter cultures for Manchego cheese production.[22] While its biotin ligase enzyme,

which has no predicted DNA-binding domain, has not been previously cloned or purified, the growth requirements of the bacterium toward biotin and various biotin analogs have been investigated.[23] While most organisms can utilize desthiobiotin, or DTB, in place of biotin, both as a substrate for biotin ligase and as a cofactor in carboxylation reactions, DTB usually supports at least two times slower growth of the organism in culture. Furthermore, the biotin analog γ -(3,4-ureylenecyclohexyl)butyric acid, in which the thiophene ring of biotin has been replaced with a cyclohexyl ring, normally inhibits growth. However, in *Leuconostoc*, DTB supports a growth rate equivalent to biotin, and the organism can grow in γ -(3,4-ureylenecyclohexyl)butyric acid when this analog is supplied at a sufficiently high concentration (100 times the minimal concentration of biotin required for growth). While this does not report directly on the substrate specificity of *Leuconostoc* biotin ligase, it is clear that the organism has a uniquely relaxed tolerance for growth on biotin analogs, and we hoped that its biotin ligase could potentially have interesting substrate specificity.

One monofunctional archaeal biotin ligase, lacking any N-terminal domain, has been previously cloned, purified, and biochemically investigated. This enzyme is from *Pyrococcus horikoshii*, an archaeon originally isolated from deep-sea vents that grows optimally at 98°C.[24]

Representing the lower eukaryotes, two yeast biotin ligase enzymes have been cloned, purified, and biochemically characterized. The baker's yeast *Saccharomyces cerevisiae* biotin ligase (yeast biotin ligase, yBL) gene was first isolated in 1995.[25] The protein does not have DNA binding capacity, though it does have a large N-terminal domain of ill-defined function. Deletion of this domain severely curtails the enzyme's catalytic activity, up to 3500-fold relative to full-

length enzyme.[26] Another monofunctional fungal biotin ligase gene was recently isolated from *Candida albicans*, a causative agent of human yeast infections.[27] *C. albicans* biotin ligase is closely related to the *S. cerevisiae* enzyme, having 39% sequence identity with its fungal relative.

In higher eukaryotes, biotin ligase is generally called holocarboxylase synthetase, in reference to its activity in biotinylating carboxylase enzymes. *Arabidopsis thaliana*, the flowering plant used as a model system for plant biologists, encodes two biotin ligase (in this case, referred to as holocarboxylase synthetase, or HCS) genes targeted to either the cytosol or to chloroplasts. Both have been cloned, purified, and characterized. This enzyme has no DNA binding domain and fails to regulate the biotin biosynthetic operon in *E. coli*. [28-32] Two isoforms of monofunctional HCS from the pea plant *Pisum sativum* have also been purified from the plant. [31, 33]

Several mammalian HCS enzymes from have been purified and subjected to preliminary characterization. The dimeric rat HCS was purified from rat liver. [34] The monomeric bovine HCS was similarly purified from bovine liver. [35] Interestingly, this enzyme can utilize CTP with an equivalent efficiency to the canonical biotin ligase nucleotide substrate, ATP.

Finally, the human HCS has been extensively characterized because mutations in the biotin-binding region of HCS decrease the affinity of the enzyme for biotin and can cause multiple carboxylase deficiency, a rare recessive disorder characterized by rashes, seizure, and death if untreated. [36] The enzyme consists of a conserved catalytic domain as well as a large N-terminal domain of ill-characterized function; expression of N-terminally truncated constructs

in *E. coli* revealed that the minimum functional enzyme consists of the last 348 amino acids of the protein and contains the catalytic domain, but that the N-terminal domain is required for full activity and substrate recognition.[37] HCS is expressed in human cells as two isoforms.[38] Both isoforms are monomeric and have no DNA binding capacity.

Many of these enzymes have been functionally and biochemically characterized, and a summary of their biochemical properties relevant to the current study is presented in Table 2-1.

Species	N-terminal domain	Bi-functional	Biotin K_m	ATP K_m	Activity toward <i>E. coli</i> BCCP	Endogenous BCCP K_m	Ordered substrate binding	Sensitivity to biotin analogs
<i>E. coli</i> [1, 2, 39]	Yes	Yes	490 nM	0.3 mM	Yes	4.39 μ M	Yes	Uses desthiobiotin (DTB) and ketone biotin as substrates
<i>B. subtilis</i> [13, 14]	Yes	Yes	NR	NR	Yes ¹	NR	NR	NR
<i>M. tuberculosis</i> [15]	No	No	420 nM	21 μ M	NR	5.2 μ M	NR	Does not use DTB as substrate
<i>A. aeolicus</i> [16, 17]	No	No	440 nM	15 μ M	NR	NR	No	NR
<i>L. dextranicum</i> [23]	NR	NR	NR	NR	NR	NR	NR	Growth not inhibited by DTB or cyclohexyl biotin analog ³
<i>B. stearrowthermophilus</i> [20]	NR	NR	75 nM	300 nM	NR	NR	NR	NR
<i>P. shermanii</i> [18]	NR	NR	2 μ M	38 μ M	NR	900 nM	NR	NR
<i>P. horikoshii</i> [40, 41]	No	No	NR; K_D is 200 nM	NR; K_D is 240 μ M	NR	NR	No	NR
<i>S. cerevisiae</i> [26]	Yes	No	67 nM	20 μ M	Yes ² ; K_m is 11 μ M	1 μ M	Yes	Not inhibited by biotin analogs ⁴
<i>C. albicans</i> [27]	NR	No	NR	NR	Yes ¹	NR	NR	NR
<i>A. thaliana</i> [29, 31, 32]	NR	No	130 nM	4.4 μ M	Yes ^{1,2}	30 μ M	Yes	Does not use DTB; not inhibited by biotin analogs ⁴
<i>P. sativum</i> [31, 33]	NR	No	28 nM	1 mM	Yes ²	NR	NR	Not inhibited by biotin analogs ⁴
Rat [34]	NR	No	50 nM	NR	NR	NR	NR	NR
Cow [35]	NR	No	13 nM	NR	NR	NR	NR	NR
Human [37, 38]	Yes	No	800 nM	47 μ M	Yes ²	21 μ M	Yes	NR

Table 2-1. Comparison of some biochemical properties of previously reported biotin ligase enzymes.

NR, not reported.

¹ *E. coli* BCCP utilization assayed by complementation of conditionally lethal heat-sensitive *E. coli* BirA strain.

² *E. coli* BCCP utilization assayed directly *in vitro*.

³ Cyclohexyl biotin analog is γ -(3,4-ureylenecyclohexyl)butyric acid.

⁴ Biotin analogs assayed for inhibition are biocytin, desthiobiotin, diaminobiotin and iminobiotin.

The biochemical properties of these enzymes are in some cases very similar and in other cases interestingly variable. For example, it is clear from both complementation studies and *in vitro* assays that these enzymes are similar in their general ability to modify biotinyl domains from other species, in particular the *E. coli* BCCP. These reactions appear to be kinetically efficient; in the case of the *S. cerevisiae* ligase, the apparent K_m for *E. coli* BCCP is only about ten times worse than its endogenous substrate. The K_m values of these enzymes for biotin vary over two orders of magnitude. Oligomerization states and the existence and function of N-terminal domains vary and have no obvious correlation to mono- or bifunctionality or kinetic parameters. As discussed in the text, examples of differential divalent cation (Zn^{2+} in place of the canonical Mg^{2+} for the *P. shermanii* enzyme) or nucleotide (CTP rather than the canonical ATP for the bovine liver enzyme) substrate specificities can be found in the literature. Most interestingly of all, while the *Leuconostoc* biotin ligase has not been previously isolated, the organism can utilize desthiobiotin and a cyclohexyl analog of biotin for growth, indicating that some organisms may have interesting flexibility in biotin metabolism. We conclude that this similarity in protein substrate specificity, but variability in structure as well as affinity for small molecule substrates, portends potential room to discover new small molecule reactivity while preserving specificity for protein and peptide substrates.

IV. Biotin ligase crystal structures

To date, biotin ligases from four species have been crystallized: *E. coli* BirA, *P. horikoshii* biotin ligase, *M. tuberculosis* biotin ligase, and *A. aeolicus* biotin ligase. The first structure determined was that of *E. coli* BirA, in its unliganded,[42] biotin bound,[42] and biotinyl-lysine bound[43]

forms. The structure of *E. coli* bound to biotinol-adenylate, an analog of the biotinyl-5'-AMP reaction intermediate, was subsequently determined.[44] The protein consists of the N-terminal DNA binding domain, a central catalytic domain housing the active site, and a C-terminal domain that has been proposed to be involved in protein substrate binding.[45] The biotin binding site features many hydrogen-bonding and hydrophobic interactions with biotin. The side chains of residues S89, T90, and Q112, as well as a backbone carbonyl group from R116, hydrogen-bond with the ureido ring of biotin.[42] The aliphatic sidechain of biotin is bound in a hydrophobic pocket.[42] In the unliganded and biotin-bound enzyme, a Gly-Arg-Gly-Arg-Arg-Gly loop is disordered; when the intermediate analog biotinol-adenylate is bound, mimicking the conditions under which BirA serves its DNA-binding function, the loop becomes ordered and is involved in high-affinity retention of the intermediate, forming a "cap" over the active site.[44] K183 is required for reaction of biotin and ATP to form the biotinyl-5'-AMP intermediate.[43]

The crystal structure of the *P. horikoshii* enzyme has been solved in its unliganded state, with biotin bound, and with various other ligands bound. While this monofunctional enzyme obviously lacks an N-terminal DNA binding domain, and furthermore dimerizes along a completely different interface than does BirA, the sequence and structure of the catalytic domain are highly conserved; in fact, the *P. horikoshii* enzyme has 31% sequence identity with the catalytic and C-terminal domains of *E. coli* BirA.[40] The *P. horikoshii* biotin ligase structure is superimposable on the catalytic and C-terminal domains of BirA with a r.m.s.d. value of 2.41 Å.[40] The catalytic lysine is conserved (K111 in *P. horikoshii* biotin ligase), as is the threonine

involved in interaction with the biotin ureido ring (T90 in BirA, T22 in *P. horikoshii* biotin ligase), as well as the glycine-rich loop that becomes ordered upon biotin binding.[40]

The crystal structure of the *A. aeolicus* biotin ligase enzyme has similarly been solved with biotin bound. Again, the disordered glycine-rich loop involved in intermediate binding is observed, as well as conservation of active-site residues including T14 (which aligns with T90 of BirA and interacts with the ureido ring of biotin).[17]

Because such high sequence and structural similarity is observed in the active sites and catalytic domains of these enzymes, we anticipate that any differences in their biochemical properties would be difficult to predict from structural analysis.

V. Study design

The varied biochemical properties of biotin ligases from across evolution made us optimistic that we might observe differences in their specificity for small-molecule substrates, and potentially discover new catalysts for ligation of our probes of interest to proteins. However, due to the similarity of reported biotin ligase structures, the basis of these differences remained mysterious, so we adopted a screening strategy wherein we would select a panel of distantly related biotin ligase enzymes in hopes of interrogating structural and functional diversity. We were also encouraged to note that we could potentially design a relatively high-throughput screening method using a single protein substrate, since biotin ligases from multiple species can recognize the *E. coli* BCCP. We next explain the probes, enzymes, and substrates that we chose for our screen, as well as the results of our screening strategy.

Results and discussion

Panel of biotin analogs

Our first prerogative was to choose a panel of probes that would extend the utility of site-specific labeling to new functional groups, but that would also be likely to be accommodated by the enzyme active site. This presumably requires significant structural similarity to biotin.

Members of the Ting lab prepared the panel of probes depicted in Figure 2-2 to meet these two design goals.

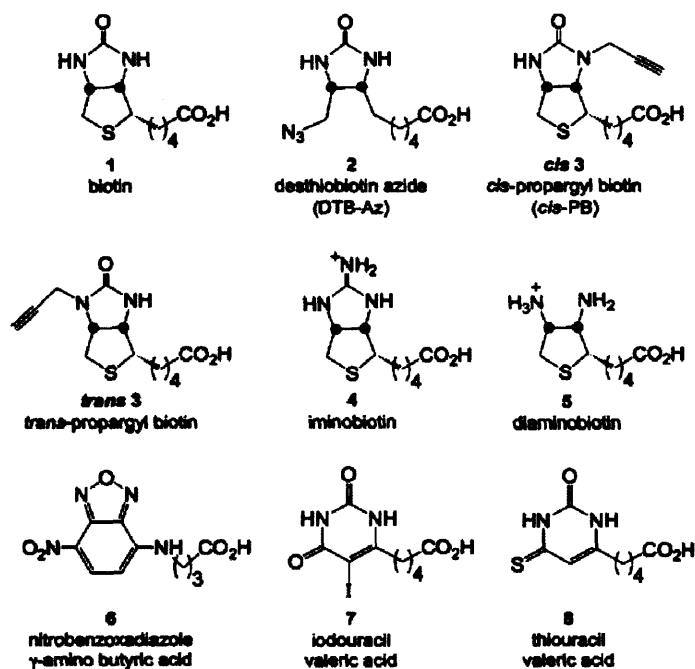


Figure 2-2. Structures of biotin analogs used in this study.

Compounds 2 and 3 were prepared by Dr. Irwin Chen, and compounds 7 and 8 were prepared by Dr. Yoon-Aa Choi.

Compound 1, included here for structural comparison, is *d*-biotin, which consists of a puckered bicyclic structure containing a ureido functionality atop a thiophene ring and an aliphatic valeric acid sidechain. Several classes of biotin analogs are represented among the remaining probes.

We first sought to incorporate bio-orthogonal functional group handles onto our probe, which would allow secondary derivatization with a wide variety of reporter structures. These functional handles are represented in compounds 2 and 3 (Figure 2-2), synthesized by Dr. Irwin Chen. Compound 2, desthiobiotin azide, contains an azide moiety in place of the thiophene ring of biotin. Azides are absent from cells and can react, as discussed in the introduction, with alkynes, phosphines, and cyclooctynes. Compound 3, *cis*- and *trans*-propargyl biotin, bears a propargyl group on the *cis*- and *trans*-ureido nitrogen atoms. Alkynes are also absent from cells and can be derivatized *via* the [3+2] azide-alkyne cycloaddition. The synthesis of these probes has been described and is summarized in Figure 2-3.[46]

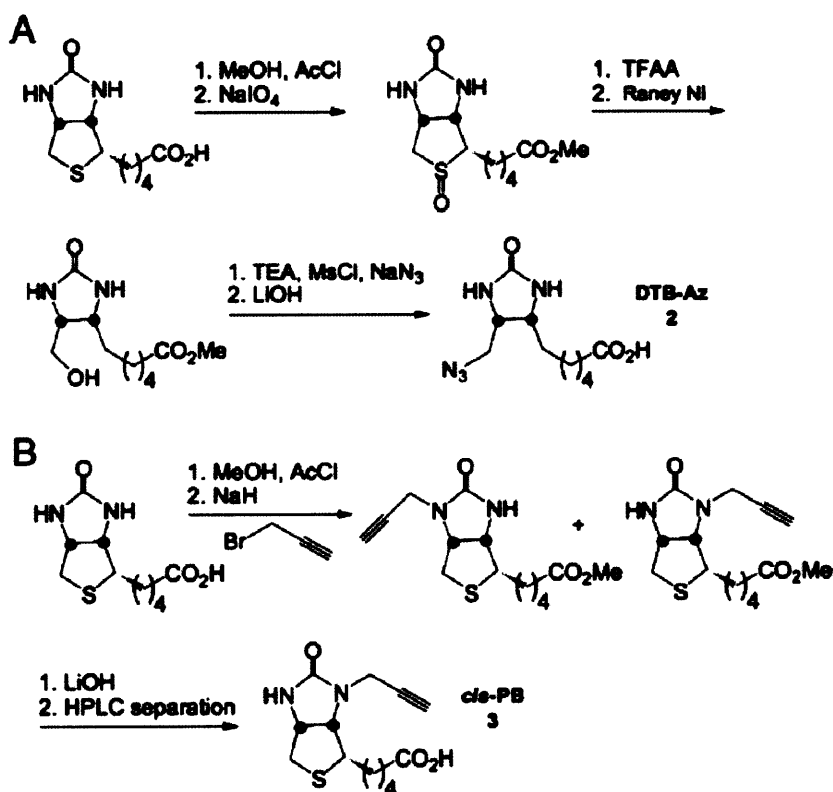


Figure 2-3. Synthetic routes to (A) desthiobiotin azide and (B) *cis*-propargyl biotin. MeOH, methanol; AcCl, acetyl chloride; Me, methyl; TFAA, trifluoroacetic anhydride; TEA, triethylamine; MsCl, mesyl chloride.

We also sought to directly incorporate probes with interesting functionality. The commercially available probes 4 and 5 (Figure 2-2), iminobiotin and diaminobiotin, display pH-dependent (strept)avidin binding properties and can be used for protein purification.[47] This pH-dependence is due to the presence of protonatable groups with acid dissociation constants in or near the physiological range. These molecules bind (strept)avidin in their deprotonated form, and upon titration to lower pH, become protonated and dissociate.[47] The pKa of the guanidino nitrogen of iminobiotin is 11.9, though it has been found to stoichiometrically bind avidin at pH 9.0.[47] The pKa values for the two amino nitrogens of diaminobiotin are 5.1 and 8.3, and are presumably so low because of the proximity of these groups to each other.[47]

Probe 6, nitrobenzoxadiazole (NBD) γ -amino butyric acid, contains the environmentally sensitive fluorophore NBD (Figure 2-2). Though NBD is planar while biotin is puckered, we hoped that the small bicyclic fluorophore could be accommodated by the enzyme active site.

Probes 7 and 8, iodouracil valeric acid and thiouracil valeric acid (Figure 2-2), synthesized by Dr. Yoon-Aa Choi, contain the photo-activatable cross-linkers iodouracil and thiouracil.[48] Such probes are useful for the study of protein-protein interactions; a complete discussion of this application can be found in Part III of Chapter 1.

For the purposes of intracellular labeling, all of these probes would have to be able to cross the cell membrane. We expect them all to have similar hydrophobicity to, if not greater hydrophobicity than, biotin, which enters cells by passive diffusion at concentrations greater than 2 μ M.[49] At lower concentrations, biotin is actively transported into cells by the sodium-dependent multivitamin transporter (SVMT); the SVMT has been shown to take up

desthiobiotin, so our biotin analogs stood a chance of being recognized as well.[50] Therefore we hypothesized that these probes could potentially be membrane permeable.

We note here that the utility of these probes inside cells would be limited by competition with endogenous free biotin, which has been demonstrated to be of approximately equivalent concentrations in bacterial and mammalian cells,[51] and in the range of 10-100 nM.[8, 52] While free biotin in cultured mammalian cells can be depleted with overnight biotin starvation,[53] it would be preferable not to subject cells to this non-physiological stress. Therefore we began our study with the knowledge that only kinetically efficient enzyme-probe pairs could become candidates for use in intracellular protein labeling.

Panel of enzymes

Our next task was to select a set of biotin ligase enzymes to probe for new reactivity toward our probes. We wished to select enzymes with maximal diversity in the hope of probing as much functional BirA sequence space as possible. We therefore selected the previously reported human, *Saccharomyces cerevisiae* (yeast), *Pyrococcus horikoshii*, and *Bacillus subtilis* biotin ligases, all of which are described in the introduction to this chapter, as well as five additional species (*Methanococcus jannaschii*, *Leuconostoc mesenteroides*, *Trypanosoma cruzi*, *Giardia lamblia*, and *Propionibacterium acnes*). We chose these species because (1) their genomes are sequenced, facilitating PCR cloning of their biotin ligase enzymes, (2) they are common model organisms or well-studied disease agents, and therefore genomic or cDNA was available, (3) at least one, *Leuconostoc mesenteroides*, had been previously been reported to have relaxed tolerance for growth on biotin analogs, and (4) they represent distantly related organisms, by

the broad measure of coming from different kingdoms of life. These species include archaea, bacteria, two protists (and infectious disease agents), a fungus, and a higher eukaryote.

We performed a ClustalW alignment[54] of all the biotin ligase enzymes utilized in this study. A portion of the alignment contained within the (though not comprising the entire) conserved central catalytic domain of each enzyme is presented in Figure 2-4. Several interesting conserved sequence elements are highlighted in the alignment. First, the disordered glycine-rich loop involved in binding the biotinyl-5'-AMP intermediate (discussed in the introduction to this chapter) is highlighted in red. A sequence involved in biotin binding, KWPND, is highlighted in blue. Finally, the catalytic lysine residue required for formation of the intermediate from biotin and ATP is highlighted in green. This alignment supports our previous assertion that the catalytic domains of these enzymes are highly conserved, and also gives us confidence that the previously unreported enzymes we chose to investigate were annotated correctly in their genome sequences.

```

P._horikoshii -----LEEGTVIVADKQTMGHGRLLNRRKWESPEG-GLWLSIVLSPKVP
B._subtilis -----NAPEGTLLVADKQTAGRGRMSRVWHSQEGNGVWMSLILRPDIP
M._jannaschii -----GKRNFIVLADKQNNGKRWGRVWYSDEG-GLYFSMVLDSKLY
E._coli -----ELKSGDACIAEYQQAGRRGRRGRKWFSPFGANLYLSMFWRLEQG
Leuconostoc -----EITKNTVIVSEMSAGIGRLGRKFFSPKNTGMYVTFALPLPVE
P._acnes -----GGTSFCLLVADHQEGRGRRFTRSWQDVPGTSLAISALVPNDRP
T._cruzi -----GAPFAVLAESQVRGRGTGERVWISPKGNMYFTLCIPEKSVT
Giardia -----NDHIVIFHTWNQTAGYGRTG-PWQSAAGNLAATWSFPVHVFS
Human EVTPPTMRLLDGLMFQTPQEMGLIVIAARQTEGKRGRGGNVWLSPVGCALSTLLISIPLRS
S._cerevisiae -----IPESTLLHVGTIQVSGRRGRGGNTWINPKGVCASTAVTTMPLQS
* * : . .

P._horikoshii QKD-----LPKIVFLGAVGVVETLKEFS-IDGRIKWPNDVLVNY-----
B._subtilis LQK-----TPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILING-----
M._jannaschii N-----PKVINLLVPICIIEVLLNYVDKELGLKFPNDIMVKVNDNY-----
E._coli PAA-----AIGLSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQD-----
Leuconostoc TIVN-----PGRLTTSTAVAVSKMVKEVFDIDIQFKWVNDLLYLN-----
P._acnes SQDWG----W--LSMVAGLAVAKVIEEVGGADRSRVTLKWPNDVLVDLDTDQ-----
T._cruzi PEIT-----PVLPLLTGLVCRAAIMSVIKGADVHVKWPNDIIYAG-----
Giardia PSS-----MYLRTSLALLRLFEYYGIASTIKWPNDIIVDG-----
Human QLGQR----IPFVQHLMSVAVVEAVRSIPEYQDINLRVKWPNDIIYSD-----
S._cerevisiae PVTNRNISVVFVQYLSMLAYCKAILSYAPGFSDIPVRIKWPNDLYALSPTYYKRKNLKLV
* * : **

P._horikoshii -----KKIAGVLVEGKGDK----IVLGIGLNVN-----N
B._subtilis -----KKTVGILTEMQAEEDRVRSVIIGIGINVNQQP----NDFPD
M._jannaschii -----KKLGGILTELTDDY----MIIGIGINVN-----NQIRN
E._coli -----RKLAGILVELTGKTGDAAQIVIGAGINMAMR-----RVEE
Leuconostoc -----KKVGGILTEAITDFESQQFSSLAVGIGMNLAT----PDGGF
P._acnes -----GGKVCGILSERVDGPAG-PHAVIGIGINVSMG----RD
T._cruzi -----KKIGGSLVESEGES----LIIGIGMNVELAPPV----TDSGR
Giardia -----EKCAGFLCEQLCPDR----VLIGVGINVAVAP-----
Human -----LMKIGGVLVNSTLMGET-FYILIGCGFNVTNSNP----TICIN
S._cerevisiae NTGFEHTKLPLGDIEPALYLKISGLLVNTHFINN-KYCLLLGCGINLTSDGPTTSLQTWID
* * * : .. *::

P._horikoshii KVPNGATSMKLELGSEVPLLSVFRSLITNLDRLYLNFLKNPMDILNLVRDNMILGVRVKI
B._subtilis ELKDIATSLSQAAGEKIDRAGVIQHILLCFEKRYRDYMTHGFTPIKLLWESYALGIGTNM
M._jannaschii EIREIAISLKEITGKELDKVEILSNFLKTFESYLEKLKNKEIDDYEILKKYKKYSITIGK
E._coli SVVNQGWITLQEAGINLDRNTLAAMLIRELRAAELFEQELAPYLSRWEKLDNFINRPV
Leuconostoc PDEISQKAGALTDEMAVSGNEVVGSLINHFFDMYQDYQDG---HYIPQYRKKVVGVGQSV
P._acnes ELPLPTATSLALCGLNHDKNELLASLLIHLDELLTVVFETGTVRDQYVARCDTIGTPVRL
T._cruzi ASTTVNEVAATLGQPKVTPAQLAELVWKHFFQMISDTALTRKILVTRFDAAMDKSLSLHR
Giardia PGTSKAPAFAKHLDPSTQTIDLVGYLTNELLAADIPDSMALELYSQACTTIGRQYQHAQ
Human DLITEYNKQHKAELKPLRADYLIARVVTVLEKLIKEFQDKGPNSVLPLYYRVWHSGQQV
S._cerevisiae ILNEERQQLHLDLLPAIKAEKLQALYMNNLEVILKQFINYGAEILPSYELWLHSNQIV
: :

```

Figure 2-4. ClustalW alignment of a portion of the catalytic domain of biotin ligase enzymes utilized in this study. Species (or genus) names are provided to the right of the sequence. Red and green text highlight conserved sequence elements important in catalysis; blue highlights a conserved sequence element required for biotin binding. Starred residues are identical in all sequences; two dots represent conserved residues, based on considerations of charge, polarity, and size; a single dot represents a weakly conserved residue.

We obtained expression plasmids for the human, yeast, and *P. horikoshii* enzymes. We cloned the *Giardia* biotin ligase from a cDNA library; the *M. jannaschii*, *Leuconostoc*, and *P. acnes* genes were cloned directly from cell lysates; and the *T. cruzi* and *B. subtilis* enzymes were cloned from purified genomic DNA. Each coding sequence was inserted into a bacterial expression plasmid with a C-terminal hexahistidine tag for affinity purification. (In a cautionary note for future scientists who wish to append affinity or epitope tags to uncharacterized proteins, our C-terminally tagged *Leuconostoc* construct was insoluble and produced inclusion bodies upon expression in *E. coli*; subsequent construction of an N-terminally hexahistidine tagged *Leuconostoc* enzyme afforded soluble, active protein upon purification.) We subsequently over-expressed each biotin ligase in *E. coli*, and purified them as described in the Methods section. Each enzyme was obtained in acceptable yield and purity (Figure 2-5), except for the human enzyme, which was significantly degraded. Nevertheless, we observed enzymatic activity in this sample (vide infra), so we made use of it for our study. We also note that the enzymes cover a satisfyingly broad range of sizes, indicating that they probably span a range of different domain structures and organizations, as desired.

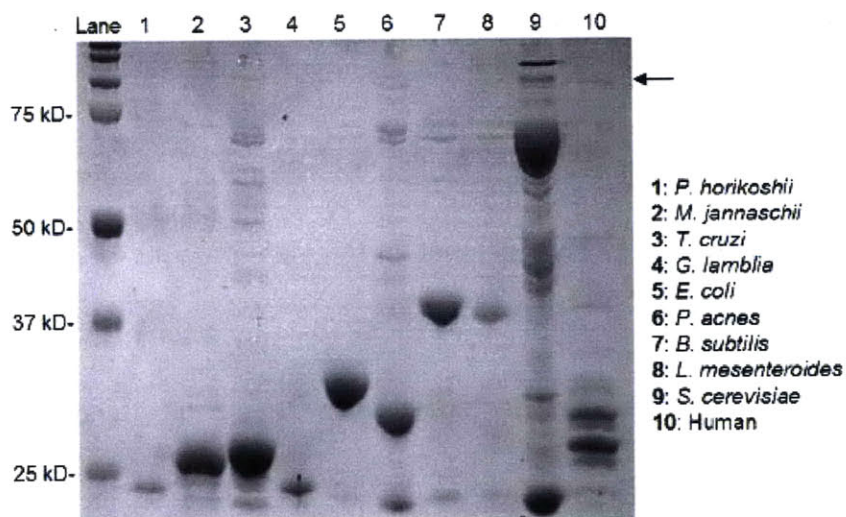


Figure 2-5. SDS-PAGE characterization of purified biotin ligases. In lanes 1-9, the major band represents the full-length ligase. In lane 10, most of the human enzyme is proteolyzed (the arrow indicates the full-length ligase), but biotinylation activity is still detected.

Substrate selection and assay design

We next had to identify suitable protein or peptide substrates for screening our enzymes. We preferred to identify a single substrate that could be recognized by all of our enzymes, in order to simplify our screening and analysis. We used an HPLC-based assay to test several of our biotin ligase enzymes for biotinylation activity toward the engineered BirA peptide substrate, the acceptor peptide (AP), and confirmed and extended what we and others had previously reported: the AP is not recognized by biotin ligase enzymes from other species, including human, yeast, and *Spodoptera frugiperda*. [55-58] We additionally observe that the *B. subtilis*, *T. cruzi*, and *P. horikoshii* enzymes do not modify the AP peptide (Figure 2-6). We therefore investigated BCCP protein substrates; it was more likely that we could identify a BCCP that could be utilized by all of our enzymes, because, as discussed above, sequence conservation

among biotinylated proteins is extremely strong at the site of biotin attachment, and biotin ligases generally recognize and biotinylate BCCPs from other species.

We first investigated the *P. shermanii* BCCP,[18] but found that it degraded after biotinylation; we then tested BCCP-87, a biotinylation-competent fragment of *E. coli* BCCP,[59] but found that it co-purified with full-length endogenous *E. coli* BCCP (data not shown). We then moved to p67, a 67-amino acid fragment of the C-terminal domain of human pyruvate carboxylase which is competent for biotinylation on a single lysine residue, K61.[60] We obtained an expression plasmid for this construct (a kind gift from the laboratory of Roy Gravel), which we found to be isolable in excellent yield and purity from *E. coli*. This protein and its modification products can be easily analyzed by HPLC and native gel-shift assays.

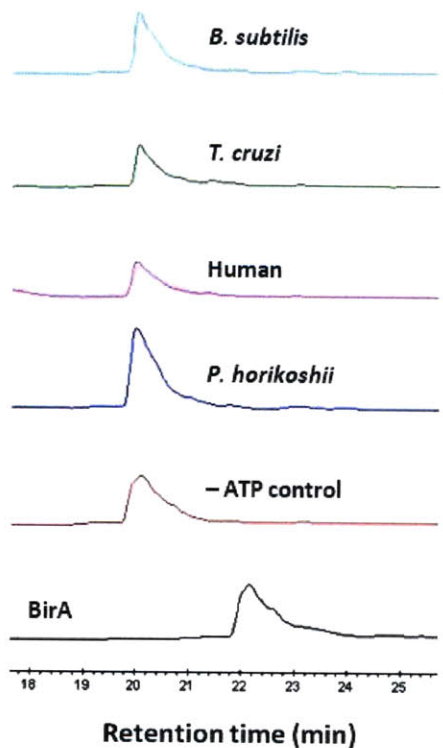


Figure 2-6. The AP peptide is not recognized by biotin ligases from other species. HPLC traces for analysis of AP reaction products with various biotin ligase enzymes in the presence of biotin are shown. Absorbance in arbitrary units is plotted on the y-axis. Each trace is labeled according to the biotin ligase enzyme used in that reaction. A negative control reaction using BirA with ATP omitted is used as a genuine standard for unmodified AP peptide. Unmodified AP peptide elutes from the column at approximately 20 minutes, while biotinylated AP peptide elutes at approximately 22 minutes. Product is formed only in the presence of BirA, and no detectable product is formed for the other enzymes.

Our assays are based on the change in charge that occurs in substrate peptides and proteins upon modification with biotin or biotin analogs. As discussed in Part II of Chapter 1, biotin ligase covalently attaches biotin to a specific lysine residue of its substrates, producing a neutral amide product. We find this change in charge to be sufficient to cause a large retention time shift in both AP peptide and p67 protein when analyzed by reversed-phase HPLC. (Though positively charged molecules like iminobiotin and diamminobiotin would not produce an overall change in charge upon ligation, they could ostensibly cause changes in polarity of the product that would be sufficient for HPLC-based resolution.) Similarly, a native gel-shift assay has been previously reported for analysis of BCCP-87.[61] Migration of BCCP-87 is governed in a native gel both by its charge and its conformation; biotinylated protein migrates more quickly due to the elimination of a positive charge after reaction, and also possibly due to conformational change. We confirmed that the same gel-shift is observable for p67 (vide infra). We would subsequently utilize both assays for analysis of p67 reaction products.

Having our biotin ligase enzymes, a suitable substrate, and working assays in hand, we needed to confirm that all of our enzymes could biotinylate our p67 substrate. We therefore tested each of our biotin ligase enzymes for biotinylation activity toward p67 using the HPLC assay, as described in the Methods section (Figure 2-7).

First, we observed that the *Giardia* enzyme did not display activity toward p67; we also tested it with our other substrates, *P.shermanii* BCCP and BCCP-87, and found that it was inactive toward all three (data not shown). We therefore eliminated the *Giardia* enzyme from our subsequent analysis. We note that the only putative annotated biotinylation substrate in the

Giardia genome, a putative transcarboxylase, diverges significantly in sequence from the canonical biotinylation sequence,[62] which could easily explain why it does not recognize a canonical substrate; it would be interesting in the future to determine if the *Giardia* enzyme is capable of biotinylating this substrate, and to try to elucidate the reason for this rare sequence divergence.

Second, we observed that all of our other biotin ligases were active toward p67, but to varying degrees. Under identical reaction conditions, the *E. coli*, *B. subtilis*, and yeast biotin ligases were highly active toward p67, with their reactions going to completion within minutes. The *P. acnes*, *Leuconostoc*, and *P. horikoshii* enzymes displayed intermediate activity, with their biotinylation reactions going to completion on a time scale of hours. The *M. jannaschii*, *T. cruzi*, and human ligases displayed low, but detectable, activity. Rough estimates of the initial rates of these reactions span a 700-fold range of velocities. The reasons for these differences could be many; the reactions were conducted under conditions optimized for BirA activity, but buffers were not individually optimized for each enzyme; reactions were conducted at 30°C, which is certainly sub-optimal for the thermophilic enzymes; and each enzyme probably varies in its affinity for p67. Regardless, we were encouraged by this result and concluded that we observed sufficient activity of each enzyme toward p67 to screen our panel of probes under sufficiently forcing reaction conditions.

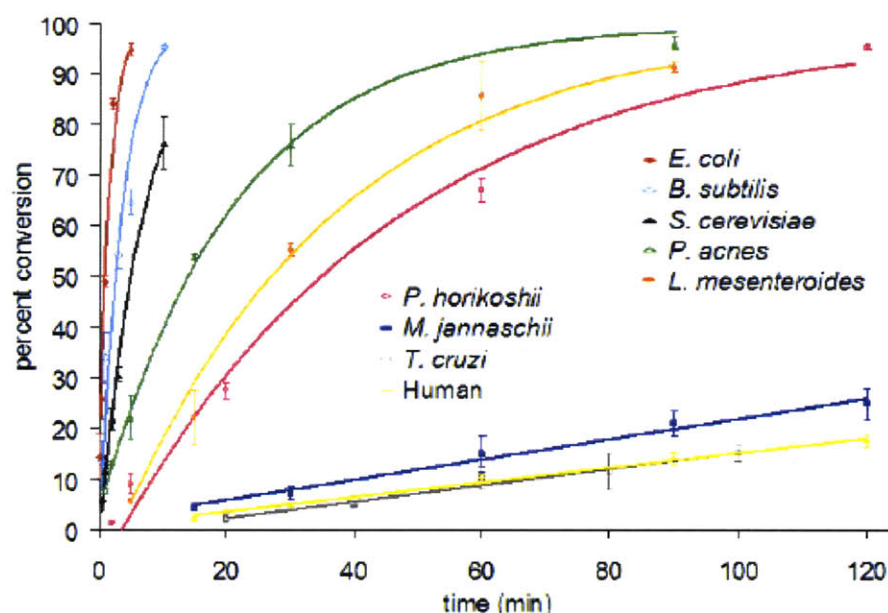


Figure 2-7. Biotinylation time courses of all enzymes under identical conditions. Reaction conditions: 1 μ M ligase, 100 μ M p67 acceptor protein, 1 mM biotin, 5 mM ATP, 5 mM $Mg(OAc)_2$, and 50 mM bicine pH 8.3 at 35 °C. Percent conversion to biotinylated p67 was measured by HPLC. Each point represents the average of three measurements. Error bars, 1 standard deviation.

Probe screening by HPLC; product identification and assessment of ligation specificity by gel-shift assay and LC-MS

Using p67 as a substrate, we assayed our eight active biotin ligase enzymes against each probe in our panel using either our HPLC assay or the gel-shift assay (Table 2-2). Notably, none of our enzymes incorporated iminobiotin, diaminobiotin, NBD γ -amino butyric acid, iodouracil valeric acid or thiouracil valeric acid. Presumably the shape or charge state of these probes precluded their association with the enzymes. However, we did observe product formation (which we simply define as any product formation above our detection limit, which for the HPLC assay is approximately 2% completion, or 2 μ M) in the reactions of *P. horikoshii* biotin ligase (PhBL) with desthiobiotin azide (DTB-Az) and propargyl biotin (PB) (a mixture of *cis* and *trans* isomers), as

well as yeast biotin ligase (yBL) with PB. (We were disappointed to find that the *Leuconostoc* enzyme did not display any new activity, despite the previously reported propensity for this organism to grow in the presence of biotin analogs; we conclude that this observation must be the result of other metabolic features of the organism than direct biotin analog ligation to carboxylase proteins.)

	Human	<i>Bacillus subtilis</i>	<i>Propionibacterium acnes</i>	<i>Leuconostoc mesenteroides</i>	<i>Trypanosoma cruzi</i>	<i>Methanococcus jannaschii</i>	<i>Pyrococcus horikoshii</i>	<i>Saccharomyces cerevisiae</i> (yeast)
Biotin	+	+	+	+	+	+	+	+
DTB-Az	-	-	-	-	-	-	+	-
<i>cis</i> -PB	-	-	-	-	-	-	+	+
<i>trans</i> -PB	-	-	-	-	-	-	-	-
Imino-biotin	-	-	-	-	-	-	-	-
Diamino-biotin	-	-	-	-	-	-	-	-
NBD γ -amino butyric acid	-	-	-	-	-	-	-	-
IU valeric acid	-	-	-	-	-	-	-	-
TU valeric acid	-	-	-	-	-	-	-	-

Table 2-2. Results of screening biotin ligase enzymes against probes. 1 μ M of each enzyme was incubated with 1 mM probe and 100 μ M p67 acceptor protein for 14 h at 30 °C in the presence of ATP. Formation of product (indicated by “+”) was detected by HPLC or native gel-shift assay. For screening, probe 3 (PB) was provided as a mixture of *cis* and *trans* isomers. Abbreviations: DTB-Az, desthiobiotin azide; PB, propargyl biotin; NBD, nitrobenzoxadiazole; IU, iodouracil; TU, thiouracil.

Significantly more product was formed under identical conditions in the PB reaction with yBL than PhBL, so we proceeded to further characterize the yBL reaction with PB. The synthetic route to PB affords both the *cis* isomer, with the *N*-propargyl group on the same side of the ring system as the alkyl side chain, and the *trans* isomer. Initial reactions were performed with this regioisomeric mixture, but we wished to determine which isomer yBL prefers in order to gain more insight into the structure of the active site. We purified the regioisomers by HPLC, then tested their yBL-mediated incorporation onto p67 separately using the native gel-shift assay (Figure 2-8). We utilize biotin as a positive control for product formation, and observe that 100% of the reaction product is a higher-mobility species on the gel (lanes 1 and 6). In negative control reactions where probe, ATP, and yBL enzyme were individually omitted, no product is formed, and all of the p67 remains in its lower-mobility apo form (lanes 3-5 and 8-10). In the presence of *trans*-PB, no product formation is formed within the detection limit of this assay (lane 7); in the presence of *cis*-PB, after a 14-hour reaction, we observe 100% conversion to product (lane 2). The enzyme clearly exclusively utilizes *cis*-PB, so we can conclude that, despite the lack of structural information about this enzyme, *trans* substituents on the biotin ring system are probably sterically excluded from binding, while the *cis* face of the biotin ring system has fewer steric clashes with the binding pocket.

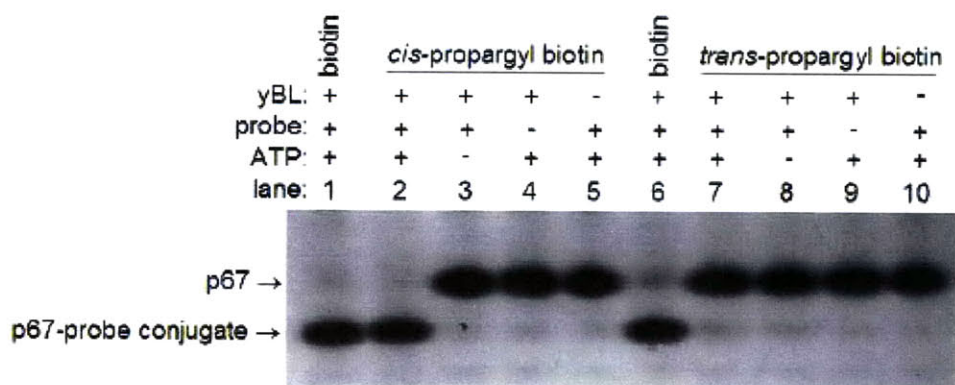


Figure 2-8. Native gel shift assay showing preferential ligation of *cis*-PB over *trans*-PB by yeast biotin ligase. Ligation products have a greater mobility in the gel than unmodified p67. Reaction conditions: 1 μ M yBL, 100 μ M p67, 1 mM *cis*- or *trans*-PB, 5 mM ATP, 5 mM Mg(OAc)₂, 50 mM bicine pH 8.3, 30 °C, 14 hours.

We then confirmed that enzymatic activity is required for our product formation in both reactions of interest, using our HPLC assay (Figure 2-9). Omission of either ATP or PhBL prevents product formation with DTB-Az; omission of ATP or yBL prevents product formation with *cis*-PB. For comparison, reactions were performed with biotin, and by inspecting the HPLC traces we observe that the p67-biotin conjugate has a unique, shorter retention time as compared to the p67-DTB-Az and p67-*cis*-PB conjugates.

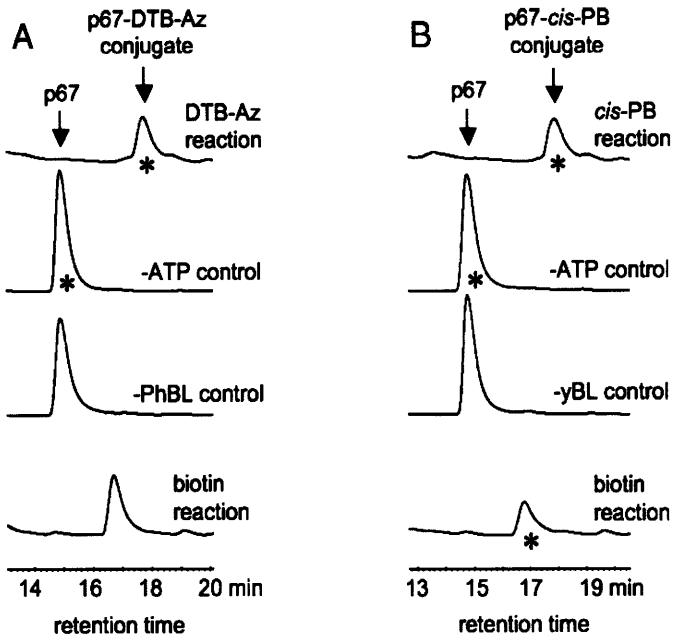


Figure 2-9. HPLC detection of (A) DTB-Az and (B) *cis*-PB ligation to p67 acceptor protein by PhBL and yBL, respectively. Negative controls are shown with ATP or enzyme omitted. The biotin-p67 conjugate has a shorter retention time than the DTB-Az or *cis*-PB conjugates. The starred peaks were collected and analyzed by mass spectrometry in a separate experiment (Figure 7). Absorbance in arbitrary units is plotted on the y-axis.

We further wished to structurally characterize our reaction products, in particular to confirm our expectation that exactly one molecule of DTB-Az or *cis*-PB had been ligated to p67, and that the reaction occurred site-specifically on the single reactive lysine residue of p67. We therefore collected the starred peaks shown in Figure 6 from our HPLC analyses and analyzed them by mass spectrometry. A comparison of the observed mass of the *P. horikoshii* ligation product of DTB-Az to p67 to apo-p67 (purified from our control reaction lacking ATP) revealed a difference consistent with exactly one DTB-Az molecule (Figure 2-10A, left). Similarly, comparison of our yBL-catalyzed *cis*-PB conjugate to apo-p67 was consistent with addition of exactly one *cis*-PB molecule to p67 by the enzyme (Figure 2-9B, left). We additionally compared the mass of the yBL-catalyzed p67 biotinylation product to our *cis*-PB conjugate, and gratifyingly found that they were distinguishable in our mass spectrometer, lending confidence to our structural assignment (Figure 2-10B, left).

In order to assay site specificity of the ligation, we generated a point mutant of p67, K61A (which we call p67(Ala)), in which the reactive lysine residue is mutated to alanine, preventing amide bond formation and eliminating site-specific ligation. We could therefore use p67(Ala) to assay for off-target modification events. Such non-specific reactivity has previously been reported for mutants of BirA at high concentrations and over long reaction times, probably due to release of the activated biotin-5'-adenylate (bio-5'-AMP) intermediate;[63] in fact, free bio-5'-AMP has been demonstrated to slowly react with the physiologically modified lysine of BCCP-87 and also the N-termini of proteins in solution due to the relatively low pKa of these primary amines.[64] Though we had already confirmed that only one probe molecule was being incorporated onto p67 in each reaction, since we were using long reaction times and high concentrations of probes, we had to further demonstrate that non-site-specific modification events were not occurring.

We therefore performed parallel reactions in which PhBL reacted with DTB-Az and p67, or in which p67 was replaced with p67(Ala); identical reactions were performed with yBL and *cis*-PB. Each reaction product was purified by HPLC (data not shown), and then analyzed by mass spectrometry. Again PhBL produces the expected 1:1 adduct of DTB-Az to wild-type p67; however, the single product of the p67(ala) reaction was identical to unmodified p67(Ala) (Figure 2-10A, right). Similarly, under conditions in which yBL catalyzes formation of the *cis*-PB-p67 adduct, p67(Ala) remains unmodified (Figure 2-10B, right). No additional peaks are visible in either p67(Ala) HPLC trace or mass spectrum.

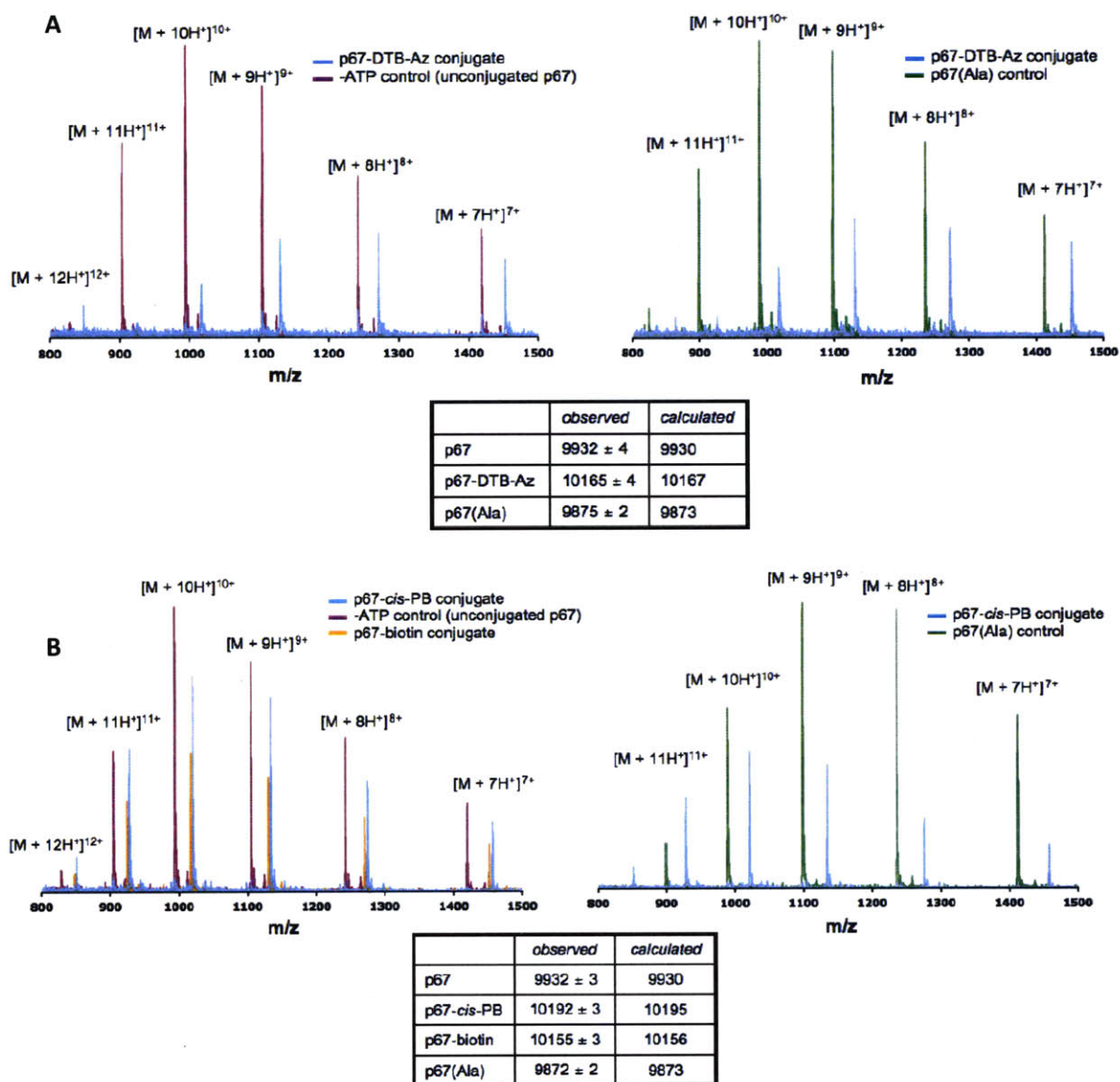


Figure 2-10. ESI-MS analysis of p67 conjugates to (A) DTB-Az and (B) *cis*-propargyl biotin. Products of negative control reactions with ATP omitted or p67(Ala) mutant in place of p67 are also shown. Charge states in each spectrum are labeled.

We therefore conclude that our ligation reactions are site-specific and dependent on enzyme activity, and therefore likely proceed by the same reaction mechanism as biotinylation.

Secondary derivatization of DTB-Az by the Staudinger ligation

Having identified and validated these new ligation reactions, we needed to demonstrate their utility for protein labeling. While alkynes can be derivatized using azide-functionalized probes *via* Click chemistry, a copper-mediated [3+2] cycloaddition reaction, as discussed in Chapter 1, this reaction was too toxic to be compatible with live-cell labeling at the time of this project; recently reported improvements have reduced toxicity from reactive oxygen species generation.[65] We therefore focused on the DTB-Az probe, as azides can be derivatized either by Click chemistry with alkynes, copper-free cycloaddition with strained cyclooctynes,[66] or Staudinger ligation with phosphines.[67]

We chose the Staudinger ligation for our *in vitro* azide derivatization. DTB-Az-p67 conjugate was enzymatically prepared, then reacted with a peptide comprising the FLAG epitope (DYKDDDDK) conjugated to a phosphine (FLAG-phosphine), which has been previously reported.[68] Subsequent immunoblotting with anti-FLAG antibody reveals the presence of the labeled protein (Figure 2-11, left). Parallel single-reagent-omission reactions were run as negative controls; treatment of these mixtures with FLAG-phosphine reveals no product formation, again confirming the specificity of our enzymatic ligation as well as the secondary derivatization step. It is important to note that the immunoreactive product runs as a slightly larger protein than the starting material visualized by Coomassie staining of the reaction mixture (Figure 2-11, right), as expected due to the addition of the octapeptide FLAG epitope. We also note that the reaction product is present at such small quantities as to be undetectable by Coomassie staining, due to the inefficiency both of the first enzymatic step (*vide infra*) and of the secondary Staudinger ligation.[66] Nevertheless, labeled product is produced and can be detected, demonstrating that we can use this chemistry for protein labeling *in vitro*.

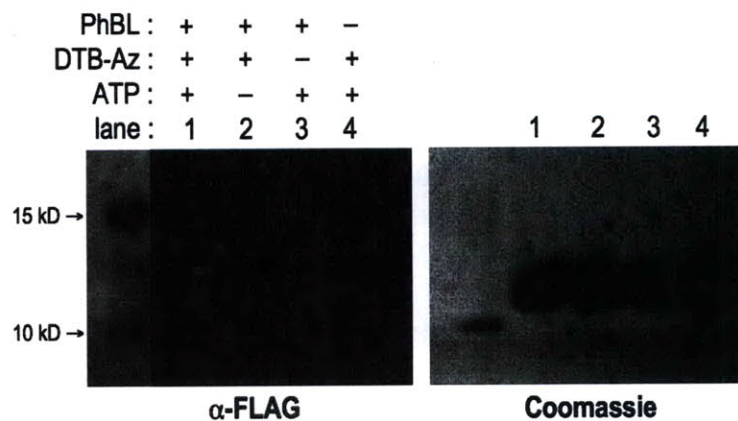


Figure 2-11. Functionalization of ligated DTB-Az by Staudinger ligation. PhBL was used to ligate DTB-Az to p67 protein. The azide was then functionalized with FLAG-phosphine. The FLAG epitope was detected with anti-FLAG immunostaining. Negative controls show omission of ligase, DTB-Az, or ATP from the reactions (lanes 2-4). Coomassie staining (right) demonstrates equal protein loading.

Kinetic analysis

We observed low yields of product formation in our initial screens, and we needed to determine the kinetics of these ligation reactions in order to determine if they would be useful for cellular labeling applications. We attempted to investigate the concentration dependence of the ratio of DTB-Az utilization by PhBL, but we were unable to saturate the initial rate of the reaction even with the maximal DTB-Az concentration that we could experimentally provide (data not shown). Therefore rather than determining k_{cat} for this reaction, we compared the rate of DTB-Az ligation to the rate of biotin ligation by PhBL under identical reaction conditions. Under these conditions, we obtained a DTB-Az ligation rate of $(1.34 \pm 0.11) \times 10^{-4} \mu\text{M s}^{-1}$ (Figure 2-12A), three orders of magnitude slower than the biotinylation rate of $0.20 \pm 0.02 \mu\text{M s}^{-1}$ (Figure 2-12B).

We obtained saturation of the initial rate of yBL-catalyzed *cis*-PB ligation to p67 at *cis*-PB concentrations above 1 mM (data not shown). Under these saturating conditions, we obtained a k_{cat} for *cis*-PB ligation to p67 of $(2.07 \pm 0.10) \times 10^{-2} \text{ s}^{-2}$ (Figure 2-12C). We also measured the k_{cat} for yBL-catalyzed ligation of biotin to p67 (Figure 2-12D); this rate is $0.28 \pm 0.4 \text{ s}^{-1}$, 14-fold faster than the rate of ligation for the unnatural probe.

Of course, these kinetic parameters are not suitable for cellular labeling. First, such high (millimolar) concentrations of probe would almost certainly be toxic to cells. Since saturating concentrations of probe would have to be delivered in order to out-compete endogenous biotin, as previously discussed, toxicity would probably be severe. Second, even if sufficient concentrations of probe could be delivered, the k_{cat} is probably still too slow to afford significant labeling except after very long incubation times. We therefore examined whether use of a different protein substrate could enhance kinetics of DTB-Az ligation by PhBL.

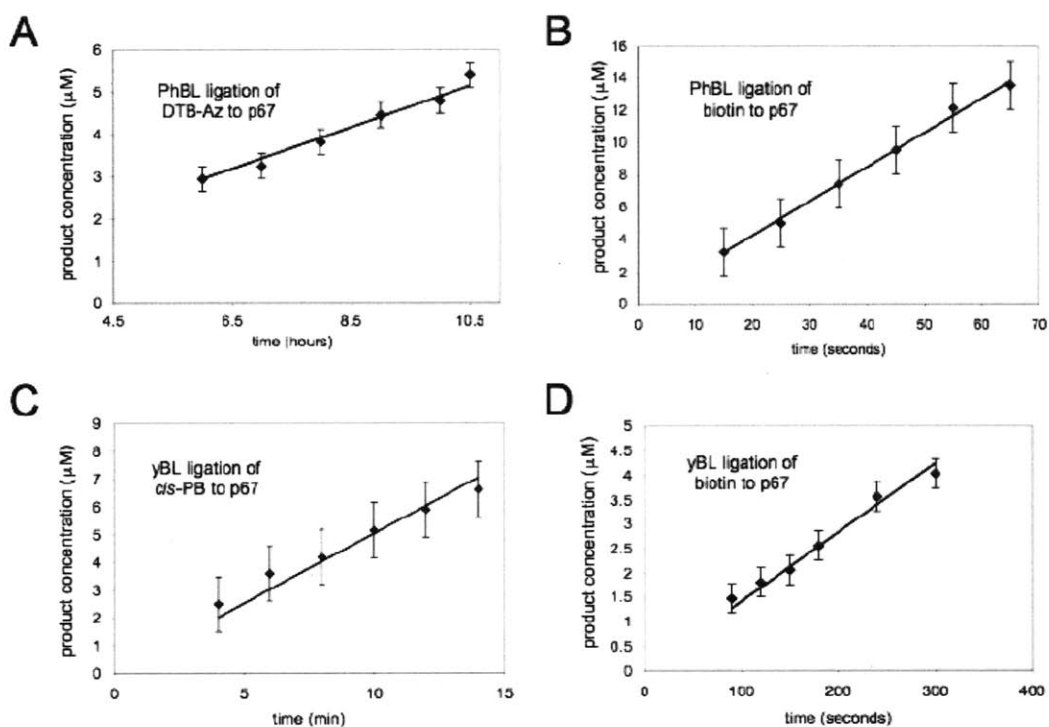


Figure 2-12. Kinetic analysis of DTB-Az and *cis*-PB ligation reactions. 1 μM PhBL, 100 μM p67, 5 mM ATP, 5 mM Mg(OAc)₂, and 50 mM bicine pH 8.3 were incubated at 35 °C with 1 mM DTB-Az (A) or 1 mM biotin (B). Samples at various timepoints were analyzed by HPLC, to determine the reaction velocities. (C) Measurement of k_{cat} for γBL-catalyzed ligation of *cis*-PB to p67. 400 nM γBL was incubated with 100 μM p67, 5 mM *cis*-PB, 5 mM ATP, 5 mM Mg(OAc)₂, and 50 mM bicine pH 8.3 at 35 °C. (D) Measurement of k_{cat} for γBL-catalyzed biotinylation of p67. 45 nM γBL was incubated with 100 μM p67, 1 mM biotin, 5 mM ATP, 5 mM Mg(OAc)₂, and 50 mM bicine pH 8.3 at 35 °C.

Attempt to improve ligation kinetics: assaying the P. horikoshii biotin carboxyl carrier protein

We first hypothesized that PhBL may exhibit improved reactivity toward its endogenous

substrate, the *P. horikoshii* BCCP (PhBCCP), derived from *P. horikoshii* acetyl CoA carboxylase.

We cloned, expressed, and purified this protein, and found that its modification products could

be analyzed by gel shift assay in much the same manner as p67. Much to our surprise, when

assayed with PhBL for DTB-Az turnover, under conditions providing 100% conversion of biotin

to product, only a small fraction of p67 became modified in the DTB-Az reaction (Figure 2-13).

While we could detect this product in our gel-shift assay (though its intensity above background was barely measurable), and later confirmed presence of the azide in the product *via* Staudinger ligation and immunoblotting (data not shown), it was clear that this reaction is not more efficient than DTB-Az ligation to p67 by PhBL. The reason for this low activity toward PhBCCP is unclear, though it is possible that the reaction may be accelerated at the higher temperatures at which *P. horikoshii* grows, a hypothesis that we left untested because we are interested only in physiologically relevant conditions for mammalian cell labeling.

PhBL:	+	+	+	+	-
Biotin:	-	+	+	-	+
DTB-Az:	+	-	-	-	-
ATP:	+	+	-	+	+
Lane:	1	2	3	4	5


Product → 

Figure 2-13. Gel-shift assay for *P. horikoshii* biotin ligase-catalyzed DTB-Az ligation to *P. horikoshii* BCCP (PhBCCP). Products have a greater mobility in the gel than apo-PhBCCP. Negative controls are shown with ATP or enzyme omitted.

Conclusion

In conclusion, we have discovered new reactivity of PhBL and yBL toward DTB-Az and *cis*-PB, respectively. These biotin analogs are suitably functionalized for secondary derivatization with bio-orthogonal probes, which we demonstrated *via* Staudinger ligation of DTB-Az with a phosphine-functionalized probe. Unfortunately, the kinetic parameters of these reactions are unsuitable for cellular labeling applications. No easy solution to this problem presents itself, aside from extensive *in vitro* evolution.

The similarity of the previously reported crystal structures of BirA, PhBL, and *A. aeolicus* biotin ligase renders the differences in small-molecule reactivity that we observe among our panel of biotin ligases toward biotin analogs somewhat mysterious, at least from a structural point of view. As more biotin ligase enzymes are crystallized in the future, it will be interesting to compare their binding sites to see if structural conservation is a general rule, and to determine any more obvious differences in active site organization.

Perhaps the most interesting differences among these enzymes come at the level of their protein and peptide substrate specificity. We have confirmed that the AP is recognized uniquely by BirA, and not by biotin ligases from several other species. We speculate that the AP may exhibit a different mode of binding to BirA than does BCCP, and that a similar binding interface is not available on other biotin ligase enzymes. This makes every enzyme in our panel a candidate for development of a peptide substrate orthogonal to the BirA-AP pair for multicolor labeling, as was reported for the yeast enzyme.[56] Perhaps the best candidate would be the *Giardia* enzyme, pending confirmation of its enzymatic activity, as it does not even recognize canonical BCCP sequences. It also confirms that the BirA and AP pair can be used for site-specific protein labeling in the context of cells from many different species.

Subsequent to this work, we serendipitously discovered that, while attempting to use *E. coli* BirA as a negative control for DTB-Az and *cis*-PB ligation, that BirA actually does ligate these probes to the p67 protein. In addition, BirA was found to ligate iminobiotin to p67. These reaction products were confirmed by HPLC and ESI-MS (data not shown). We never previously observed this reactivity because all previous BirA assays with biotin analogs were performed

with the AP peptide. We confirmed that, according to our initial observations, BirA does not detectably ligate these probes to the AP peptide. We speculate that the differential reactivity of BirA toward the AP peptide and p67 protein with respect to biotin analogs may be due to our proposed different binding modes of peptide and protein. More specifically, p67 probably binds to the enzyme in a similar manner to BCCP due to high sequence similarity, while the AP probably binds differently. If our biotin analogs are also oriented slightly differently in the active site than biotin, they may be incompetent for reaction with the AP. Unfortunately these BirA ligation reactions are still too kinetically inefficient to be utilized for cellular labeling. While this result was surprising to us, it does not invalidate the major conclusions of our study, which is that the enzyme homologs from different species can have differential substrate permissivity.

In parallel with this work, the Ting research group developed the enzyme *E. coli* lipoic acid ligase, or LplA, for site-specific protein labeling applications, as described in Chapter 1. Since LplA has proved to be an efficient and highly specific catalyst for labeling peptide-tagged proteins with azides as well as photo-affinity probes and coumarins (see Chapters 1 and 4 for a complete discussion), and our biotin ligase enzymes were found to be inefficient catalysts, we chose not to continue to pursue biotin ligases as an approach to developing new small-molecule probes for cellular labeling.

Methods

Cloning of biotin ligases

The expression plasmid for human biotin ligase (also called holocarboxylase synthetase), fused to glutathione S transferase, in the pDEST15 vector, was a gift from Roy Gravel. The *Pyrococcus*

horikoshii biotin ligase (PhBL) gene in pET11a was a gift from Mitsuaki Sugahara and Naoki Kunishima. The *Saccharomyces cerevisiae* (yeast) biotin ligase (yBL) gene in pET16b was a gift from Stephen Polyak and John Wallace. The expression plasmid for *E. coli* biotin ligase (BirA) has been described.[69]

The bacterial expression plasmids for biotin ligases from *Trypanosoma cruzi*, *Bacillus subtilis*, *Giardia lamblia*, and *Propionibacterium acnes* were prepared by PCR-amplifying the genes and cloning into the NheI and XhoI sites of the pET21a vector. This cloning introduces a C-terminal hexahistidine tag (from the pET vector). *Methanococcus jannaschii* biotin ligase was cloned in a similar fashion, between NheI and XhoI of pET21a, but a stop codon was introduced before the hexahistidine tag. *Leuconostoc mesenteroides* biotin ligase was cloned into the NdeI and BamHI sites of pET15b. This introduced an N-terminal hexahistidine tag from the vector. Primers used for all the PCR reactions are given in the table below. The following PCR program was used for amplification of all genes: 95 °C for 5 minutes, followed by 30 rounds of [95 °C for 30 seconds, 55 °C for 60 seconds, and 72 °C for 90 seconds]. PCR reactions contained 2 μM of each primer, 500 μM dNTPs, 25 ng template DNA or a minimal inoculation of cells, 1X Thermo pol buffer (New England Biolabs), and 5 units of Taq polymerase (New England Biolabs).

Species	Forward Primer	Reverse Primer	Template Source
<i>Trypanosoma cruzi</i> Y strain	5'cctagctagcatgccaatggatgtgcccca	5'aaaaagcttaaaaaggtattccgcatgag	genomic DNA, a gift from Barbara Burleigh
<i>Bacillus subtilis</i> PY79	5'cctagctagcatgcggtcaacattaagaaaa	5'aaaactcgaggccaattcgatatcggcaga	genomic DNA, a gift from Alan Grossman
<i>Giardia lamblia</i> WB/1267 (ATCC 50582)	5'cctagctagcatggaaggtaattagcgca	5'aaaactcgagacagactggcgaaagatcctg	cDNA library, a gift from Heidi Elmendorf
<i>Propionibacterium acnes</i> (ATCC 11827)	5'cctagctagcgtgccgtccacacctcgct	5'aaaactcgagcctggttcgcaatgatgaac	heat-generated cell lysate
<i>Methanococcus jannaschii</i> (DSMZ 2661)	5'cctagctagcatggaattatacatttaagt	5'aaaactcgagtattcttacatggatgcaaat	cell lysate generated by autolysis after exposure to air
<i>Leuconostoc mesenteroides</i> (DSMZ 20241)	5'aaaacatatgagtacagctgataaactgttag	5'aaaaggatccctaaccctgtattcattatttg	heat-generated cell lysate

Table 2-3. Cloning primers and template sources for biotin ligase enzymes.

Expression and purification of biotin ligases

In general, expression plasmids were transformed into *E. coli* BL21 DE3 cells, and individual colonies were amplified in 500 mL Luria Broth (LB) supplemented with 10 µg/mL ampicillin until OD(optical density) 0.6. Protein expression was induced with 420 µM isopropylthiogalactoside (IPTG) for 3 hours at 30 °C. Cells were harvested and stored overnight at -80 °C. For yBL, expression was induced with 42 µM instead of 420 µM IPTG, for 2 hours at 30 °C. For human

holocarboxylase synthetase and PhBL, expression was induced with 420 μ M IPTG for 12 hours at 25 °C. For hexahistidine-tagged proteins, cells were lysed by sonication in binding buffer (50 mM Tris-HCl pH 7.8, 300 mM NaCl) in the presence of 1 mM phenyl-methyl-sulfonyl fluoride (PMSF). Clarified lysate was loaded onto 1 mL nickel-nitrilotriacetic acid agarose resin (Qiagen) by gravity flow, then washed with binding buffer followed by washing buffer (binding buffer + 30 mM imidazole). An exception was yBL, for which the washing buffer contained only 10 mM imidazole. Fractions containing the highest concentrations of ligase were pooled and dialyzed into phosphate-buffered saline (PBS) pH 7.4. Purified ligases were stored in aliquots at -80 °C.

For *M. jannaschii* and *P. horikoshii* biotin ligases, purification was carried out by heat-selective precipitation following the protocol of Bagautdinov *et al.*[40] Briefly, cells were resuspended in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM DTT, and 1 mM PMSF, then lysed by sonication. Clarified lysate was heated to 90 °C for 11.5 minutes, then denatured proteins were removed by centrifugation at 17,000 g for 20 minutes. Supernatant containing the desired enzyme was dialyzed into PBS pH 7.4, then aliquoted and stored at -80 °C.

For human holocarboxylase synthetase, cells were lysed by sonication in loading buffer (50 mM HEPES pH 7.4, 300 mM NaCl) in the presence of 1 mM PMSF. Clarified lysate was loaded onto glutathione agarose resin (Sigma), washed with loading buffer, then eluted with loading buffer containing 6.4 mg/mL reduced glutathione. Concentrated fractions were pooled and dialyzed into PBS.

HPLC analysis of AP peptide modification

The following reaction conditions were used: 1 μ M enzyme was incubated with 1 mM biotin, 4 mM ATP, 50 μ M K₃AP (a previously reported AP peptide with 3 lysine residues appended to improve HPLC peak shapes), [2] 5 mM Mg(OAc)₂, 50 mM bicine buffer (pH 8.0). Reactions were incubated for 15 hours at 30°C. 50 μ L of the reaction mixture was injected onto a C18 reverse phase column (Microsorb-MV 300-5, 250 x 4.6 mm, Varian) and separated with a mobile phase gradient of 10 to 40 % acetonitrile in water with 0.1% trifluoroacetic acid over 25 minutes. Absorption was monitored at 210 nm. Unmodified K₃AP had a retention time of 20 minutes, and biotinylated K₃AP had a retention time of 22 minutes.

Expression and purification of p67 acceptor protein

p67 is a protein domain comprising the 67 C-terminal amino acids of human propionyl CoA carboxylase.[60] The p67 expression plasmid, in pDEST11 vector, was a gift from Roy Gravel. The plasmid was transformed into *E. coli* BL21 DE3 cells and individual colonies were amplified in 6 liters of LB with 100 μ g/mL ampicillin, until OD 0.8. Expression was induced with 420 μ M IPTG for 3 hours at 30 °C. Cells were harvested and lysed as described above, and p67 was purified using nickel affinity chromatography with 200 mM imidazole for the final elution.

Before use in biotinylation assays, we removed pre-biotinylated p67 (caused by endogenous BirA during the protein expression) by incubating the entire batch of purified p67 with 750 μ L

streptavidin-agarose resin (Novagen) at 4 °C for 1 hour with rocking. The slurry was then filtered through a fritted column (Biorad) by gravity. Biotinylated p67 was retained on the column, and the non-biotinylated p67 was stored in aliquots at -80 °C.

The K61A mutant of p67, called p67(Ala), was generated by QuikChange with the primer 5'gggcaagaaatttgtgtgattgaagccatggcaatgcagaatagtagacgc and its reverse complement.

Screening of biotin ligases against biotin analogs

The following reaction conditions were used in the screening assays: 1 μM of biotin ligase enzyme, 1 mM biotin analog, 100 μM p67 protein, 5 mM Mg(OAc)₂, 5 mM ATP, and 50 mM bicine pH8.3. The reaction mixture was incubated for 14 hours at 30 °C. Formation of product was detected by either HPLC (for *P. acnes*, yeast, human, *M. jannaschii*, *P. horikoshii*, *T. cruzi*, and *B. subtilis* biotin ligases) or native gel shift assay (for *L. mesenteroides* and yeast biotin ligases), as described below.

HPLC assay to detect probe conjugation to acceptor protein

Reactions for DTB-Az ligation contained 5 μM of *P. horikoshii* biotin ligase, 10 mM DTB-Az, 100 μM p67, 5 mM Mg(OAc)₂, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions for *cis* PB ligation contained 1 μM of yeast biotin ligase, 8 mM *cis*-PB, 100 μM p67, 5 mM Mg(OAc)₂, 5 mM ATP, and 50mM bicine pH 8.3. Reactions were incubated for 14 hours at 30 °C. 50 μL of the reaction mixture was injected onto a C18 reverse phase column (Microsorb-MV 300-5, 250 x 4.6 mm, Varian) and separated with a mobile phase gradient of 30 to 50 % acetonitrile in water with

0.1% trifluoroacetic acid over 30 minutes. Absorption was monitored at 210 nm. Unmodified p67 had a retention time of 14.9 minutes, and the probe-p67 conjugates ranged from 16.9 to 17.9 minutes.

Native gel shift assay to detect probe conjugation to acceptor protein

20 μ L of each reaction mixture, prepared as above, was loaded onto a non-denaturing 20% polyacrylamide gel and electrophoresed at 100 V for 3.5 hours. The electrophoresis apparatus was submerged in ice to prevent protein unfolding. Gels were stained with Coomassie brilliant blue. Modified p67 demonstrated a characteristic shift toward greater mobility relative to unmodified p67.

Mass-spectrometric analysis of probe conjugation to acceptor protein

Reactions of *P. horikoshii* biotin ligase with DTB-Az contained 1 μ M enzyme, 100 μ M p67, 10 mM DTB-Az, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions of *S. cerevisiae* biotin ligase with *cis*-PB contained 1 μ M enzyme, 100 μ M p67, 5 mM *cis*-PB or 1 mM biotin, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions of *P. horikoshii* biotin ligase with DTB-Az and K61A p67 (to test site specificity) contained 5 μ M enzyme, 100 μ M K61A p67, 10 mM DTB-Az, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions of yeast biotin ligase with *cis*-PB and K61A p67 contained 1 μ M enzyme, 8 mM *cis*-PB, 100 μ M p67, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions were incubated at 30 $^{\circ}$ C for 14 hours, and separated by HPLC as described above. Product fractions were collected by hand

then injected directly onto an Applied Biosystems 200 QTRAP mass spectrometer with electrospray ionization at a flow rate of 10 $\mu\text{L}/\text{min}$ with detection in positive ion mode.

Staudinger ligation to detect ligated DTB-Az

PhBL was used to label p67 with DTB-Az in a reaction containing 1 μM ligase, 100 μM p67, 500 μM DTB-Az, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions were incubated at 30 $^\circ\text{C}$ for 14 hours. For the Staudinger ligation, FLAG-phosphine[68] was directly added to the reaction mixture to a final concentration of 1.5 mM, then incubated at 25 $^\circ\text{C}$ for 12 hours. The reaction mixture was divided in half and analyzed on two 20% SDS-PAGE gels. One gel was stained with Coomassie brilliant blue to demonstrate equal protein loading. The contents of the other gel were transferred to nitrocellulose. The FLAG epitope was detected by staining with anti-FLAG (M2) antibody-horseradish peroxidase conjugate (Sigma, 1:2000 dilution). The blot was developed with SuperSignal West Pico Chemiluminescent Substrate (Pierce).

Kinetic analysis of DTB-Az and cis-PB ligations

We compared the rates of DTB-Az and biotin ligations to p67, catalyzed by PhBL under identical conditions. The measured DTB-Az ligation rate does not represent the maximal ligation velocity (V_{max}), because we found that we could not saturate the ligase with DTB-Az at the highest experimentally achievable DTB-Az concentration (10 mM) (data not shown). Reactions contained 1 μM PhBL, 100 μM p67, 1 mM biotin or DTB-Az, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH8.3. Reactions were incubated at 35 $^\circ\text{C}$, then quenched at various time points and

analyzed by HPLC as described above. Measurements were performed in triplicate, and error bars represent one standard deviation.

The V_{\max} and k_{cat} of *cis*-PB and biotin ligations to p67, catalyzed by yBL, were determined as follows. For biotin, reactions contained 45 nM yBL, 100 μM p67, 1 mM biotin, 5 mM $\text{Mg}(\text{OAc})_2$, 5mM ATP, and 50 mM bicine pH 8.3. For *cis*-PB, reactions contained 400 nM yBL, 100 μM p67, 5 mM *cis*-PB, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions were incubated at 35 °C, then quenched at various time points and analyzed by HPLC as described above.

Measurements were performed in triplicate, and error bars represent one standard deviation.

Cloning, purification, and assay for P. horikoshii BCCP

Purified *P. horikoshii* genomic DNA (ATCC) was obtained and used as the cloning template. *P. horikoshii* BCCP was PCR-amplified from the genomic DNA using the forward primer AAAAGGATTCATGATGAGGATGAAAGTGAAAGTTGT and the reverse primer TTTTAAGCTTCCCCAATTCTATTAGTGGTTGTCCT, and cloned into the BamHI and HindIII sites of pET21a to append a C-terminal hexahistidine tag. PhBCCP was overexpressed in *E. coli* and purified as described for hexahistidine-tagged biotin ligase enzymes and its concentration was determined using the BCA assay.

PhBL was used to label PhBCCP with DTB-Az in a reaction containing 1 μM PhBL, 100 μM PhBCCP, 500 μM DTB-Az, 5 mM $\text{Mg}(\text{OAc})_2$, 5 mM ATP, and 50 mM bicine pH 8.3. Reactions were incubated at 30 °C for 14 hours. As a positive control for mobility, biotin replaced DTB-Az

in the reaction mixture; single-omission negative controls were performed. Reaction products were analyzed by gel shift assay exactly as described above.

References

1. Howarth, M., et al., *Targeting quantum dots to surface proteins in living cells with biotin ligase*. Proc.Natl.Acad.Sci.U.S.A, 2005. **102**(21): p. 7583-7588.
2. Chen, I., et al., *Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase*. Nat.Methods, 2005. **2**(2): p. 99-104.
3. Chen, I. and Massachusetts Institute of Technology. Dept. of Chemistry., *Site-specific labeling of cellular proteins with unnatural substrates of biotin ligases*. 2007. p. 2 v. (308 leaves).
4. Choi, Y.-A. and Massachusetts Institute of Technology. Dept. of Chemistry., *Molecular engineering of new protein labeling methodology based on rational design and in vitro evolution*. 2010. p. 171 p.
5. Jager, S., et al., *A versatile toolbox for variable DNA functionalization at high density*. J.Am.Chem.Soc., 2005. **127**(43): p. 15071-15082.
6. Kopp, M., et al., *SorF: a glycosyltransferase with promiscuous donor substrate specificity in vitro*. Chembiochem., 2007. **8**(7): p. 813-819.
7. Cronan, J.E., Jr., *Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins*. J.Biol.Chem., 1990. **265**(18): p. 10327-10333.
8. Beckett, D., *Biotin sensing: universal influence of biotin status on transcription*. Annu.Rev.Genet., 2007. **41**: p. 443-464.
9. Streaker, E.D. and D. Beckett, *Coupling of protein assembly and DNA binding: biotin repressor dimerization precedes biotin operator binding*. J.Mol.Biol., 2003. **325**(5): p. 937-948.
10. Streaker, E.D. and D. Beckett, *The biotin regulatory system: kinetic control of a transcriptional switch*. Biochemistry, 2006. **45**(20): p. 6417-6425.
11. Rodionov, D.A., A.A. Mironov, and M.S. Gelfand, *Conservation of the biotin regulon and the BirA regulatory signal in eubacteria and archaea*. Genome Research, 2002. **12**(10): p. 1507-1516.
12. Earl, A.M., R. Losick, and R. Kolter, *Ecology and genomics of Bacillus subtilis*. Trends in Microbiology, 2008. **16**(6): p. 269-275.
13. Bower, S., et al., *Cloning and characterization of the Bacillus subtilis birA gene encoding a repressor of the biotin operon*. J.Bacteriol., 1995. **177**(9): p. 2572-2575.
14. Bower, S., et al., *Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon*. Journal of Bacteriology, 1996. **178**(14): p. 4122-4130.
15. Purushothaman, S., et al., *Ligand specificity of group I biotin protein ligase of Mycobacterium tuberculosis*. PLoS.One., 2008. **3**(5): p. e2320.
16. Clarke, D.J., et al., *Biotinylation in the hyperthermophile Aquifex aeolicus*. Eur.J.Biochem., 2003. **270**(6): p. 1277-1287.
17. Tron, C.M., et al., *Structural and functional studies of the biotin protein ligase from Aquifex aeolicus reveal a critical role for a conserved residue in target specificity*. J.Mol.Biol., 2009. **387**(1): p. 129-146.
18. Shenoy, B.C. and H.G. Wood, *Purification and properties of the synthetase catalyzing the biotination of the aposubunit of transcarboxylase from Propionibacterium shermanii*. FASEB J., 1988. **2**(8): p. 2396-2401.

19. Dupuis, C., C. Corre, and P. Boyaval, *Lipase and Esterase-Activities of Propionibacterium-Freudenreichii Subsp Freudenreichii*. Applied and Environmental Microbiology, 1993. **59**(12): p. 4004-4009.
20. Cazzulo, J.J., et al., *Synthesis of pyruvate carboxylase from its apoenzyme and (+)-biotin in Bacillus stearothermophilus. Purification and properties of the apoenzyme and the holoenzyme synthetase*. Biochem.J., 1971. **122**(5): p. 653-661.
21. Brison, Y., et al., *Synthesis of dextrans with controlled amounts of alpha-1,2 linkages using the transglucosidase GBD-CD2*. Applied Microbiology and Biotechnology, 2010. **86**(2): p. 545-554.
22. Nieto-Arribas, P., et al., *Genotypic and technological characterization of Leuconostoc isolates to be used as adjunct starters in Manchego cheese manufacture*. Food Microbiology, 2010. **27**(1): p. 85-93.
23. Whiteside-Carlson, V., et al., *Non-specificity of biotin activity for leuconostoc*. Proc.Soc.Exp.Biol.Med., 1951. **77**(2): p. 344-348.
24. Gonzalez, J.M., et al., *Pyrococcus horikoshii sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough*. Extremophiles, 1998. **2**(2): p. 123-130.
25. Cronan, J.E., Jr. and J.C. Wallace, *The gene encoding the biotin-apoprotein ligase of Saccharomyces cerevisiae*. FEMS Microbiol.Lett., 1995. **130**(2-3): p. 221-229.
26. Polyak, S.W., et al., *Biotin protein ligase from Saccharomyces cerevisiae. The N-terminal domain is required for complete activity*. J.Biol.Chem., 1999. **274**(46): p. 32847-32854.
27. Pardini, N.R., et al., *Biotin protein ligase from Candida albicans: expression, purification and development of a novel assay*. Arch.Biochem.Biophys., 2008. **479**(2): p. 163-169.
28. Alban, C., *Is plant biotin holocarboxylase synthetase a bifunctional enzyme?* C.R.Acad.Sci.III, 2000. **323**(8): p. 681-688.
29. Denis, L., et al., *Molecular characterization of a second copy of holocarboxylase synthetase gene (hcs2) in Arabidopsis thaliana*. J.Biol.Chem., 2002. **277**(12): p. 10435-10444.
30. Puyaubert, J., L. Denis, and C. Alban, *Dual targeting of Arabidopsis holocarboxylase synthetase1: a small upstream open reading frame regulates translation initiation and protein targeting*. Plant Physiol, 2008. **146**(2): p. 478-491.
31. Tissot, G., R. Douce, and C. Alban, *Evidence for multiple forms of biotin holocarboxylase synthetase in pea (Pisum sativum) and in Arabidopsis thaliana: subcellular fractionation studies and isolation of a cDNA clone*. Biochem.J., 1997. **323** (Pt 1): p. 179-188.
32. Tissot, G., et al., *Purification and properties of the chloroplastic form of biotin holocarboxylase synthetase from Arabidopsis thaliana overexpressed in Escherichia coli*. Eur.J.Biochem., 1998. **258**(2): p. 586-596.
33. Tissot, G., et al., *Protein biotinylation in higher plants: characterization of biotin holocarboxylase synthetase activity from pea (Pisum sativum) leaves*. Biochem.J., 1996. **314** (Pt 2): p. 391-395.
34. Xia, W.L., J. Zhang, and F. Ahmad, *Biotin Holocarboxylase Synthetase - Purification from Rat-Liver Cytosol and Some Properties*. Biochemistry and Molecular Biology International, 1994. **34**(2): p. 225-232.
35. Chiba, Y., et al., *Purification and Properties of Bovine Liver Holocarboxylase Synthetase*. Archives of Biochemistry and Biophysics, 1994. **313**(1): p. 8-14.
36. Dupuis, L., et al., *Clustering of mutations in the biotin-binding region of holocarboxylase synthetase in biotin-responsive multiple carboxylase deficiency*. Human Molecular Genetics, 1996. **5**(7): p. 1011-1016.
37. Campeau, E. and R.A. Gravel, *Expression in Escherichia coli of N- and C-terminally deleted human holocarboxylase synthetase. Influence of the N-terminus on biotinylation and identification of a minimum functional protein*. J.Biol.Chem., 2001. **276**(15): p. 12310-12316.

38. Ingaramo, M. and D. Beckett, *Distinct amino termini of two human HCS isoforms influence biotin acceptor substrate recognition*. J.Biol.Chem., 2009. **284**(45): p. 30862-30870.
39. Chapman-Smith, A., et al., *Molecular recognition in a post-translational modification of exceptional specificity. Mutants of the biotinylated domain of acetyl-CoA carboxylase defective in recognition by biotin protein ligase*. J.Biol.Chem., 1999. **274**(3): p. 1449-1457.
40. Bagautdinov, B., et al., *Crystal structures of biotin protein ligase from Pyrococcus horikoshii OT3 and its complexes: structural basis of biotin activation*. J.Mol.Biol., 2005. **353**(2): p. 322-333.
41. Daniels, K.G. and D. Beckett, *Biochemical Properties and Biological Function of a Monofunctional Microbial Biotin Protein Ligase*. Biochemistry, 2010.
42. Wilson, K.P., et al., *Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains*. Proc.Natl.Acad.Sci.U.S.A, 1992. **89**(19): p. 9257-9261.
43. Weaver, L.H., et al., *Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator*. Proc.Natl.Acad.Sci.U.S.A, 2001. **98**(11): p. 6045-6050.
44. Wood, Z.A., et al., *Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution*. J.Mol.Biol., 2006. **357**(2): p. 509-523.
45. Chapman-Smith, A., et al., *The C-terminal domain of biotin protein ligase from E. coli is required for catalytic activity*. Protein Sci., 2001. **10**(12): p. 2608-2617.
46. Slavoff, S.A., et al., *Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species*. J.Am.Chem.Soc., 2008. **130**(4): p. 1160-1162.
47. Green, N.M., *Thermodynamics of the binding of biotin and some analogues by avidin*. Biochem.J., 1966. **101**(3): p. 774-780.
48. Meisenheimer, K.M. and T.H. Koch, *Photocross-linking of nucleic acids to associated proteins*. Crit Rev.Biochem.Mol.Biol., 1997. **32**(2): p. 101-140.
49. Gore, J., C. Hoinard, and P. Maingault, *Biotin uptake by isolated rat intestinal cells*. Biochim Biophys Acta, 1986. **856**(2): p. 357-61.
50. Park, S. and P.J. Sinko, *The blood-brain barrier sodium-dependent multivitamin transporter: a molecular functional in vitro-in situ correlation*. Drug Metab Dispos, 2005. **33**(10): p. 1547-54.
51. Waller, J.R., *Increased sensitivity of the microbiological assay for biotin by Lactobacillus plantarum*. Appl Microbiol, 1970. **20**(3): p. 485-91.
52. Cicmanec, J.F. and H.C. Lichstein, *Uptake of extracellular biotin by Escherichia coli biotin prototrophs*. J Bacteriol, 1978. **133**(1): p. 270-8.
53. Fernandez-Suarez, M., T.S. Chen, and A.Y. Ting, *Protein-protein interaction detection in vitro and in cells by proximity biotinylation*. J.Am.Chem.Soc., 2008. **130**(29): p. 9251-9253.
54. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. Bioinformatics, 2007. **23**(21): p. 2947-8.
55. Grosveld, F., et al., *Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry*. Ann.N.Y.Acad.Sci., 2005. **1054**: p. 55-67.
56. Chen, I., Y.A. Choi, and A.Y. Ting, *Phage display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins*. J.Am.Chem.Soc., 2007. **129**(20): p. 6619-6625.
57. Duffy, S., K.L. Tsao, and D.S. Waugh, *Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides*. Analytical Biochemistry, 1998. **262**(2): p. 122-8.
58. Athavankar, S. and B.R. Peterson, *Control of gene expression with small molecules: biotin-mediated acylation of targeted lysine residues in recombinant yeast*. Chem Biol, 2003. **10**(12): p. 1245-53.

59. Nenortas, E. and D. Beckett, *Purification and characterization of intact and truncated forms of the Escherichia coli biotin carboxyl carrier subunit of acetyl-CoA carboxylase*. J.Biol.Chem., 1996. **271**(13): p. 7559-7567.
60. Leon-Del-Rio, A. and R.A. Gravel, *Sequence requirements for the biotinylation of carboxyl-terminal fragments of human propionyl-CoA carboxylase alpha subunit expressed in Escherichia coli*. J Biol Chem, 1994. **269**(37): p. 22964-8.
61. Chapman-Smith, A., et al., *Expression, biotinylation and purification of a biotin-domain peptide from the biotin carboxy carrier protein of Escherichia coli acetyl-CoA carboxylase*. Biochem.J., 1994. **302 (Pt 3)**: p. 881-887.
62. Jordan, I.K., et al., *Phylogenomic analysis of the Giardia intestinalis transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes*. J Mol Microbiol Biotechnol, 2003. **5**(3): p. 172-89.
63. Cronan, J.E., *Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase*. J.Nutr.Biochem., 2005. **16**(7): p. 416-418.
64. Streaker, E.D. and D. Beckett, *Nonenzymatic biotinylation of a biotin carboxyl carrier protein: unusual reactivity of the physiological target lysine*. Protein Sci., 2006. **15**(8): p. 1928-1935.
65. Hong, V., et al., *Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation*. Angew Chem Int Ed Engl, 2009. **48**(52): p. 9879-83.
66. Agard, N.J., et al., *A comparative study of bioorthogonal reactions with azides*. ACS Chem.Biol., 2006. **1**(10): p. 644-648.
67. Saxon, E. and C.R. Bertozzi, *Cell surface engineering by a modified Staudinger reaction*. Science, 2000. **287**(5460): p. 2007-10.
68. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*. Proc Natl Acad Sci U S A, 2002. **99**(1): p. 19-24.
69. Kwon, K. and D. Beckett, *Function of a conserved sequence motif in biotin holoenzyme synthetases*. Protein Sci., 2000. **9**(8): p. 1530-1539.

Chapter 3: Delivering quantum dots into the cytosol of living cells

Dr. Takashi Kawakami performed intracellular protein labeling with quantum dots delivered using streptolysin O.

Introduction

Single-molecule imaging using quantum dots (QDs) has proved exceptionally powerful in the study of the trafficking of cell-surface receptors. The advent of biotin ligase (BirA)-mediated targeting of streptavidin-functionalized QDs to biotinylated acceptor peptide (AP)-tagged proteins has simplified QD targeting and facilitated long-term single molecule imaging.[1, 2] However, because QDs are large (at least 20 nm in diameter for commercial dots),[3] they cannot cross the cell membrane, so their potential for investigating intracellular proteins at the single-molecule level remains untapped. As discussed in Chapter 1, many methods have been explored for delivery of QDs into living cells; however, most methods are subject to toxicity, QD aggregation, or trapping of QDs in endocytic vesicles. Microinjection is currently widely utilized, and microinjected QDs have begun to make contributions to intracellular single-molecule imaging.[4, 5] However, microinjection is technically difficult and requires specialized equipment, so is generally accessible only to the accomplished cell biologist. Aside from microinjection, only osmotic lysis of pinocytic vesicles, and streptolysin O permeabilization have thus far afforded freely diffusible QDs that are well-distributed in the cytosol, and then with varying degrees of characterization and confidence.[6-9] However, pinosome lysis and streptolysin O are subject to toxicity and are relatively poorly characterized as QD delivery systems. Therefore a simple, inexpensive, non-toxic method to efficiently deliver QDs into cells, then target them to biotinylated proteins, is still needed. In this chapter we describe our efforts toward finding a method to deliver non-cell permeable imaging reagents, and in particular QDs, into the cytosol of living cells as a step toward this goal. We present this chapter in two parts, each describing one delivery method we investigated. In Part I, we discuss our attempts to

harness viral cell entry mechanisms for delivery of QDs. In Part II, we discuss our validation, carried to fruition by Dr. Takashi Kawakami, of the bacterial pore-forming toxin streptolysin O for QD delivery into the cytosol of living cells.

Part I: Investigation of Influenza Virosomes for Quantum Dot Delivery

Introduction

Viral delivery has been used extensively in transfection and gene therapy. This is because viruses are exceedingly adept at efficiently accessing the interior of cells. During infection, viruses must be internalized into a host cell, escape into the cytosol, then translocate their genetic material into the nucleus in order to activate host cell transcription and translation of the viral genome. The mechanism of this process has been particularly well characterized for influenza virus.[10]

Influenza is an enveloped virus, consisting of a ribonucleoprotein genome encapsulated in a lipid bilayer derived from the host cell. The major protein substituents of the lipid bilayer are hemagglutinin (HA) and neuraminidase, which are both required for infectivity.[11] Enveloped viruses enter cells *via* membrane fusion, either at the cell surface or with endosomal membranes. The influenza viral particle, or virion, binds to the surface of a host cell through interaction of HA with sialic acid residues on cell-surface glycoproteins.[12] The virus is then internalized *via* both clathrin-dependent and -independent endocytosis.[13] As the endosome matures, the pH decreases, causing HA to undergo a conformational change,[14, 15] exposing the HA2 fusion peptide that inserts into the endosomal membrane.[16] The viral membrane then fuses with the endosomal membrane, first forming a hemi-fusion intermediate in which the outer layers of both membranes fuse, followed by collapse to the fully fused, mixed membrane.[17] This fusion allows the internal contents of the virion to escape into the cytosol.

The mechanism of influenza virus cell entry has previously been exploited in delivering non-cell-permeable cargo into cells by creating artificial viruses called virosomes. Virosomes have been prepared by inserting purified hemagglutinin into artificial liposomes, by fusing viruses with liposomes, and by reconstituting viral membranes.[18] By encapsulating cargo inside virosome membranes, intracellular delivery can be affected; virosomes have been prepared containing and used to deliver DNA, siRNA, and protein toxins into cells.[19-22] We note that cell transfection and killing by protein toxins requires only very low delivery efficiencies into the cytosol, so the real efficiency of cargo delivery by virosomes remained ill-characterized. Furthermore, to our knowledge, no attempt to deliver QDs into cells by viral means has been reported; we therefore proposed to utilize influenza virosomes to deliver QDs, as depicted in Figure 3-1. We expected that it is physically possible for at least one QD to be encapsulated per virosome, as reconstituted virosomes have been measured by electron microscopy to be 100-200 nm in diameter.[23] Because we had access to the assistance of kind collaborators in the laboratories of Xiaowei Zhuang at Harvard and Antoine van Oijen at Harvard Medical School who were at the time beginning to investigate influenza virosomes, we elected to begin our investigation of QD delivery methods with the influenza virosome system.

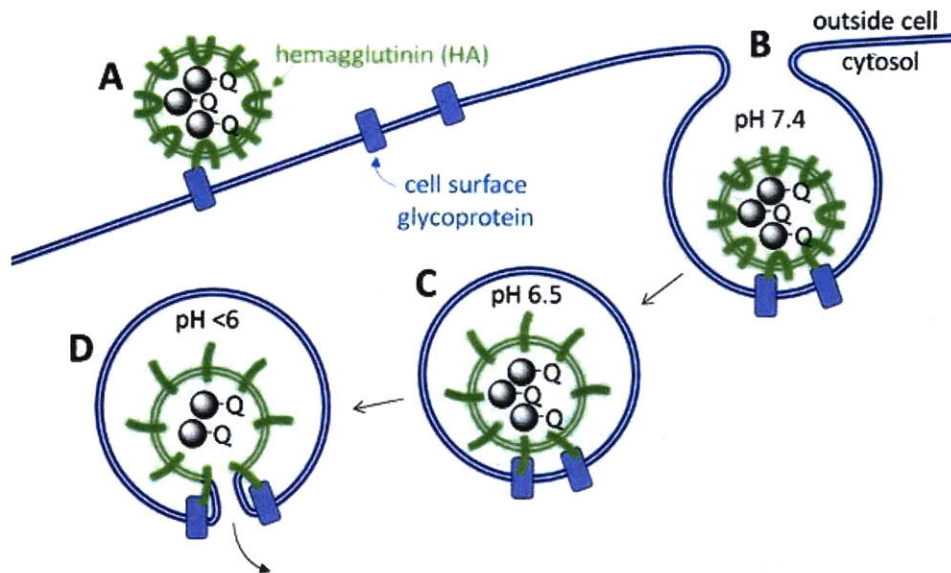


Figure 3-1. Proposed mechanism of virosome-mediated quantum dot (QD) delivery. (A) Virosomes, or reconstituted influenza viral envelopes, containing QDs bind to cell surface glycoprotein-displayed sialic acid residues and become internalized in endocytic vesicles (B). (C) As endosomes mature, the pH decreases, causing a conformational change in the influenza hemagglutinin protein. (D) The hemagglutinin conformational change causes fusion of the virosome membrane with the endosomal membrane, freeing the contents of the virosome into the cytosol.

While cell entry by influenza virus is efficient, it is not perfect; approximately 25-50% of virions that initially attach to the cell surface eventually reach the cytosol.[13] Since viral delivery, along with most delivery methods, was originally formulated for gene delivery, the actual efficiency of cargo release from endosomes is ill-characterized; for gene transfection, only one copy of DNA needs to reach the nucleus, and any trapped material remains invisible. Recalling that any imaging reagents that remain trapped in endocytic vesicles will be a source of background that cannot be removed in imaging experiments, any virosomes that cannot escape from vesicles could potentially be problematic. We therefore approached this project with an eye toward determining whether influenza virosomes produce endocytically trapped QDs, or free cytosolic QDs.

Results, discussion, and conclusions

In order to prepare QD-containing virosomes, we utilized a protocol originally reported by Stegmann *et al.* in 1987,[11] with the kind assistance of the laboratory of Professor Xiaowei Zhuang at Harvard and the laboratory of Professor Antoine van Oijen at Harvard Medical School. The protocol is described in detail in the Methods section of this chapter. Briefly, influenza virus is solubilized in non-denaturing detergent and the viral ribonucleoprotein particle is removed by ultracentrifugation; this removal of the viral genome eliminates infectivity. A rhodamine green-stearic acid conjugate (a kind gift from Professor Antoine van Oijen) or fluorescein-phosphatidylethanolamine is added to the solubilized membranous material; these dyes become incorporated into the membrane upon reconstitution and act as markers for viral particles. Commercial red CdSe core QDs with polyethylene glycol-conjugated, amine-functionalized ligands for biocompatibility and solubilization (Invitrogen) are also added to this solubilized mixture. The viral membranes are then reconstituted by vortexing in the presence of sorbent beads that rapidly remove the detergent, forcing the membrane substituents (e.g. lipids and membrane proteins such as hemagglutinin) to associate *via* hydrophobic interactions. Cargo, in this case QDs, should be encapsulated randomly. The intact virosomes are then purified by ultracentrifugation on a sucrose step gradient. We determined through experimentation that the optimal concentration of QDs to add to the reconstitution mixture is 50 nM (data not shown); lesser concentrations produce little apparent encapsulation, and higher concentrations interfere with membrane reconstitution. A schematic of the virosome

preparation protocol is provided in Figure 3-2.

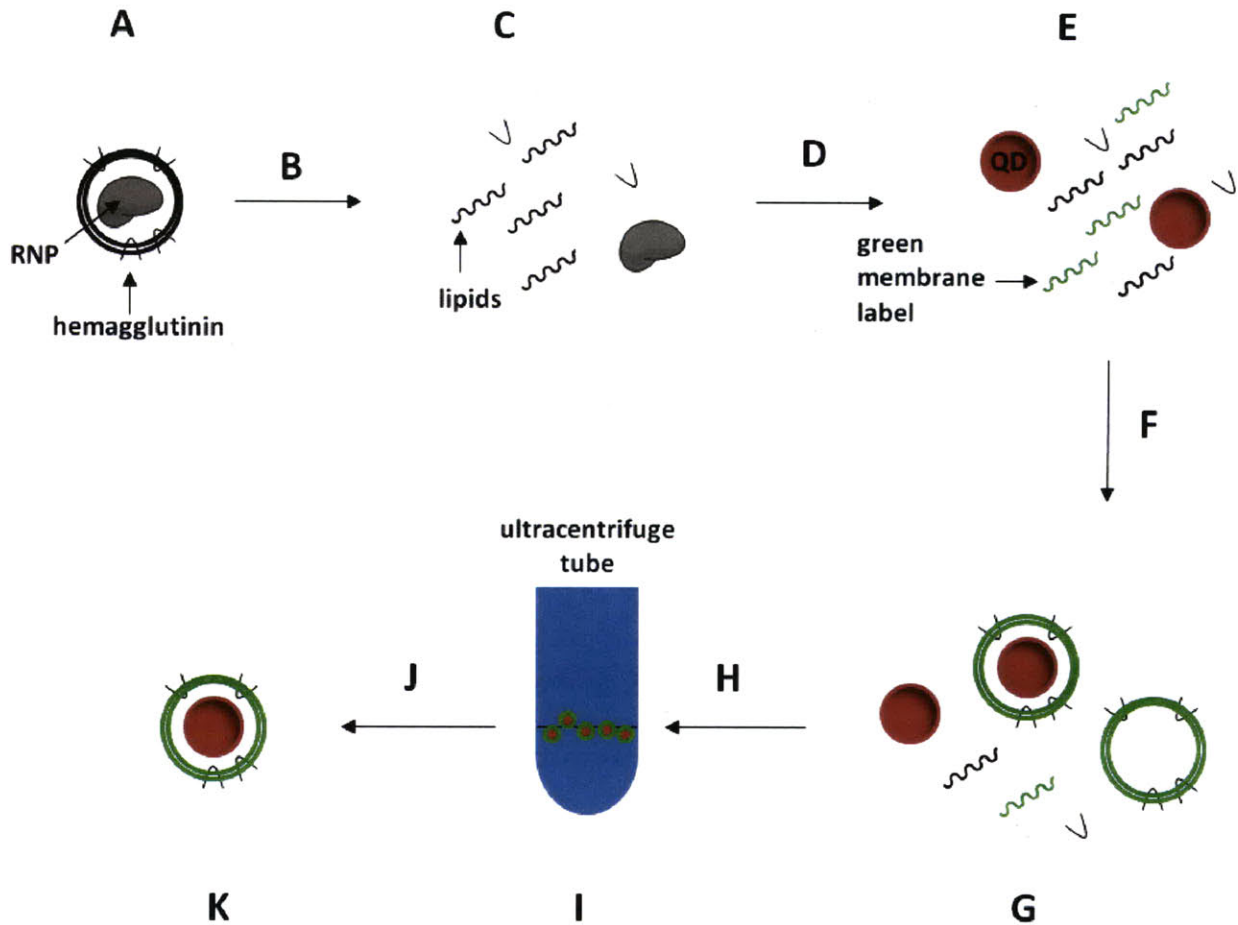


Figure 3-2. Schematic of virosome preparation protocol. (A) Native influenza virus consists of a lipid bilayer containing hemagglutinin protein, represented as a U-shaped molecule, enveloping a ribonucleoprotein (RNP) genome (shown in gray). (B) Solubilization of the virion with non-denaturing detergent affords (C) a mixture of viral membrane lipids (shown as wavy lines), membrane proteins including hemagglutinin, and the RNP. (D) Ultracentrifugation affords removal of the RNP and elimination of infectivity. (E) The resulting mixture of lipid and protein is doped with a green lipophilic membrane dye to report on eventual membrane reconstitution as well as QDs. (F) Addition of sorbent beads facilitates rapid removal of the non-denaturing detergent, yielding (G) a mixture of re-formed viral envelopes, viral envelopes that stochastically encapsulated QDs (the desired result), and free QDs and membrane substituents. (H) Ultracentrifugation on a sucrose step gradient purifies virosomes and QD-containing virosomes at the interface, depicted here on a schematic centrifuge tube (I); optimizing the amount of QDs added gives the maximal yield of virosome-encapsulated QDs at this step and minimizes contaminating free QDs. The final product (K) is intended to be primarily QD-containing viral envelopes, labeled with the green membrane dye and containing fusion-active hemagglutinin.

If we successfully reconstituted virosomes, then imaging should show that the products of our preparation are individual particles. Furthermore, if the products of our preparation protocol are exactly the virosomes reported in the literature, we expect them to be 100-nm-diameter particles, and so we would expect to see monodisperse, diffraction-limited spots under the microscope; these spots would be green, because of incorporation of the green membrane dye. Finally, imaging could tell us whether the virosome particles are associated with QDs; this would be visualized as co-localization of the green particles with the red QD fluorescence. We therefore used an imaging assay to determine if we had successfully produced individual virosome particles, and if QDs were associated with, and hopefully encapsulated within, our virosomes (Figure 3-3). First, we note that we observe green particles, as expected; however, their size is not monodisperse. Some spots are larger than single pixels under the 40x objective, indicating that they are larger than diffraction-limited spots; smaller particles are also present. This indicates that larger aggregates may be formed by this protocol. Essentially complete co-localization of green and red spots indicates that most of the virosomes are associated with QDs. Finally, while individual QDs exhibit a phenomenon known as blinking,[2] in which their fluorescence spontaneously turns on and off, we note that we do not observe blinking of our virosomes, meaning that each is associated with more than one QD.

We note that this assay does not report on QD encapsulation in well-formed lipid bilayers, by any means; co-localization could be due to QDs associating with the outside of lipid bilayers, or alternatively QDs sticking to aggregates of lipid and protein. Indeed, the varied size of the virosomes indicates inhomogeneity in the products of the preparation, probably including some

larger aggregates. (For semantic ease we continue to refer to the products of our preparation as virosomes while acknowledging this inhomogeneity.)

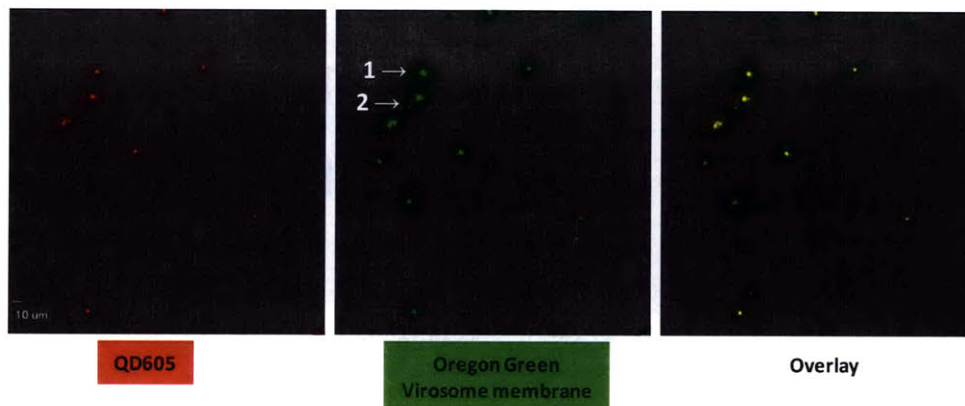


Figure 3-3. Imaging virosomes reconstituted in the presence of QDs. Influenza viral membranes were purified and reconstituted in the presence of Oregon Green-derivatized stearic acid, to dye the virosome membrane, and QD605. The QD channel is shown on the left in red, the virosome membrane channel is shown in the center in green, and an overlay of the two is shown on the right. Imaging reveals that the preparation affords individual virosome particles (Oregon Green channel), and that many of these virosomes have QD signal associated (QD605 channel and overlay). However, the presence of (1) larger and (2) smaller spots indicates inhomogeneity and potential aggregation in the preparation. Scale bar, 10 μ M.

Since virosomes are reported to be smaller than the diffraction limit (~ 100 nm, whereas the diffraction limit is ~ 250 nm, as discussed in Chapter 1), further structural characterization (e.g., confirming QD encapsulation, comparing virosome sizes, and determining if they consist of well-formed, circularized lipid bilayers) would require electron microscopy. Rather than further structurally characterizing the virosomes, we elected to characterize their activity in terms of membrane fusion and quantum dot delivery, as we reasoned that, whatever their composition, the most important characteristic is whether they can fuse with biological membranes and deliver QDs into the cytosol.

The first question to ask of our virosomes is whether they have membrane fusion activity, presumably mediated by the hemagglutinin protein. In order to simplify this question, we utilized a surface fusion assay, in which artificial acidification is used to force the viral membrane to fuse with the plasma membrane of the cell (Figure 3-4). This simplifies analysis by eliminating endocytosis steps. The first step of the experiment is to allow virosomes to bind to the surface of HEK cells, which we do by treating the cells with virosomes in ice-cold phosphate-buffered saline (PBS), such that endocytosis does not occur. Then any unbound material is gently washed away using more ice-cold PBS, so that only virosomes that can specifically bind to the cell surface remain. If these cells are then incubated in PBS at pH 7.5 on ice, imaging reveals that the virosomes remain bound to the cell surface (Figure 3-4, top panels). Both the red fluorescence from the QDs and the green membrane dye are clearly localized to the plasma membrane. This serves as our negative control, showing that virosomes can remain stably bound to the cell surface at physiological pH if cells are kept at 4°C. Then, to induce fusion, we allowed virosomes to bind to the cell surface on ice and washed away unbound material as previously, but then switched the cells into ice-cold PBS at pH 5.5. In other words, these cells were treated identically to the negative control cells except that the pH of the incubation buffer was decreased to 5.5. This decreased pH forces hemagglutinin to undergo its fusogenic conformational change while still bound to the cell surface, and should force membrane fusion to occur at the plasma membrane. Imaging (Figure 3-4, bottom panels) surprisingly reveals that, after this treatment, the green fluorescence from the virosome membrane dye has is no longer detectable under identical imaging conditions. We expect this to be the case if membrane fusion has occurred and the dye has been diluted into the HEK cell membrane, although we do

not have direct evidence that this has occurred. Second, we observe that, while some QD fluorescence appears to remain at the cell surface, there are QDs observable in the cytosol (indicated by white arrows, Figure 3-4, bottom left). Note that, because of their large size, these QDs are excluded from passive diffusion into the nucleus and remain localized in the cytosol. We conclude from this experiment that, under harsh and forcing conditions, a fraction of our virosomes (though probably not the entire population) is capable of binding to the cell surface and fusing with biological membranes.

We note that there is batch-to-batch variation in fusogenicity of our virosomes. For example, a subsequent batch of virosomes exhibited cell surface binding, but clearly remained exclusively bound to the cell surface and un-fused (in particular because the green dye signal remained, un-attenuated) after cold acidification (data not shown). We will discuss possible sources of this batch-to-batch variability and propose methods to address it in the future in more detail at the end of this section (*vide infra*).

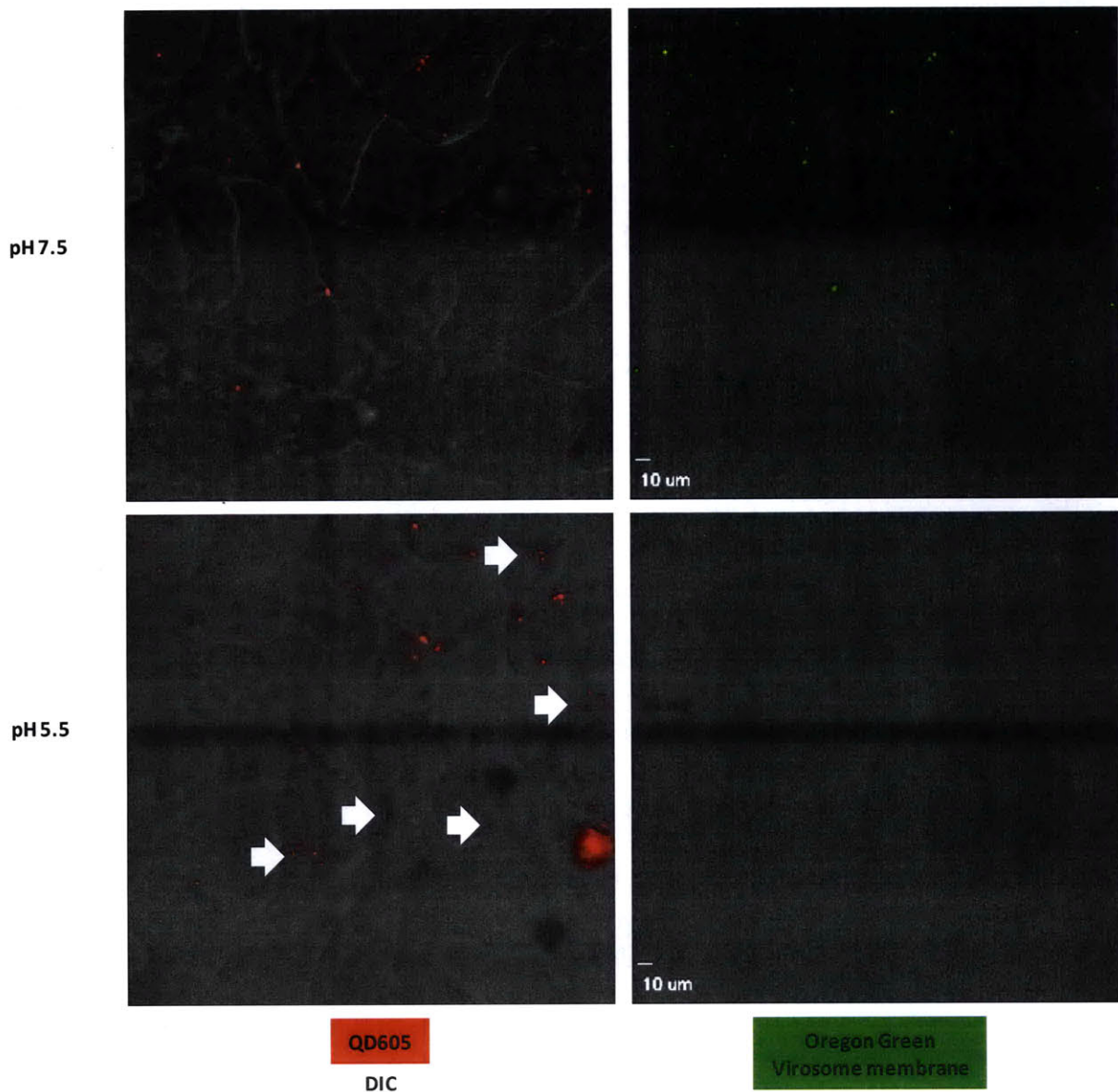


Figure 3-4. Cell-surface acidification assay for virosome fusogenicity. HEK cells were treated with QD-containing influenza virosomes on ice to prevent endocytosis. Incubating the cells on ice at physiological pH causes the virosomes to remain bound at the cell surface, as evidenced by the membrane localization of the QD and virosome membrane dye puncta (top row). Exchanging the cells into ice-cold buffer at pH 5.5 artificially forces the virosome membranes to fuse with the plasma membrane, as evidenced by the disappearance of the green membrane dye (possibly due to dilution into the plasma membrane of the HEK cells) and the intracellular localization of the QD fluorescence (bottom row). Some intracellular ODs are indicated by white arrows (bottom left).

Having established that some fraction of our virosomes has apparent fusogenic activity, we then wanted to see if QDs could be delivered into HEK cells by our virosomes during the course of normal endocytosis. We therefore interrogated the ability of our virosomes to be internalized after binding to the cell surface, and examined what happens to the QDs during this process (Figure 3-5). As in the previous experiment, in order to effect specific binding and remove unbound material, QD-containing virosomes were applied to cells on ice, allowing cell-surface binding to occur but preventing endocytosis, then unbound material was washed off on ice. (This was important because, at physiological temperatures, ill-formed, non-fusogenic virosomes that cannot associate with the cell surface could nonetheless be non-specifically pinocytosed, showing up as background in endosomes and preventing us from querying what happens to any cell-surface-bound virosomes, which presumably have active hemagglutinin.) As a negative control, as previously, cells were incubated in ice-cold PBS for 5 minutes and then imaged. Under these conditions, as seen in the top row of images in Figure 3-5, virosomes can associate with the cells and remain primarily bound to the cell surface (though during the imaging time, which is conducted at room temperature, some internalization does occur). The effects of endocytosis were then investigated. Another set of cells was treated with virosomes on ice, then washed and incubated in room-temperature PBS for 5 minutes. As seen in the second panel row of Figure 3-5, some of the virosomes have been internalized after this treatment, while most remain at the cell surface, much like our cold control, probably because endocytosis is occurring very slowly at this temperature. Finally, a third set of cells was treated with virosomes on ice, then incubated in 37°C PBS for 5 minutes. Under these conditions, essentially all of the virosomes are internalized (though some remain at the cell surface), as

seen in the third row of Figure 3-5. Importantly, under all conditions, the green fluorescence of the virosome membrane dye remains visible, unlike in our forced fusion experiment. As a negative control, HEK cells were treated with free quantum dots under identical conditions (that is, loading with quantum dots on ice, washing, and incubation at 37°C). Under these conditions, no free quantum dots bound to the cell surface or were internalized (Figure 3-5, bottom). The localization of the virosomes (apparently perinuclear, in many cells in the third row of Figure 3-5) after internalization at 37°C provides evidence that they are internalized rather than simply aggregated on the cell surface, as we will discuss in our next experiment. These experiments demonstrate that the viral membrane components of our virosome preparation are necessary to promote QD binding to and internalization into cells.

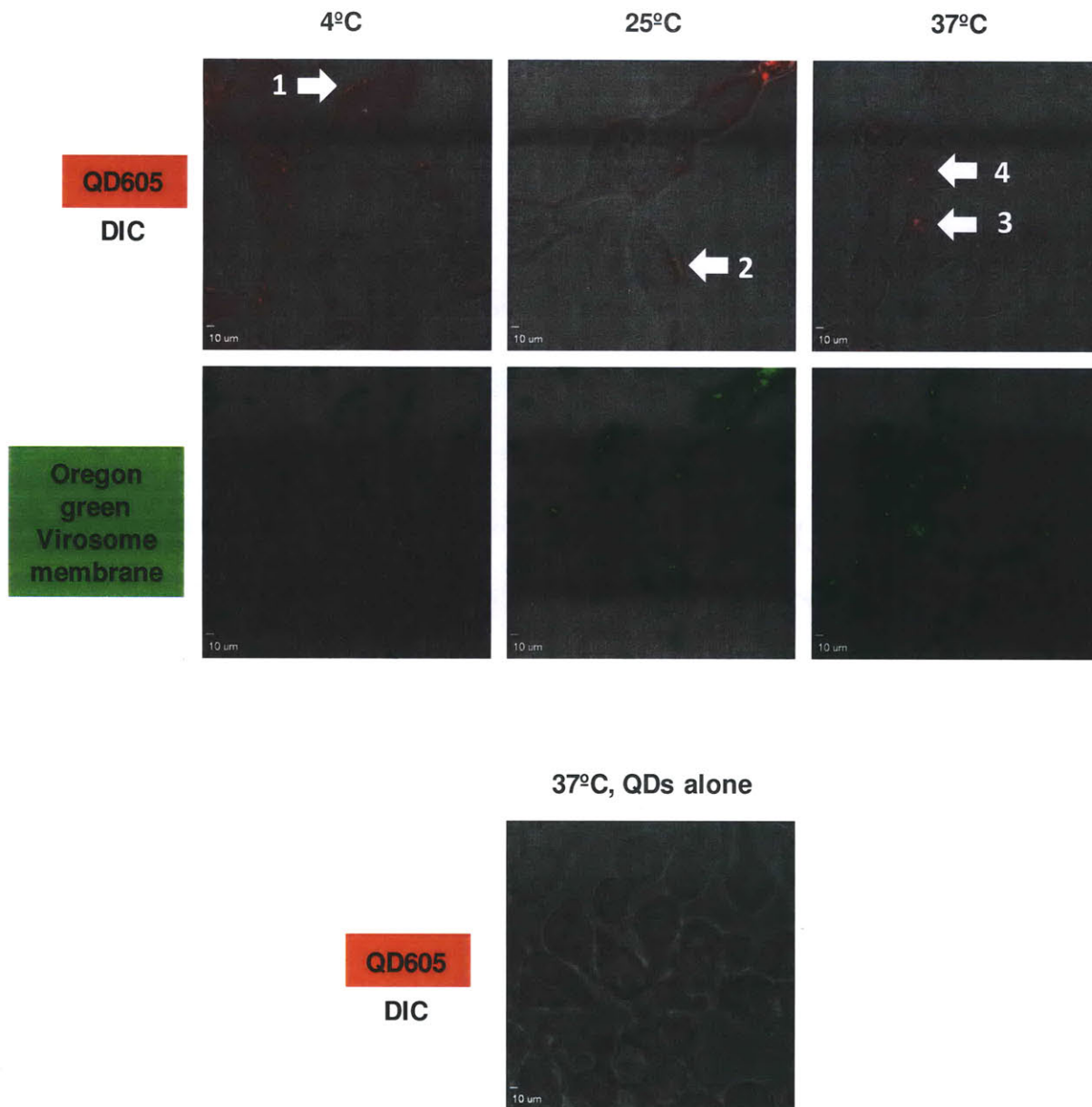


Figure 3-5. Internalization of QD-containing virosomes into HEK cells. The QD channel is shown in red superimposed on the DIC image, left. The virosome membrane dye channel is shown in green on the right. Top, QD-containing virosomes were allowed to bind to the surface of HEK cells on ice to prevent internalization. Imaging reveals primarily cell-surface-associated virosomes, as indicated by arrow 1. Second row, virosomes were bound to the surface of HEK cells, then incubated at 25°C for five minutes prior to imaging. Virosomes remaining bound to the cell surface are indicated by arrow 2. Bottom row, virosomes were bound to the surface of HEK cells, then incubated at 37°C for five minutes. Internalized virosomes are indicated with arrows 3 and 4.

We then sought to determine whether these internalized quantum dots are freely distributed in the cytosol, as desired, or if they are trapped in endocytic vesicles, at sufficiently long times after internalization. In order to determine QD localization, we needed a *bona fide* marker of endocytic vesicles. For this purpose we utilized FM 1-43, a styryl dye that emits green fluorescence.[24] FM 1-43 can intercalate into the plasma membrane, then be internalized into endosomal membranes; subsequent wash-out of excess dye from the plasma membrane allows selective visualization of endocytic vesicles.[25] Treatment of our HEK cells with FM 1-43 under this protocol generates perinuclear fluorescence, as expected for a population of late endosomes (Figure 3-6). We then allowed virosomes to bind to HEK cells on ice, washed away unbound material, and allowed internalization to occur for a longer time, 30 minutes at 37°C. (As compared to the previous experiment, this longer internalization time not only allows for endocytosis, but also for the virosomes to get to their final destination inside the cell.) The virosomes exhibited a clear perinuclear localization (Figure 3-6), identical to our late endosomal marker. Again, the green fluorescence of the virosome membrane dye is visible as internalized puncta, which we interpret as an indication that membrane fusion and subsequent dye dilution has not occurred. While it is possible that individual virosomes are fusion-active and release a small proportion of the total population of QDs into the cytosol, it is clear that the vast majority remain localized in endosomes.

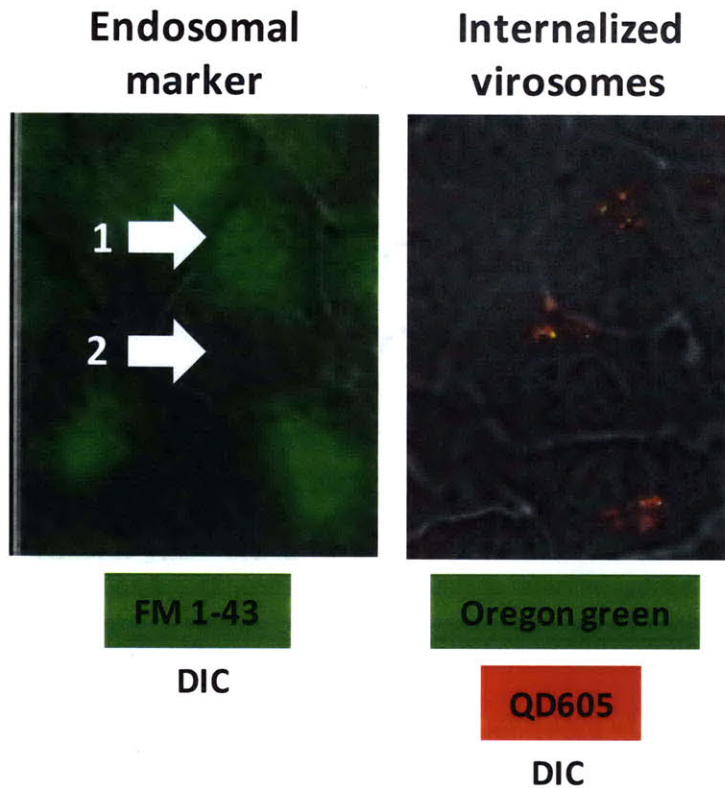


Figure 3-6. Comparison of internalized virosome localization to a marker of late endosomes reveals endosomal trapping. Left, HEK cells were labeled with FM1-43 to visualize late endosomes, which display a punctuate, perinuclear localization. The perinuclear FM1-43 fluorescence is indicated by white arrow 1, and the nucleus of that cell is indicated with white arrow 2. FM1-43 fluorescence is shown in green superimposed on the DIC image. Right, HEK cells were allowed to endocytose cell surface-bound QD-containing virosomes, which are restricted to a perinuclear localization, indicating that they are trapped in late endosomes. QD (red) and virosome membrane dye (green) fluorescence are superimposed on the DIC image.

We therefore conclude that our functional assays on these virosomes indicate that they are not fusion-active during endocytosis, as we had hoped. Under artificial, forcing conditions, some fraction of our virosomes seem to be able to fuse with the plasma membrane, but this cannot be used as a general delivery method, as treating cells with cold acidic buffer is certainly not a step one would wish to employ prior to investigating biological phenomena. Furthermore, there is high batch-to-batch variability in this fusogenic activity, indicating that virosomes are difficult to prepare in active form and will not be a general tool for use by non-experts in QD delivery.

Further optimization of the virosome preparation could have been carried out, and would certainly have required additional structural characterization of the virosomes. However, at this time, personal communication with our collaborators in the Zhuang and van Oijen laboratories indicated that virosomes proved not to have significant fusogenic activity in their hands either, and furthermore that electron microscopy had revealed the results of this preparation not to be well-formed lipid bilayers, but rather larger aggregates of hydrophobic membrane substituents (Dan Floyd, personal communication). Formation of aggregates rather than lipid bilayers has been reported for virosomes reconstituted by a similar protocol from Sendai virus.[26] Furthermore our imaging data are indicative of aggregates, as most apparent virosomes are larger than diffraction-limited spots. It is likely that the variation between batches of virosomes depends on the fraction of well-formed virosomes formed relative to the amount of inactive aggregates. The physical source of this variability is almost certainly variation in sorbent loading and quality between preparations. We note that even the best reported data in the literature indicate that only 25% of virosomes prepared by this method are fusion-active,[27] and further note that influenza virosomes have not become a general tool for intracellular delivery, despite having been first reported in 1987; unfortunately, our results go a long way toward explaining why this is the case.

We therefore abandoned this influenza virosome formulation rather than attempting to further optimize it. However, viral delivery may yet prove useful for delivering imaging reagents into cultured cells and tissues. A “slow reconstitution” influenza virosome preparation protocol, in which a short-chain phospholipid detergent is used for solubilization, then removed by dialysis rather than sorbent beads, has recently been reported.[23] This method reportedly generates

better-quality and more reproducible virosomes with up to 75% fusion efficiency.[23] Shortly after this initial report, these dialyzable virosomes were reported to efficiently encapsulate and deliver siRNA[20, 21] and plasmid DNA into cells.[19] The elimination of the sorbent-dependent step could conceivably reduce the batch-to-batch variability and aggregate formation plaguing the previous formulation. Therefore, this improved influenza virosome preparation may prove more useful for the delivery of imaging reagents. However, if this slow-reconstitution influenza virosome preparation is investigated in the future, we suggest that more complete characterization of QD-associated virosomes be performed prior to cellular experiments; in particular, electron microscopy should be performed prior to any functional assays in order to verify that virosomes consist of well-formed lipid bilayers rather than hydrophobic aggregates of membrane components, and furthermore that QDs are actually encapsulated.

Additionally, the non-enveloped adenovirus has been extensively utilized for gene delivery due to its high efficiency of cell entry. Rather than relying on encapsulation of cargo, adenovirus causes endosomal rupture, allowing any material, including fluorescently labeled high molecular weight dextrans, that is co-endocytosed with the virus to escape into the cytosol.[28-30] Despite the fact that it is no longer utilized in clinical applications due to its immunogenicity, we anticipate that adenovirus could be developed into an efficient tool for delivery imaging reagents into cultured cells for microscopy studies. Because convincing imaging data for adenovirus-mediated endosomal lysis exists, this seems to be a more promising avenue for future investigation than influenza virosomes.

These improved viral delivery methods may be investigated in the Ting lab in the future. However, we decided that more rigorous methods for determining the intracellular fate of delivered QDs must be validated prior to trying to develop new delivery methods. In particular, we needed a “positive control” of *bona fide* cytosolic QDs for comparison to our QDs internalized *via* any newly developed delivery methods. We therefore set aside viral delivery method development in order to validate a previously reported (but ill-characterized) method for QD delivery, reversible permeabilization with the bacterial toxin streptolysin O.

Part II: Streptolysin O-mediated delivery of quantum dots into the cytosol of living cells

Introduction

Streptolysin O (SLO) is a bacterial toxin that forms pores in biological membranes. The protein consists of a dagger-shaped monomer which intercalates into biological membranes by binding cholesterol.[31] After the initial binding event, additional monomers associate; oligomerization leads to formation of large circular pores up to 35 nm in diameter.[32] If a large number of pores are opened in a given cell, cell death inevitably occurs; however, if the number of pores is controlled (to two or fewer per cell), re-sealing of the lesions can occur after activation of a Ca^{2+} -dependent signaling pathway.[32, 33] This “reversible permeabilization” with SLO has been utilized for delivery of macromolecules into cells.[32] Cells subjected to reversible permeabilization followed by re-sealing have been reported to remain viable.[32, 34] It is important to note that SLO exhibits toxicity even at doses low enough to permit significant re-sealing; under conditions providing 70-80% of cells permeabilized, only two-thirds of the permeabilized cells survive.[32] Therefore, SLO affords cytosolic delivery, but a careful balance must be maintained between sufficient percentage of cells permeabilized and toxicity. Our proposed mechanism of delivery utilizing SLO is depicted in Figure 3-7.

The mechanism of SLO lesion repair after reversible permeabilization has been investigated. Upon pore formation, calcium from the extracellular medium enters the cell, and initiates calcium-dependent signaling that is responsible for initiating the downstream repair processes.[32, 33] This activates the transcription factor NF- κ B, which promotes expression of inflammatory cytokines.[33] Subsequent membrane lesion repair occurs in a microtubule-

dependent process.[32] Evidence has been found that, although the lesions have presumably been healed (because after re-sealing of SLO pores, cells no longer take up propidium iodide), the SLO monomer persists in the membrane of treated cells.[32, 33] This repair process only occurs if one to two lesions are formed per cell; if SLO is applied at high concentrations such that more pores are formed in each cell, lesion healing cannot occur and the cells die.[32] Therefore, the literature indicates that SLO can be used for reversible permeabilization of cells, and furthermore that SLO pores can be healed if they are formed in limited numbers. However, transcriptional activation and stress response signaling, as well as cytoskeletal remodeling, occur during the healing process, so the biological relevance of the state of the cell must be taken into consideration if the reagents delivered with SLO are intended to probe a normal physiological function.

SLO has been reported to be an effective method for imaging reagent delivery, including dye-conjugated RNA molecular beacons.[35] While SLO has been previously reported to deliver monodisperse QDs into the cytosol of living cells, limited imaging data has appeared in the literature.[8, 9] Importantly, no negative controls showing that the observed QD localization is due to SLO treatment have been reported. Furthermore, the QDs delivered by SLO have not been demonstrated in the literature to be able to label intracellular proteins. We wished to investigate SLO-mediated delivery in greater detail, and to determine if streptavidin-QDs and fluorophore-labeled streptavidin could be specifically targeted to biotinylated proteins inside the cell.

Extracellular medium

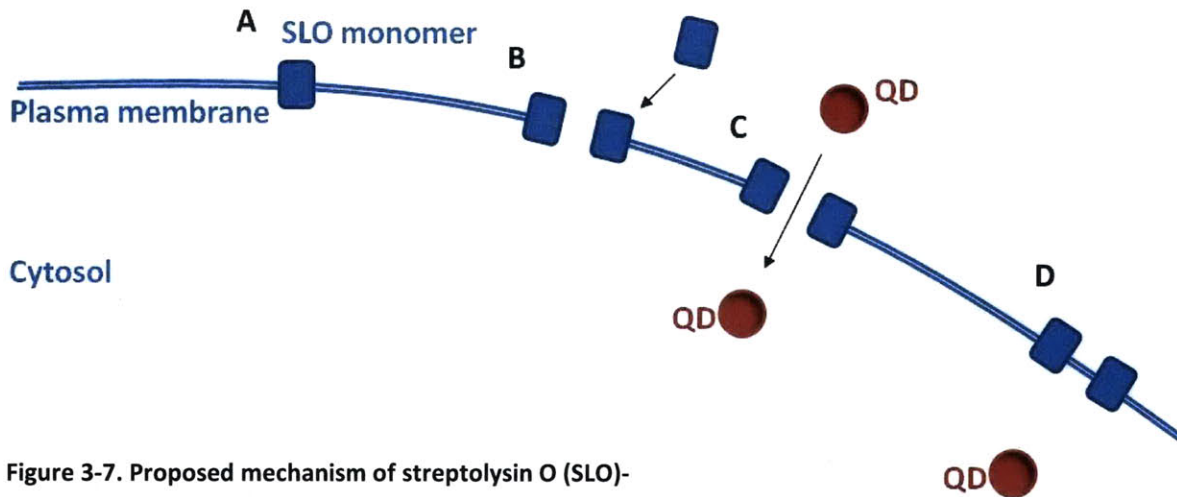


Figure 3-7. Proposed mechanism of streptolysin O (SLO)-mediated QD delivery. (A) An individual SLO monomer initially binds to cholesterol in the cell membrane. (B) After the initial binding event, additional SLO monomers associate, oligomerizing to form a pore of up to 35 nm diameter. (C) Once pores are formed, nanoparticulate cargo such as QDs can be added to the extracellular medium, and enter through the pores *via* passive diffusion. (D) If fewer than one or two pores are formed per cell, calcium-dependent signaling can activate active remodeling processes that repair SLO pores, trapping QDs inside the cytosol and allowing cell survival. This process of pore generation followed by re-sealing is referred to as reversible permeabilization.

Results and discussion

Delivering cell-impermeant cargo into the cytosol of living cells with streptolysin O

Streptolysin O had been previously reported to provide QD delivery into the cytosol of living cells[9], and generally to facilitate delivery of non-cell-permeable cargo *via* reversible membrane pore formation[32]. However, QD delivery by SLO was poorly characterized and no controls were done. We reasoned, therefore, that, should we demonstrate that SLO delivery

really does deliver QDs into the cytosol, not only would we then have in hand a positive control for QD delivery against which to assay any future methods we should develop, we would also have validated SLO as a method for QD delivery which could be used in the meantime, until better methods could be developed.

We first had to develop a general protocol for SLO delivery. We noted that the previously reported protocol called for delivery of QDs into SLO-permeabilized cells at 37°C for one hour.[9] Because we previously observed that incubating cells with QDs at this temperature readily leads to QD endocytosis (Mark Howarth, unpublished results), we reasoned that this protocol would result in a mixture of endocytically trapped and free QDs inside the cells. We therefore devised a modified SLO delivery protocol, wherein we treated cells with SLO to pre-form pores, then incubated the permeabilized cells at 4°C with the molecule we wished to deliver, allowing diffusion into cells while preventing endocytosis. Finally, rescue in calcium-containing buffer should permit re-sealing of the lesions. With this modified protocol in hand, we proceeded to query SLO-mediated delivery of various imaging reagents.

We also note that our initial experiments were conducted with commercially available SLO. This reagent contains surface cysteines, which after purification and storage form inter-molecular disulfides, inactivating the protein; therefore, prior to use in cell permeabilization, the protein must be “activated” by incubating it at 37°C with concentrated reductant solutions, such as dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine hydrochloride (TCEP).[9, 32, 35] Most reported protocols then apply this protein-reductant solution directly to cells for permeabilization, a condition we therefore also employed.[9] We note that treating cells with

reductant is non-physiological and obviously not ideal, and could contribute significantly to toxicity. An alanine mutant of SLO at this position has previously been reported,[32] which produces active protein without the disulfide formation problem. We did not at the time have access to this clone, though subsequent experiments conducted by Dr. Takashi Kawakami in our research group have made use of this mutant protein to great effect, as we will discuss later (vide infra).

Because of differences between cell lines and the delicate balance between SLO permeabilization and cell killing (requiring exactly one or two pores to be formed per cell),[32] it is recommended that cells be titrated with varying concentrations of SLO until the optimal balance between delivery and toxicity is achieved.[32] Additionally, SLO activities are reported in units; one unit is usually defined in terms of the amount of hemolysis (that is, lysis of suspended red blood cells) it causes over a certain time.[32] However, SLO from different sources is assayed for hemolytic activity under different conditions, so the unit definition does not necessarily translate from source to source.[32] Therefore, each batch of SLO should be titrated against the cells of interest to empirically determine the concentration of SLO that provides the desired amount of permeabilization activity empirically.[32]

We therefore began our experiments by determining efficacious, non-toxic SLO concentrations that would be required to deliver fluorescein, a small organic fluorophore that cannot cross cell membranes due to its negative charge. We performed this SLO titration on three common laboratory cell lines (HEK, HeLa, and COS7) and determined that concentrations of SLO between 50-100 units per mL afforded efficient fluorescein delivery (Figure 3-8). Importantly, a negative

control demonstrates that fluorescein uptake in HeLa cells is dependent on SLO treatment. Furthermore, after healing, no significant morphology change is evident in cells treated with SLO relative to cells left untreated, indicating that SLO toxicity is not acute under the conditions employed.

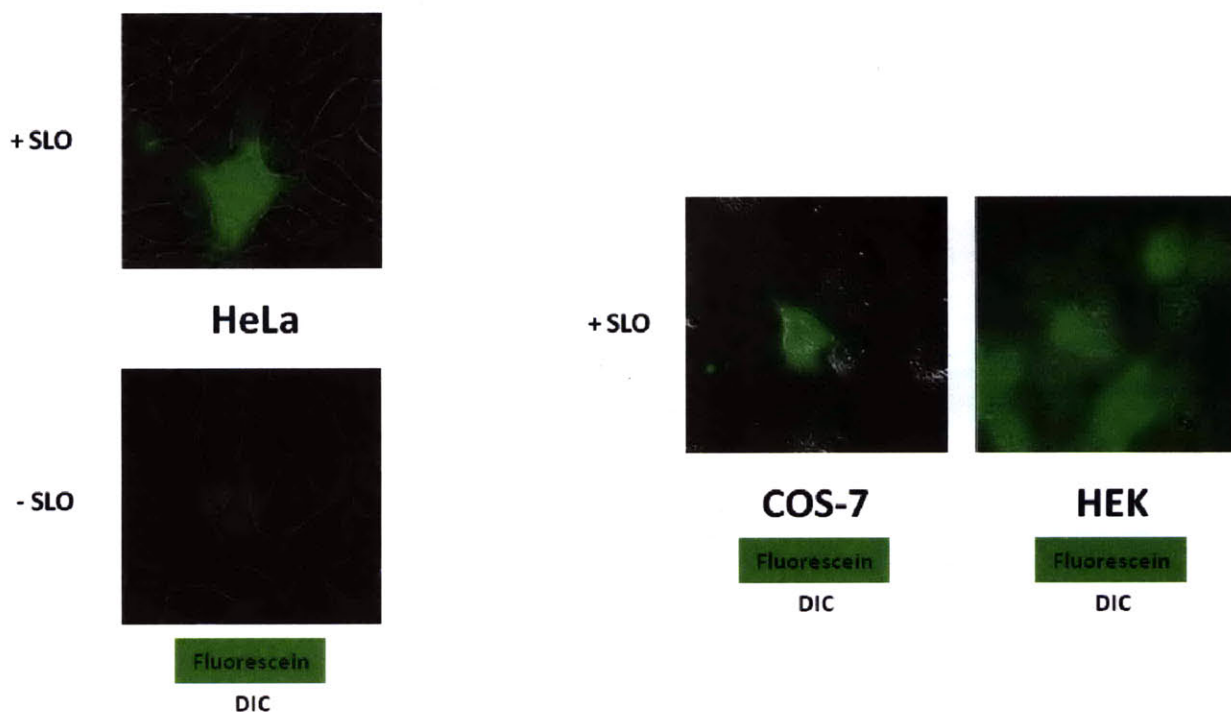


Figure 3-8. Delivering fluorescein into cell lines with streptolysin O. HeLa, COS-7, and HEK cells were permeabilized for 15 minutes with 50-100 units per mL of SLO, then fluorescein was delivered for 5 minutes at room temperature. Cells were imaged live. Uniform, diffuse fluorescein signal was observed in permeabilized cells. A negative control was performed for HeLa cells with SLO omitted to show that under normal conditions, fluorescein cannot enter cells. HeLa images are normalized on the same intensity scale. Fluorescein intensity (green) is superimposed on the DIC image for each.

It is interesting to note that our cell lines exhibit highly variable susceptibility to SLO permeabilization, with nearly 100% of HEK cells taking up fluorescein. It is possible that these cell lines have differential cholesterol content in their plasma membranes, or more generally

different membrane composition, causing differences in the ability of SLO monomers to intercalate into their membranes.

Next we attempted to deliver proteins into HEK cells with SLO. mCherry is a monomeric red fluorescent protein[36] that serves in this case as a model membrane-impermeant nanoparticle. We also attempted to deliver streptavidin conjugated to Alexa Fluor 568, a reagent commonly used for imaging biotinylated proteins.[37] Both were efficiently delivered into HEK cells after SLO permeabilization, and displayed uniform cytoplasmic localization (Figure 3-9). No evidence of punctuate or endocytic localization is evident.

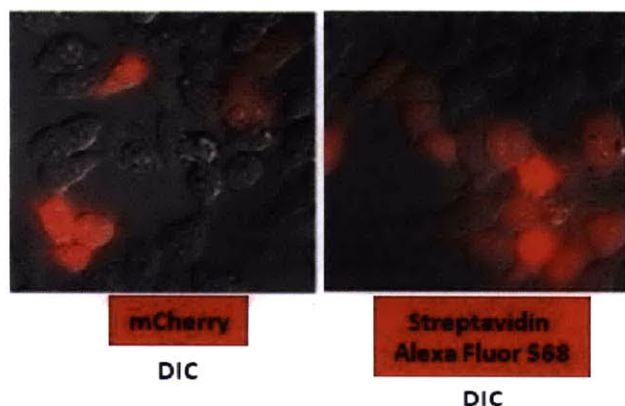


Figure 3-9. Delivering proteins into HEK cells with streptolysin O. mCherry, left, and streptavidin-Alexa Fluor 568, right, were delivered into HEK cells after SLO permeabilization. Diffuse cytosolic fluorescence is observed in each case. In each image, the protein fluorescence (red) is superimposed on the DIC image.

We then tested QD delivery with SLO, utilizing commercial red QDs with amine-functionalized ligands. These dots are large, approximately 20 nm, and commonly used for single-molecule imaging.[3] QD605 were efficiently delivered into HeLa cells after permeabilization with 20 units per mL of SLO, and in the absence of SLO, no QD internalization occurred (Figure 3-10).

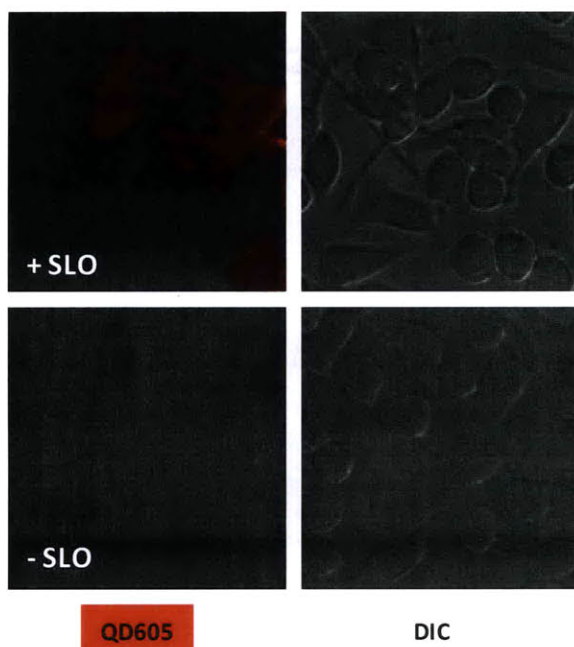


Figure 3-10. Delivering QDs into HeLa cells with streptolysin O. HeLa cells were permeabilized with SLO, then QD605 were delivered, top. Delivered QDs are generally disperse in the cytosol and appear to be individual molecules rather than larger aggregates. If SLO is omitted, QDs do not enter cells under these conditions, bottom. The QD channel (red) is shown on the left, and the DIC image is shown to the right in all images. HeLa images are normalized on the same intensity scale.

We therefore confirmed that SLO affords efficient cytosolic delivery of organic fluorophores, proteins, and QDs under conditions where endocytosis is not expected to occur, providing good evidence that these molecules should be free in the cytosol. The next step is to use them for labeling and single-molecule imaging of biotinylated proteins.

Targeting QDs to specifically biotinylated intracellular proteins

We then needed to demonstrate that these SLO-delivered QDs can be specifically targeted to intracellular proteins. We reasoned that we could utilize the biotin ligase (BirA)-mediated labeling method described in detail in Chapter 1 to specifically biotinylate intracellular proteins. Briefly, the protein of interest could be genetically fused to the acceptor peptide (AP) tag, and co-expressed inside cells with BirA. Treatment of the cells with biotin would then allow BirA to biotinylate the AP-tagged protein (since ATP is available inside cells). Subsequent SLO permeabilization and delivery of streptavidin-functionalized QDs should permit the QDs to bind

to the biotinylated protein. It has previously been demonstrated that BirA specifically biotinylates AP-tagged proteins (to the exclusion of mammalian proteins) in the context of mammalian cells by various methods, including mass spectrometry,[38] immunoblotting,[2, 39] and imaging.[39]

We reasoned that the easiest way to determine specific targeting of QDs to a protein of interest was to select model proteins with well-characterized intracellular localizations. We initially investigated two membrane-targeted proteins: AP-tagged cyan fluorescent protein (CFP) targeted to the plasma membrane by appending a farnesylation sequence; and the transmembrane receptor EphA3 bearing a C-terminal, intracellular AP tag (EphA3-AP). We confirmed that each of these proteins can be specifically biotinylated by BirA inside cells by immunoblotting and immunofluorescence, respectively (data not shown).

Unfortunately, in our hands, SLO-delivered streptavidin-Alexa Fluor conjugates were not observed to be targeted to these constructs in living HEK cells. This is likely to be due to the fact that, under the SLO-permeabilization conditions utilized, cells remained healthy and viable, but statistically few cells received streptavidin-Alexa Fluor as well as being transfected and biotin-labeled. More stringent permeabilization conditions could possibly have been employed while still maintaining cell health to give higher permeabilization efficiencies; however, in our hands, higher SLO concentrations were toxic (data not shown). Second, though we confirmed that biotinylation of EphA3 was occurring, we had no assay in hand to confirm that excess biotin was fully washed out of the cells prior to streptavidin delivery. Excess free biotin inside the cells and in the cell growth media would occupy the streptavidin binding site and prevent targeting to

the biotinylated protein inside the cell. Finally, SLO-mediated permeabilization is most efficient for cells that are not contact-inhibited, so more careful optimization of confluency could have improved our chances of observing targeting.

At this point the project was taken over by Dr. Takashi Kawakami. He performed independent optimization of the intracellular biotinylation and SLO delivery protocols (see Methods), and has subsequently demonstrated SLO-mediated targeting of streptavidin-functionalized QDs not only to EphA3-AP, but also to the cytoskeletal proteins actin and vimentin (Figures 3-11 and 3-12). Briefly, the AP-tagged protein is co-expressed with BirA, a short biotin labeling step is employed, and excess biotin is removed by washing. Cells are then permeabilized with the C530A mutant of SLO, which does not require activation with reducing agents, then streptavidin-QDs are delivered and cells are imaged immediately after the delivery step in buffer containing added calcium, which promotes lesion healing.

Dr. Kawakami has demonstrated that streptavidin-QDs to EphA3 and vimentin-AP reveals the expected intracellular localizations (Figure 3-11). QDs targeted to EphA3-AP are localized to the plasma membrane, as expected for this transmembrane protein, but are actually predominantly localized in intracellular puncta, presumably endocytic vesicles containing internalized EphA3-AP (Figure 3-11, top). This is not surprising, as QDs display many copies of streptavidin and therefore can cross-link biotinylated EphA3-AP molecules; this type of clustering has been previously demonstrated to promote EphA3 internalization.[1] Targeting streptavidin-QDs to vimentin, a cytoskeletal protein that forms intermediate filaments, reveals wide, twisting filaments (Figure 3-11, bottom). These experiments demonstrate that

intracellular biotinylation and QD targeting is specific, and more importantly, that SLO-delivered QDs are freely diffusible in the cytosol of living cells and can “find” their biotinylated target proteins.

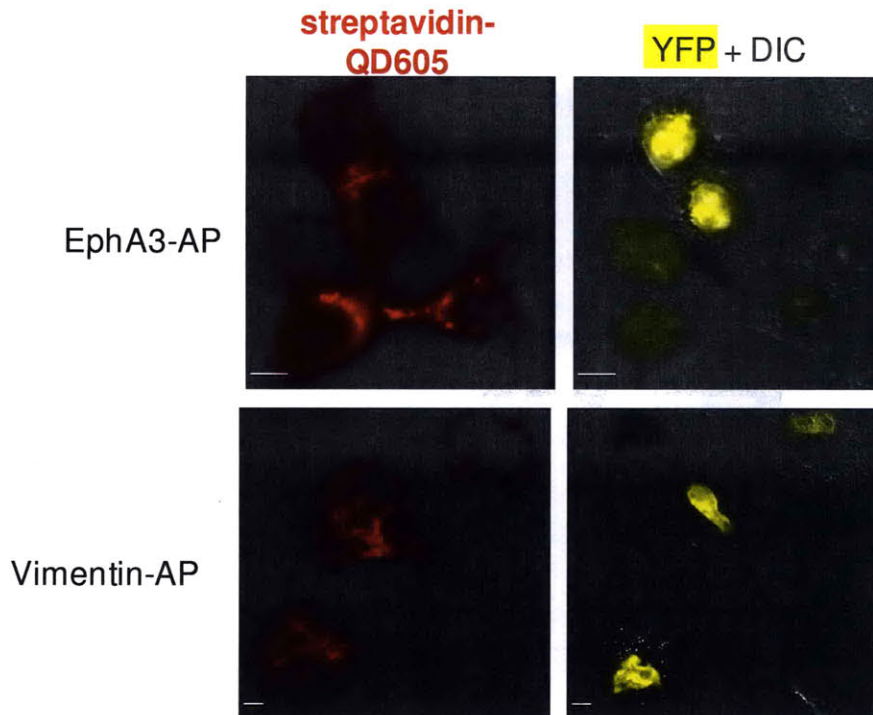


Figure 3-11. SLO-mediated targeting of QDs to EphA3-AP and vimentin-AP. Top, HEK cells were transfected with EphA3-AP and BirA, as well as a yellow nuclear transfection marker. After biotinylation and SLO permeabilization, streptavidin-functionalized QDs were targeted to this cell surface receptor. The resulting QD localization (shown in red, top left) is distributed between the cell surface and recycling endosomes, as expected. Bottom, COS7 cells were transfected with vimentin-AP, BirA, and the yellow nuclear transfection marker. After biotinylation and SLO-mediated QD delivery, the QD fluorescence (red, bottom left) delineates the large, twisted fibers characteristic of this cytoskeletal protein. The yellow transfection marker is superimposed on the DIC image on the right for both experiments. Scale bars, 10 μ m.

Finally, Dr. Kawakami performed labeling of AP-tagged actin, a cytoskeletal protein involved in microfilament formation, and also conducted a series of controls to demonstrate the specificity of QD targeting. SLO-delivered QDs can be targeted to biotinylated AP-actin in HEK cells (Figure 3-12A), revealing the thin filaments characteristic of actin localization. If biotin treatment is omitted (Figure 3-12B), BirA is not expressed in the cells (Figure 3-12C), or a mutation in the AP

that converts the specifically biotinylated lysine to alanine is introduced (Figure 3-12D), no QD labeling occurs; furthermore, this protocol fortuitously allows the delivered QDs to diffuse back out of the cells if they are not bound to an intracellular protein, rather than trapping the untargeted QDs inside the cells upon lesion healing, although this exciting result was not expected *a priori*.

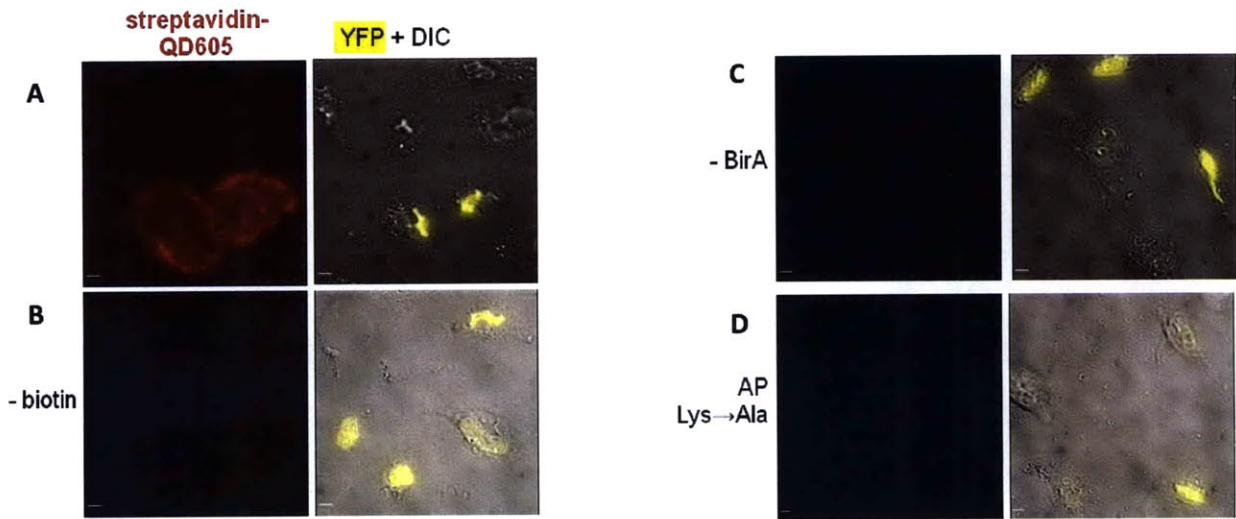


Figure 3-12. SLO-mediated targeting of QDs to actin and specificity controls. HeLa cells were transfected with AP-actin, BirA, and a yellow nuclear transfection marker. Subsequent to biotinylation and SLO-mediated permeabilization, streptavidin-QDs were delivered into the cells. Imaging of these cells (A) reveals QD fluorescence that is targeted to apparent actin filaments (shown in red, left) in transfected cells (shown by the yellow transfection marker superimposed on the DIC image, right). Negative controls reveal that this targeting is dependent on biotinylation of the target protein. Omitting biotin (B), BirA (C), or using a mutant of AP-actin where the specifically biotinylated lysine of the AP tag is mutated to alanine (D) abrogates QD targeting inside transfected cells. Scale bars, 10 μm .

The work of Dr. Kawakami has therefore successfully demonstrated that intracellular proteins can be specifically biotinylated, and that SLO-mediated delivery of streptavidin-QDs can afford QD labeling of these intracellular proteins. This validates SLO as a delivery method that can provide freely diffusible, targetable cytosolic QDs. Of course, SLO delivery can therefore be utilized for QD targeting in the near term. In the longer term, we do not intend SLO to be a

permanent solution to the delivery problem because of its limitations in terms of perturbing cell physiology and health (see Conclusion); we can, however, use SLO delivery to be a positive control for cytosolic QD delivery against which to evaluate any new QD delivery methods developed in the Ting lab.

Conclusion

We have established that SLO can efficiently deliver organic fluorophores, proteins, and QDs into living cells. Dr. Takashi Kawakami has subsequently demonstrated that SLO-delivered QDs can be utilized for protein labeling and imaging. SLO-mediated delivery may therefore prove to be an accessible near-term alternative to microinjection for intracellular QD targeting. This method could be accessible to more researchers and applicable to more problems than microinjection because the SLO reagent is relatively inexpensive and commercially available, permeabilizes many cells at a time rather than just a few, and requires no special training prior to use.

However, we note that this technology suffers serious problems that limit its applicability to real biological systems. First, as discussed in the introduction to this section, the balance between efficient permeabilization and toxicity with SLO is delicate[32], meaning that it probably cannot be applied to primary cells or living animals. Even in robust laboratory cell lines, SLO concentration must be carefully controlled. Second, the biological relevance of any observations made after SLO treatment may well be questionable, as the cell is undergoing massive cytoskeletal remodeling and membrane repair processes to repair the SLO lesions[33], which is certainly not a normal physiological condition. Third, the commercially available SLO

protein must be activated with reductants such as TCEP or DTT prior to use, due to inter-molecular disulfides formed between surface cysteines that inactivate the protein; treating cells with solutions containing reductant is, obviously, non-physiological and toxic. While the alanine mutant of SLO at this position can be prepared,[32] the mutant protein is not commercially available and must be recombinantly expressed and purified by the researcher; it is more likely that the non-expert researcher will elect to utilize the commercially available reagent. These limitations make it clear that SLO serves as a temporary solution until a more favorable delivery system is devised. More importantly, though, SLO provides us a benchmark against which to evaluate QD delivery methods we develop in the future. It provides us a genuine sample of QDs delivered into the cytosol of living cells, against which we can compare the localization, fate, and diffusion properties of QDs delivered by new methods.

Methods

Mammalian cell culture

HEK, HeLa, CHO A7 and COS-7 cells were cultured in growth media, consisting of Dulbecco's modification of Eagle's medium (DMEM, Cellgro) supplemented with 10% fetal bovine serum (FBS, PAA Laboratories), 50 units/mL penicillin, and 50 $\mu\text{g}/\text{mL}$ streptomycin (Cellgro). Cells were maintained at 37°C under an atmosphere of 5% CO_2 . For cellular imaging, cells were grown on glass coverslips. HeLa, CHO A7 and COS-7 cells were grown directly on the glass substrate. HEK cells were grown on glass pre-treated with 50 $\mu\text{g}/\text{ml}$ fibronectin (Millipore).

Fluorescence imaging

Cells were imaged in Dulbecco's phosphate buffered saline (DPBS) on glass coverslips. A Zeiss AxioObserver inverted microscope with a 40x or 63x oil-immersion objective was used for epifluorescence imaging. QD605 (405 broad excitation, 585 dichroic, 605/30 emission), Fluorescein/Oregon Green/FM1-43 (493/16 excitation, 506 dichroic, 525/30 emission), mCherry/Alexa Fluor 568 (570/20 excitation, 585 dichroic, 605/30 emission), and differential interference contrast (DIC) images were collected and analyzed with SlideBook software (Intelligent Imaging Innovations). Acquisition times ranged from 20 milliseconds to 5 seconds.

Virosome preparation

Purified H3N2 influenza virus (A, X31, Aichi/68) was obtained from Charles River Labs (1 mL of 2 mg/mL virus). The virus was washed by adding 4.5 mL of "Hepes 145 buffer" (consisting of 50 mM Hepes pH 7.4, 145 mM NaCl, 0.1 mM EDTA) to 500 μ L virus suspension, then virus was harvested by ultracentrifugation at 35,000 rpm and 4°C for 90 minutes. Fluorescein isothiocyanate-phosphatidylethanolamine (FITC-PE) in methylene chloride (3.1 μ L, Avanti Polar Lipids) was dried under nitrogen and added to 1 mL 100 mM octaethylene glycol monododecyl ether (C12E8, Anatrace) in Hepes 145 buffer; alternatively, 40 μ L of 2 mg/mL ethanolic solution of Oregon Green stearic acid, a kind gift of Dan Floyd and Antoine van Oijen, was added to 1 mL of 100 mM C12E8 in Hepes 145. These surfactant-dye mixtures were used to resuspend the viral pellet, followed by virion solubilization with sonification at room temperature for 20 minutes. The viral genome was then pelleted by ultracentrifugation at 43,000 rpm and 4°C for 40 minutes and discarded. The supernatant was reserved and Qdot ITK amino (PEG) (Invitrogen) was added to a final concentration of 50 nM. BioBeads SM2 (BioRad) were washed with Hepes 145 buffer and 30 mg wet beads were added to the supernatant-QD mixture. The

reconstitution mixture was vortexed at 2,000 rpm for 1 hour at 4°C for detergent removal. The reconstituted sample was layered on a sucrose step gradient of 5% to 40%, then separated by ultracentrifugation at 43,000 rpm and 4°C for 60 minutes. Reconstituted virosomes were recovered by pipetting from the visible interface between the sucrose layers.

To assay virosome formation and QD encapsulation, virosomes were diluted 1000 times in phosphate buffered saline (PBS) and imaged in epifluorescence mode with a 40x objective in the fluorescein and QD605 channels.

Surface acidification assay

Purified virosomes were diluted at a 1:25 ratio in 250 µL of DPBS-Mg and incubated on 90% confluent HEK cells at 4°C for 5 minutes. Cells were rinsed 4 times with ice-cold DPBS to remove unbound virosomes. Cells were then either incubated in ice-cold DPBS, pH 5.5, for 5 minutes on ice, or ice-cold DPBS, pH 7.5, for 5 minutes on ice as a negative control. Cells were exchanged back into ice-cold DPBS-Mg, pH 7.5, for imaging.

Virosome binding and internalization assays

Purified virosomes were diluted at a 1:25 ratio in 250 µL of DPBS-Mg (Dulbecco's modification of phosphate buffered saline with 1 mM MgCl₂ added), and incubated on 90% confluent HEK cells at 4°C for 5 minutes. (This dilution ratio was empirically determined to produce cell-surface virosome binding.) Cells were rinsed 4 times with ice-cold DPBS to remove unbound virosomes. Cells were then incubated in pre-temperature-equilibrated DPBS-Mg at 4°C, room

temperature (approximately 25°C), or 37°C for 5 minutes. Imaging was performed at room temperature in DPBS-Mg.

Visualizing late endosomes with FM1-43

Dye loading was performed by incubating 90% confluent HEK cells with 50 µg/mL FM1-43 in DPBS-Mg for 20 minutes at 37°C. The cells were then incubated in DMEM for 10 minutes at 37°C to remove excess dye from the plasma membrane prior to imaging.

SLO permeabilization and delivery

Streptolysin O (Sigma) was diluted in DPBS to the desired concentration, then activated by adding 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and incubating at 37°C for 30 minutes.[35] For permeabilization, 80% confluent cells (HeLa, HEK, COS-7 or CHO A7) were treated with 250 µL per 95 cm² of 20-100 units per mL SLO in DPBS for 15 minutes at 37°C. Cells were then rinsed 4 times with DPBS.

For fluorescein delivery, 10 µM fluorescein in DPBS was incubated with the permeabilized cells for 5 minutes at room temperature. Cells were rinsed three times with DPBS then immediately imaged.

For delivery of mCherry and streptavidin-Alexa Fluor conjugates, 24 µM mCherry protein or 10 µM streptavidin-Alexa Fluor 568 in DPBS-Mg were incubated with permeabilized cells for 60 minutes at 4°C. Cells were rinsed with DPBS 4 times and imaged.

For delivery of QD605, 10 nM QD605 in DPBS-Mg was incubated with permeabilized cells for 60 minutes at 4°C. Cells were rinsed with DPBS 4 times and imaged.

Unsuccessful QD targeting protocol

HEK cells were co-transfected with 400 ng 3YF-EphA3-AP (a kind gift of Martin Lackmann at Monash University) along with 25 ng of mCherry-BirA (pcDNA3) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions, then debiotinylated overnight (12-16 hours) with 50 µg/mL streptavidin (w/v in the growth media). Labeling with 50 µM biotin was performed for 60 minutes, then excess biotin was allowed to wash out for 60 minutes in DPBS-Mg. SLO was then added in DPBS-Mg to a final concentration of 100 U/mL and incubated on the cells for 15 minutes at 37°C. After washing off excess SLO, streptavidin-Alexa fluor 647 was diluted 1:25 in DPBS-Mg and incubated on the cells at 32°C for 60 minutes. Cells were then placed in DMEM (no serum) and incubated at 37°C for 60 minutes to allow lesion healing, then imaged.

Successful QD targeting protocol

HeLa, HEK, or COS7 cells were co-transfected with BirA (pcDNA3), the AP fusion protein (EphA3-AP, AP-actin, or AP-vimentin), and H2B-YFP (a nuclear-localized co-transfection marker) using Lipofectamine 2000 according to the manufacturer's instructions. No debiotinylation was carried out. The following day, biotinylation was carried out by incubating the cells with 10 µM biotin in growth media at 37°C for 10 minutes. Excess biotin was washed out with 3 applications of fresh growth media for 10 minutes each. Cells were then permeabilized with 0.1 mg/mL SLO (C530A mutant, prepared as previously described[32]) in cold DPBS for 3 minutes, then cells were rinsed with DPBS to remove excess SLO. Delivery of QDs (Qdot 605 streptavidin conjugate, Invitrogen) was accomplished by incubating the permeabilized cells with 10 nM QD in cold

DPBS with 5 mM MgCl₂ and 1 mM EGTA for 20 minutes. Cells were washed with DPBS with 1 mM CaCl₂ added (DPBS-Ca, which promotes lesion healing), then imaged in DPBS-Ca.

References

1. Howarth, M., et al., *Monovalent, reduced-size quantum dots for imaging receptors on living cells*. Nat Methods, 2008. **5**(5): p. 397-9.
2. Howarth, M., et al., *Targeting quantum dots to surface proteins in living cells with biotin ligase*. Proc.Natl.Acad.Sci.U.S.A, 2005. **102**(21): p. 7583-7588.
3. Liu, W., et al., *Compact biocompatible quantum dots functionalized for cellular imaging*. J Am Chem Soc, 2008. **130**(4): p. 1274-84.
4. Zamir, E., et al., *Fluorescence fluctuations of quantum-dot sensors capture intracellular protein interaction dynamics*. Nat Methods, 2010. **7**(4): p. 295-8.
5. Janes, P.W., et al., *Cytoplasmic Relaxation of Active Eph Controls Ephrin Shedding by ADAM10*. Plos Biology, 2009. **7**(10): p. -.
6. Derfus, A.M., W.C.W. Chan, and S.N. Bhatia, *Intracellular delivery of quantum dots for live cell labeling and organelle tracking*. Advanced Materials, 2004. **16**(12): p. 961-+.
7. Courty, S., et al., *Tracking individual kinesin motors in living cells using single quantum-dot imaging*. Nano Letters, 2006. **6**(7): p. 1491-1495.
8. Xing, Y., et al., *Molecular profiling of single cancer cells and clinical tissue specimens with semiconductor quantum dots*. Int J Nanomedicine, 2006. **1**(4): p. 473-81.
9. Agrawal, A. and S.M. Nie, *Real-time detection of single quantum dots inside living cells*. Nanobiophotonics and Biomedical Applications II, 2005. **5705**: p. 152-158
- 294.
10. Harrison, S.C., *Viral membrane fusion*. Nat Struct Mol Biol, 2008. **15**(7): p. 690-8.
11. Stegmann, T., et al., *Functional Reconstitution of Influenza-Virus Envelopes*. Embo Journal, 1987. **6**(9): p. 2651-2659.
12. Gamblin, S.J. and J.J. Skehel, *Influenza haemagglutinin and neuraminidase membrane glycoproteins*. J Biol Chem, 2010.
13. Lakadamyali, M., et al., *Visualizing infection of individual influenza viruses*. Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9280-5.
14. Bullough, P.A., et al., *Structure of influenza haemagglutinin at the pH of membrane fusion*. Nature, 1994. **371**(6492): p. 37-43.
15. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution*. Nature, 1981. **289**(5796): p. 366-73.
16. Stegmann, T., et al., *The HA2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion*. J Biol Chem, 1991. **266**(27): p. 18404-10.
17. Kemble, G.W., T. Danieli, and J.M. White, *Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion*. Cell, 1994. **76**(2): p. 383-91.
18. Kaneda, Y., *Virosomes: evolution of the liposome as a targeted drug delivery system*. Adv Drug Deliv Rev, 2000. **43**(2-3): p. 197-205.
19. de Jonge, J., et al., *Cellular gene transfer mediated by influenza virosomes with encapsulated plasmid DNA*. Biochem J, 2007. **405**(1): p. 41-9.
20. Huckriede, A., et al., *Cellular delivery of siRNA mediated by fusion-active virosomes*. J Liposome Res, 2007. **17**(1): p. 39-47.

21. de Jonge, J., et al., *Reconstituted influenza virus envelopes as an efficient carrier system for cellular delivery of small-interfering RNAs*. *Gene Ther*, 2006. **13**(5): p. 400-11.
22. Bron, R., A. Ortiz, and J. Wilschut, *Cellular cytoplasmic delivery of a polypeptide toxin by reconstituted influenza virus envelopes (virosomes)*. *Biochemistry*, 1994. **33**(31): p. 9110-7.
23. de Jonge, J., et al., *Use of a dialyzable short-chain phospholipid for efficient solubilization and reconstitution of influenza virus envelopes*. *Biochim Biophys Acta*, 2006. **1758**(4): p. 527-36.
24. Gaffield, M.A. and W.J. Betz, *Imaging synaptic vesicle exocytosis and endocytosis with FM dyes*. *Nat Protoc*, 2006. **1**(6): p. 2916-21.
25. Bertrand, C.A., et al., *Methods for detecting internalized, FM 1-43 stained particles in epithelial cells and monolayers*. *Biophys J*, 2006. **91**(10): p. 3872-83.
26. Bagai, S. and D.P. Sarkar, *Fusion-mediated microinjection of lysozyme into HepG2 cells through hemagglutinin neuraminidase-depleted Sendai virus envelopes*. *J Biol Chem*, 1994. **269**(3): p. 1966-72.
27. Stegmann, T., et al., *Functional reconstitution of influenza virus envelopes*. *Embo Journal*, 1987. **6**(9): p. 2651-9.
28. Brabec, M., et al., *Opening of size-selective pores in endosomes during human rhinovirus serotype 2 in vivo uncoating monitored by single-organelle flow analysis*. *Journal of Virology*, 2005. **79**(2): p. 1008-16.
29. FitzGerald, D.J., et al., *Adenovirus-induced release of epidermal growth factor and pseudomonas toxin into the cytosol of KB cells during receptor-mediated endocytosis*. *Cell*, 1983. **32**(2): p. 607-17.
30. Wiethoff, C.M., et al., *Adenovirus protein VI mediates membrane disruption following capsid disassembly*. *Journal of Virology*, 2005. **79**(4): p. 1992-2000.
31. Gilbert, R.J., *Pore-forming toxins*. *Cell Mol Life Sci*, 2002. **59**(5): p. 832-44.
32. Walev, I., et al., *Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O*. *Proc Natl Acad Sci U S A*, 2001. **98**(6): p. 3185-90.
33. Walev, I., et al., *Resealing of large transmembrane pores produced by streptolysin O in nucleated cells is accompanied by NF-kappaB activation and downstream events*. *FASEB J*, 2002. **16**(2): p. 237-9.
34. Fawcett, J.M., S.M. Harrison, and C.H. Orchard, *A method for reversible permeabilization of isolated rat ventricular myocytes*. *Exp Physiol*, 1998. **83**(3): p. 293-303.
35. Santangelo, P.J., et al., *Dual FRET molecular beacons for mRNA detection in living cells*. *Nucleic Acids Res*, 2004. **32**(6): p. e57.
36. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein*. *Nat Biotechnol*, 2004. **22**(12): p. 1567-72.
37. Howarth, M., et al., *A monovalent streptavidin with a single femtomolar biotin binding site*. *Nat. Methods*, 2006. **3**(4): p. 267-273.
38. Grosveld, F., et al., *Isolation and characterization of hematopoietic transcription factor complexes by in vivo biotinylation tagging and mass spectrometry*. *Ann.N.Y.Acad.Sci.*, 2005. **1054**: p. 55-67.
39. Fernandez-Suarez, M., T.S. Chen, and A.Y. Ting, *Protein-protein interaction detection in vitro and in cells by proximity biotinylation*. *J. Am. Chem. Soc.*, 2008. **130**(29): p. 9251-9253.

Chapter 4: Imaging intracellular protein-protein interactions with proximity lipoylation and proximity coumarin ligation

Dr. Justin Cohen performed kinetic analysis of coumarin ligation to the LAP1 peptide. Dr. Amar Thyagarajan assisted with the development of the neuron labeling protocol and performed neuron labeling and imaging.

Introduction

The behavior of proteins inside cells is governed by their interactions with other biomolecules, and protein-protein interactions in particular underlie many signaling pathways. Many methods of detecting protein-protein interactions (PPIs) in the context of the cell have been reported, and have been powerful in both the focused study of interactions of interest as well as profiling whole interactomes, as discussed in detail in Chapter 1. However, each method has its own set of strengths and weaknesses. Biochemical methods, such as co-immunoprecipitation, and yeast genetic methods, such as the yeast two-hybrid, have been extensively applied in the identification of new PPIs, but cannot provide spatial and temporal information about PPIs in their endogenous context. Protein complementation assays (PCAs), such as bimolecular fluorescence complementation (BiFC), have proved powerful for this purpose, but still have significant limitations, such as trapping the interaction complex. We therefore set out to develop improved methods for imaging PPIs in mammalian cells.

Similar to PCAs, where each half of a split reporter is fused to interacting proteins, we can create a two-component reporter comprised of a ligase enzyme and its substrate peptide. We design the reporter such that labeling of the peptide by the enzyme only occurs in the case of an interaction. We call this general concept proximity ligation. Figure 4-1 provides a schematic of the general proximity ligation concept, and we use this figure to explain the ideal kinetic parameters of a proximity ligation system that will produce the greatest dynamic range for PPI detection. For two (inducibly or transiently) interacting proteins A and B, protein A is fused to the enzyme and protein B is fused to the peptide. If A and B do not interact, no labeling should

occur. In the case that A and B do interact, the enzyme and peptide are brought into proximity and labeling can occur. Maximal dynamic range will occur when background labeling in the absence of an interaction is zero or minimal, and labeling in the case of an interaction is maximal. We accomplish this by treating the PPI as a “kinetic switch”.

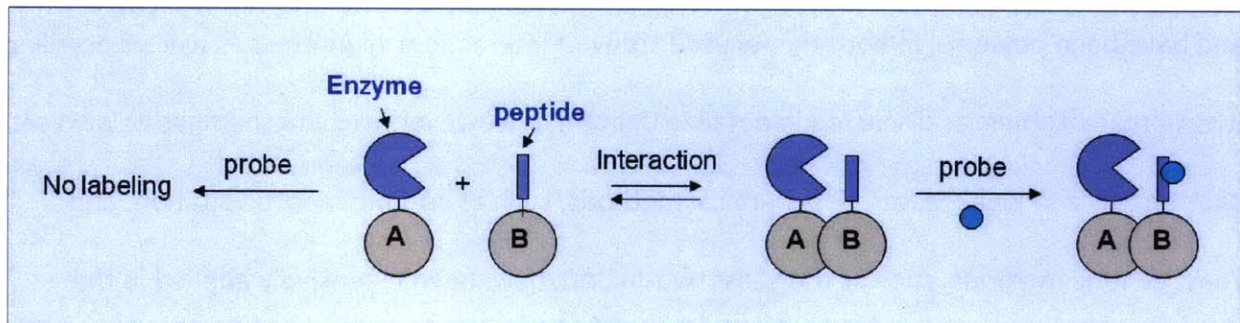


Figure 4-1. Schematic of proximity ligation. A and B are proteins of interest. A is fused to the enzyme ligase and B is fused to the peptide substrate. Probe ligation occurs only if A and B interact. Probe is represented as a blue circle.

When we say kinetic switch, we mean that we can engineer a system that kinetically discriminates between the cases when A and B are interacting and when they are not interacting, based on the change in local concentration created upon interaction. This can be accomplished by specifying the appropriate Michaelis-Menten parameters for enzymatic probe ligation to the peptide substrate. We briefly review the Michaelis-Menten treatment of steady-state enzyme kinetics and the relevant measurable kinetic parameters. For our steady-state system, K_m is empirically the concentration of peptide at which the rate of enzymatic ligation is half-maximal (assuming all other small-molecule substrates, i.e., probe and ATP, are provided at saturating concentrations). The maximal rate of enzymatic probe ligation when the peptide is provided at saturating concentrations is described by k_{cat} . When the enzyme and peptide are

present in the reaction at low concentrations, the rate of the overall reaction is controlled by the bimolecular rate constant k_{cat}/K_m and similarly depends on the concentration of the peptide. When the enzyme and peptide are present at high concentrations, and the peptide concentration is saturating, the rate of the pseudo-zero-order reaction is governed by k_{cat} .

Let us now consider how the occurrence or absence of an interaction can provide a “kinetic switch” for the rate of enzymatic probe ligation to the peptide. If enzyme and peptide are fused to proteins A and B and no interaction occurs, and if both enzyme and peptide are present in the cell at concentrations well below K_m , the reaction rate is bimolecular and depends on k_{cat}/K_m as well as the concentration of enzyme and peptide. If K_m for the enzyme and peptide is sufficiently high, that is, much higher than the concentrations of these constructs that can be achieved in the cell under most experimental conditions, then probe ligation under these conditions (that is, no interaction) will be kinetically inefficient. In fact, with sufficiently low protein concentrations and/or sufficiently high K_m , undetectable background labeling can theoretically be accomplished because the bimolecular rate will be extremely slow. Now consider the case of a PPI. If A and B do interact, enzyme and peptide are forcibly brought into very high local concentration, providing the previously mentioned “kinetic switch” that creates pseudo-zero-order conditions; under this condition, the rate of the reaction will be governed exclusively by the value of k_{cat} . Given a sufficiently high value for k_{cat} , therefore, the rate of probe ligation during the interaction will be maximally high. It is intuitive that, for labeling to occur during the time of interaction, k_{cat} must be fast enough that one enzymatic turnover can occur during two half-lives of the PPI. The faster the k_{cat} achievable for the enzyme and peptide, the better sensitivity of labeling that can be accomplished during a given labeling time, and the

more general the method will be, especially for transient PPIs. Therefore, in order to achieve maximal dynamic range for our kinetic switch reporter, K_m of the enzyme and peptide must be as high as possible, which maximizes the range of enzyme and peptide concentrations under which the bimolecular reaction is very slow (and, optimally, undetectable), providing low background labeling in the absence of an interaction; simultaneously, k_{cat} of probe ligation to the peptide should be maximally high in order to provide a fast ligation rate and therefore high sensitivity when an interaction occurs. The occurrence of a PPI therefore provides this kinetic switching by forcibly shifting the rate of enzymatic probe ligation to the peptide from a slow bimolecular rate to a fast pseudo-zero-order rate by providing forced proximity, or high local concentration. We set out to create a reporter of maximal dynamic range, therefore, by engineering K_m to be as high as possible, and k_{cat} to be as high as possible.

Our research group has previously reported a PPI sensor based on the proximity ligation concept called proximity biotinylation.[1] Proximity biotinylation utilizes the enzyme *E. coli* biotin ligase, BirA, and a truncation of its acceptor peptide, the AP(-3), as two halves of a reporter which are fused to interacting proteins. Interaction promotes biotinylation of the AP(-3) by BirA, which is subsequently detected with streptavidin. Design of the AP(-3) was a first attempt to access the desirable kinetic parameters we just described. The K_m of BirA for the original AP is 25 μM ,[2] and for the engineered AP(-3) is 345 μM ,[1] which provides for low background in the absence of forced proximity (that is, in the concentration regime governed by the bimolecular rate constant). However, the k_{cat} for this reaction is perturbed. The k_{cat} for biotinylation of AP by BirA is 12 min^{-1} ,[2] while the k_{cat} for biotinylation of AP(-3) is 0.53 min^{-1} ,[1] which severely curtails the sensitivity of the method. We reasoned that we could expand

the dynamic range of detection if we could find an enzyme-peptide pair having increased k_{cat} , but similarly high (or higher) K_m . A second limitation to the method is that proximity biotinylation is restricted to the surface of living cells or the cytoplasm of fixed cells due to the requirement for streptavidin-based detection.

We therefore sought to improve on the dynamic range of proximity biotinylation, and also to extend the applicability of enzyme-mediated proximity labeling to the cytoplasm of living cells. Because of our extensive experience with enzyme ligases and their peptide substrates, we chose to build on the general proximity ligation paradigm for the design of our system. Unfortunately we cannot currently address both of these issues with one reporter design, but we will describe two methods that attempt to address first improved dynamic range, then live-cell imaging.

We present this chapter in three parts. Part I describes our efforts to improve on dynamic range of PPI detection in fixed cells and lysates using the enzyme *E. coli* lipoic acid ligase, or LplA, and its ligation of its natural substrate, lipoic acid, to an engineered acceptor peptide called LAP1. We then describe our design and validation of a reporter for imaging PPIs in living cells using a mutant of LplA that can attach a blue fluorescent coumarin probe to the LAP1 peptide. Part II will delineate our application of proximity coumarin ligation to investigate the interaction of neuroligin-1 and the post-synaptic density resident protein PSD-95 in neurons. In Part III we present our conclusion for this body of work.

Part I. Design and validation of proximity lipoylation and proximity coumarin ligation protein-protein interaction reporters

Introduction

Kinetic advantage of lipoic acid ligation by LplA

Our first goal was to try to improve on the dynamic range of proximity biotinylation by reducing background and improving signal. We already had access to a system with potentially ideal kinetics for proximity labeling. Fernández-Suárez *et al.* previously reported an engineered peptide substrate for the enzyme LplA. LplA catalyzes site-specific attachment of lipoic acid to this first-generation LplA acceptor peptide, or LAP1. LplA lipoylates LAP1 with a k_{cat} of 3 min^{-1} , significantly faster than BirA modifies the AP(-3) and therefore hopefully providing an increase in sensitivity for PPI detection. Furthermore, this system should also provide low background, because the K_m of LplA for LAP1 is extremely high, initially estimated at greater than $200 \mu\text{M}$. [3, 4] (We later determined the K_m to be $678 \mu\text{M}$, *vide infra*).

While it is clear that proximity lipoylation has the potential to provide an improved dynamic range of detection over proximity biotinylation, due to the higher k_{cat} and K_m of LplA for its peptide substrate, lipoic acid must be detected using an antibody and so is restricted to immunoblotting and immunofluorescence detection, which must be performed on lysates or fixed cells, respectively. We therefore additionally sought a method for live-cell imaging of PPIs. (Note that we can rationalize the existence of a commercially available anti-lipoic antibody, as free serum concentrations of lipoic acid are essentially too low to measure. [5] This is because lipoic acid is biosynthesized in cells *via* the following pathway: octanoic acid produced by the

fatty acid biosynthesis pathway on an acyl carrier protein scaffold is transferred to dehydrogenase proteins *via* an acyl-enzyme intermediate by the enzyme LipB; the sulfur atoms are then directly installed by the lipoyl synthetase LplA.[6] Therefore, no free lipoic acid is formed during biosynthesis of the cofactor.)

Live-cell labeling of PPIs with coumarin

As discussed in Chapter 1, mutants of LplA at the W37 “gatekeeper” residue at the back of the lipoate binding pocket of the enzyme have been shown to ligate a blue fluorescent coumarin probe to protein targets inside living cells.[7] We therefore set out to develop a live-cell PPI reporter using the same general experimental paradigm as depicted in Figure 1, using coumarin as our probe.

The structure of lipoic acid and the coumarin probe are shown in Figure 4-2. Of course, the 7-hydroxycoumarin probe depicted has up to two negative charges at physiological pH and therefore does not efficiently cross cell membranes, so modifications to the structure are necessary for cellular labeling. Chayasith Uttamapinant has synthesized the di-acetoxymethyl ester derivative, denoted as (AM)₂.coumarin, which can efficiently enter cells in its uncharged state (Figure 4-2).[7] Once inside the cell, non-specific esterases cleave the ester groups, and subsequent spontaneous elimination of formaldehyde offers the de-protected, fluorescent coumarin.[8] Free anionic coumarin in the cytoplasm washes out over time due to the action of non-specific anion transporters.[8] (While it is the case that formaldehyde is toxic, this method of delivery has been generally applied to delivery of non-cell-permeable anionic molecules, perhaps most famously to the commercially available Fura-2 calcium dye.[9] Considering that

these molecules could otherwise only be introduced into cells *via* microinjection, the researcher must consider whether the toxicity tradeoff is acceptable.)

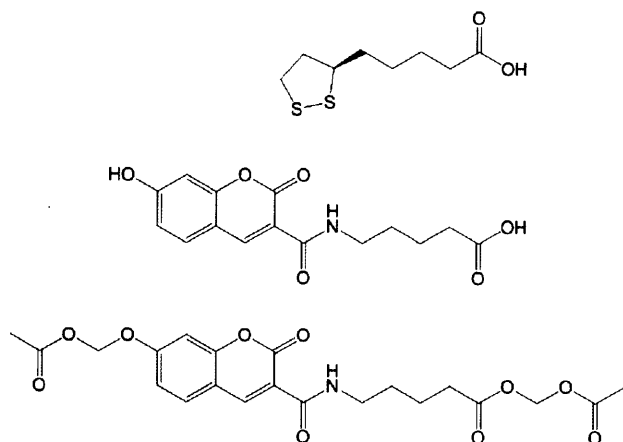


Figure 4-2. Structures of probes utilized in protein-protein interaction (PPI) reporter design. Top, lipoic acid, the endogenous *E. coli* lipoic acid ligase, or LplA, substrate. Center, 7-hydroxycoumarin probe utilized by W37 mutants of LplA. Bottom, diacetylmethyl ester (AM)₁ coumarin structure utilized for cell loading.

We were faced with several choices of enzyme and peptide reporter constituents. Several mutations of LplA have been reported to confer efficient coumarin ligase activity, in particular W37I and W37V. These enzymes have slightly different kinetic parameters for coumarin ligation. LplA(W37I) turns over coumarin with a k_{cat} of 0.96 min^{-1} but a very high K_m of $261 \mu\text{M}$. [7] LplA(W37V) has more favorable parameters, a k_{cat} of 1.14 min^{-1} and K_m of $56 \mu\text{M}$. [7] These enzymes have similarly been shown to have differential sensitivities for cellular labeling of proteins, with W37V providing stronger signal at low expression levels, but also higher background under high expression levels. [7] We would empirically determine which of these enzymes works better for proximity labeling (*vide infra*).

While the kinetic parameters of LAP1 should be equally suitable for proximity coumarin ligation as they are for proximity lipoylation, it was not obvious *a priori* that LAP1 would work in this system. Coumarin ligation onto LAP1 had never previously been observed by other members of the Ting research group, *in vitro* or in cells, probably because both coumarin and LAP1 are both

impaired substrates in terms of the bimolecular rate constant relative to the endogenous LplA substrates lipoic acid and H protein (or a sub-domain of H-protein, called E2p). For this reason, a kinetically efficient LplA peptide substrate, LAP2, was evolved using yeast display.[4] The K_m of wild-type LplA for LAP2 is 13 μM . [4] Coumarin labeling of LAP2 has been demonstrated *in vitro* and *in vivo*. [7] However, we theorized that LAP2 would not suffice for our proximity labeling reporter, because as long as both proteins are present at concentrations higher than 13 μM (which we generally find to be the case when using transient transfection of recombinant constructs, *vide infra*), interaction-independent coumarin labeling would occur; that is, this peptide would show high background labeling. We further reasoned that, although coumarin labeling of LAP1 had never been demonstrated due to its kinetic unfavorability, the forced proximity of LplA and LAP1 that would occur during a PPI might overcome the kinetic barriers to labeling.

The relevant kinetic parameters for LplA labeling of each previously characterized probe and peptide, which inform our reporter design, are summarized in Table 4-1.

	K_m for acceptor peptide	kcat	Reference
BirA + AP + biotin	25 μ M	12 min ⁻¹	[2]
BirA + AP(-3) + biotin	345 μ M	0.53 min ⁻¹	[1]
LplA + LAP1 + 8-azidooctanoic acid	> 300 μ M	0.048 s ⁻¹	[3]
LplA + LAP2 + lipoic acid	13 μ M	0.22 s ⁻¹	[4]
LplA(W37V) + LAP1 + coumarin	vide infra	vide infra	This work
LplA(W37V) + LAP2 + coumarin	Not determined	0.019 s ⁻¹	[7]

Table 4-1. Previously characterized kinetic parameters for LplA enzymes, LplA acceptor peptides (LAPs), and small molecule probes.

We therefore proposed that we could construct a proximity ligation PPI sensor using ligation of coumarin to LAP1 by a W37 mutant of LplA. We then sought model systems in which to investigate and validate our new PPI sensors.

Model system: Rapamycin-dependent interaction of FRB and FKBP

We began our study with a well-characterized model system, the rapamycin-dependent association of FKBP (FK506-binding protein) with a fragment of the mammalian target of rapamycin, mTOR, called FRB, that binds FKBP and rapamycin. Rapamycin is a small molecule derived from *Streptomyces hygroscopicus* that chemically dimerizes FRB and FKBP; it is named for Easter Island, or Rapa Nui, where it was first isolated.[10] In the presence of rapamycin, the ternary complex is stable; the dissociation constant for FRB binding to the FKBP-rapamycin

complex has been reported to be in the range of 2.5 nM[11] to 12 nM.[12] In the absence of rapamycin, the association between FRB and FKBP cannot be measured. Furthermore, a crystal structure of the ternary complex has been reported, facilitating fusion construct design (Figure 4-3). It is clear from examination of the crystal structure that the C termini of FRB and FKBP are only 18 Å apart.[13] Therefore, we expect fusions of LplA and LAP1 at the C-termini of FRB and FKBP to afford labeling because the enzyme and substrate should be accessible to one another.

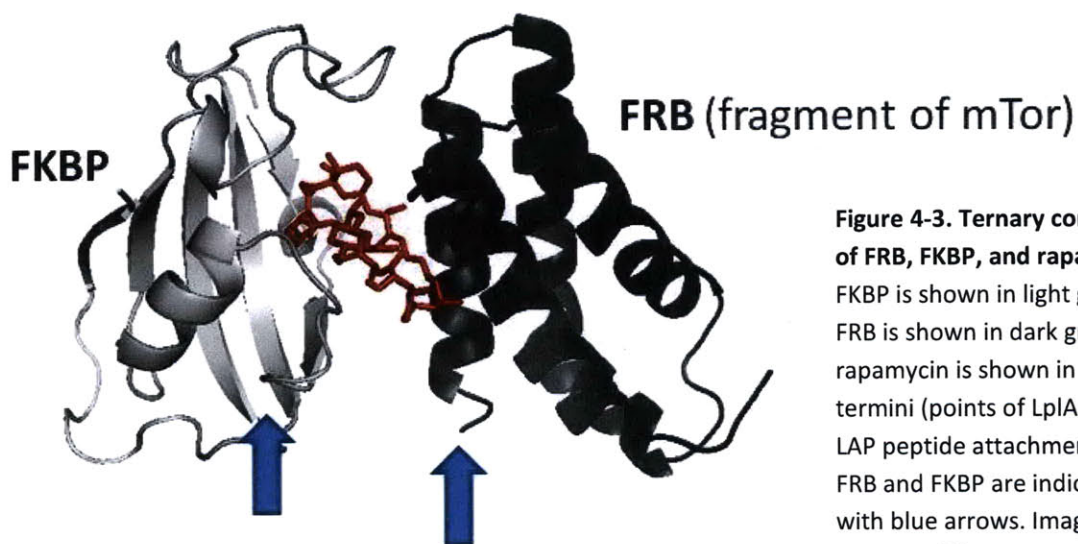


Figure 4-3. Ternary complex of FRB, FKBP, and rapamycin. FKBP is shown in light gray, FRB is shown in dark gray, and rapamycin is shown in red. C-termini (points of LplA and LAP peptide attachment) of FRB and FKBP are indicated with blue arrows. Image generated from PDB ID 1FAP.

Results and Discussion

Proximity lipoylation to detect PPIs *in vitro* and in cells

Construct design and in vitro proximity lipoylation

We first attempted to create a new PPI reporter based on lipoylation of LAP1 by LplA. Our first task was to design our FRB and FKBP fusion constructs. In particular, we had to determine the optimal LAP sequence, because several LAP1 peptides have been reported. A minimal 17-amino

acid sequence is required for LplA recognition; however, this peptide is recognized well as an N-terminal fusion, but poorly as a C-terminal fusion.[3] This may be due to peptide termination with a proline residue, which could influence its secondary structure. Therefore a 22-amino acid LAP1 peptide, with a 5-amino acid extension at its C-terminal end, has been generally utilized as an LplA substrate.[3, 14] However, we hypothesized that the 17-mer peptide would be sufficient for labeling in the context of fusions to interacting proteins because any effects of poor affinity might be overcome by high local concentration in the case of interaction. We quantitatively confirmed the suitability of the 17-mer peptide in later studies (vide infra). The *in vitro* data reported here utilize the 17-mer LAP1. The initial fusion pair we chose was FRB-LplA, with LplA fused to the C-terminus of FRB, and FKBP-LAP1, with the 17-mer LAP1 fused to the C-terminus of FKBP (Figure 4-4). We began by investigating whether proximity lipoylation can work *in vitro*. FRB-LplA and FKBP-LAP1 were over-expressed in *E. coli* and purified by nickel affinity chromatography (Figure 4-5).

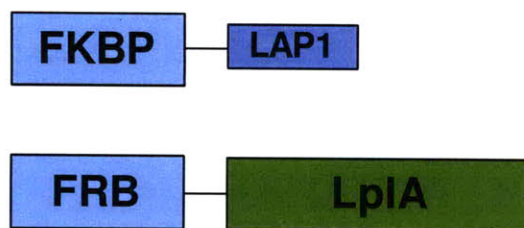


Figure 4-4. Domain structures of proximity lipoylation constructs. LAP1 is fused to the C-terminus of FKBP, and LplA is fused to the C-terminus of FRB.

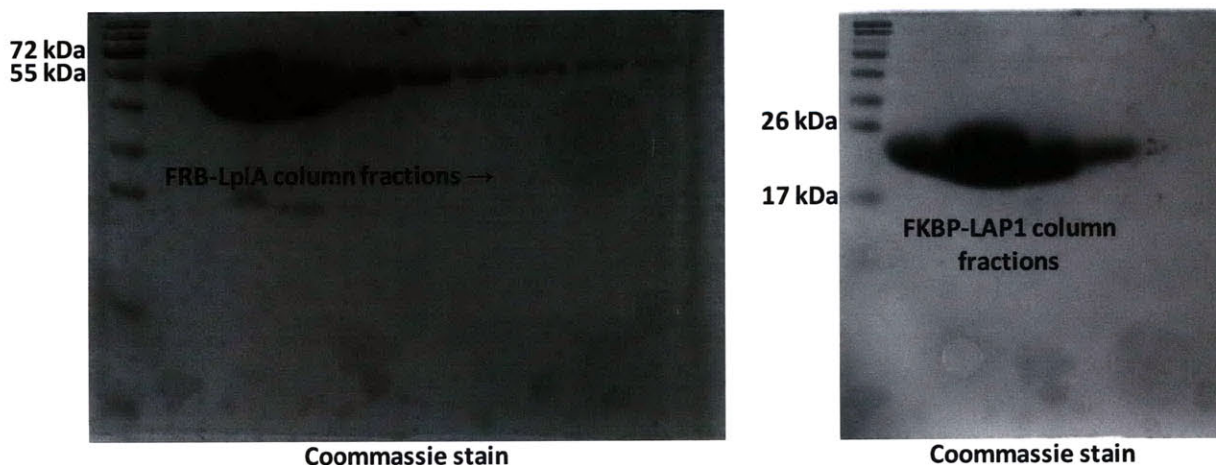


Figure 4-5. Purification of FRB-LplA and FKBP-LAP. Hexahistidine-tagged FRB-LplA and FKBP-LAP1 were expressed in *E. coli* and purified by nickel affinity chromatography as described in Methods. Purity was analyzed by electrophoresis on 16% SDS-PAGE gels. The Coomassie-stained gels are shown.

We investigated the signal-to-noise ratio of *in vitro* proximity lipoylation using an immunoblotting assay. The proteins were combined at 10 μ M and pre-incubated either with rapamycin, to induce the interaction, or without rapamycin as a negative control. We then performed labeling by adding lipoic acid and ATP for varying times, then quenched the reaction with the divalent metal cation chelator ethylenediaminetetraacetic acid (EDTA). (EDTA affords quenching because LplA requires Mg^{2+} for ATP hydrolysis. Subsequent blotting to nitrocellulose and detection with anti-lipoic acid antibody revealed selective lipoic acid labeling in the presence of rapamycin, with low (but measurable) background when rapamycin is omitted from the reaction mixture (Figure 4-6). With labeling times of both 1 minute and 4 minutes, the signal-to-background ratio of FKBP-LAP1 labeling, determined by spot densitometry, is approximately 12.5:1.

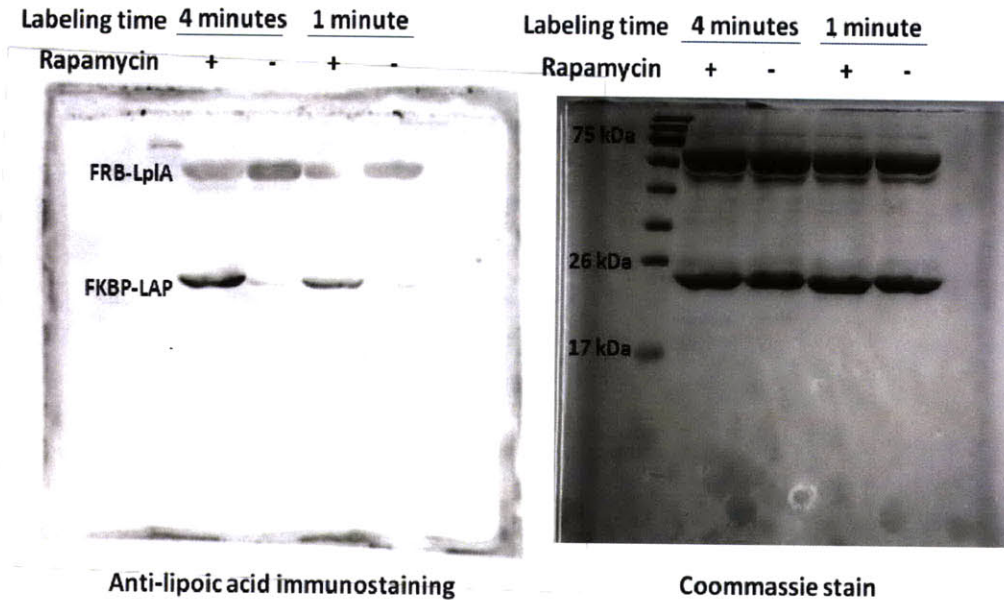


Figure 4-6. *In vitro* proximity lipoylation analyzed by immunoblotting. Purified FRB-LplA and FKBP-LAP1 were combined at 10 μ M in the presence or absence of rapamycin, then treated with ATP and lipoic acid to afford labeling. FKBP-LAP1 lipoylation was assessed by immunoblotting with anti-lipoic acid antibody (left). SDS-PAGE and Coomassie staining of the same reactions serves as a loading control (right). The left-most lane of both the immunoblot and the gel is a molecular weight ladder.

Unfortunately, we also observe significant LplA self-labeling, a general problem of both LplA and BirA, which may contribute background in experiments where enzyme is not separated from substrate, as is the case in cellular imaging experiments. The mechanism of BirA and LplA self-labeling is unclear. In particular, the primary rate constants for the two-step reaction of LplA have not yet been reported in the literature, to our knowledge. However, BirA has been reported to slowly release the weakly amine-reactive intermediate biotinyl-5'-AMP into solution (k_{off} of 0.00027s^{-1}), and this released intermediate can non-specifically react with bovine serum albumin in solution.[15, 16] Purified free biotinyl-5'-AMP can react with proteins, albeit very slowly, in solution in the absence of enzyme.[17] Furthermore, "promiscuous" mutants of BirA have been reported that release biotinyl-5'-AMP at accelerated rates; these enzymes exhibit greater self-biotinylation, even during their expression in *E. coli*.[15, 16] It has

therefore been postulated that self-labeling is a side reaction that is the result of the slow release of the activated intermediate; the relatively high local concentration of the activated intermediate (either biotinyl- or lipoyl-5'-AMP) in the vicinity of the enzyme means that enzymatic lysine residues (or the N-terminus) are more likely than other proteins (and solvent) in the cell to react with it. While our methods would be most effective if this enzymatic self-labeling could be entirely eliminated, in particular in the case of BirA, the rate of biotin transfer to the peptide substrate (0.1 s^{-1})[2] is much faster than the rate of intermediate release (0.00027 s^{-1}).[18] This means that as long as labeling times are as short as possible (and for biotin and lipoic acid labeling, as long as the enzyme is produced in cells under conditions of minimal free biotin or lipoic acid as necessary, in order to prevent self-labeling during expression), the background of enzyme self-labeling relative to the peptide labeling signal will be minimal.

Proximity lipoylation in living cells with immunoblotting detection

Proximity lipoylation worked so well *in vitro*, we decided to test the method inside cells. We further wished to test our premise that the poor- K_m LAP1 peptide is required for low background in the absence of a PPI to provide our kinetic switch. In order to do so, we utilized three peptide or protein fusions to FKBP: LAP1, for which LplA has a poor K_m (which we measure to be $678 \mu\text{M}$, *vide infra*); LAP2, for which LplA has a lower K_m of $13 \mu\text{M}$; and E2p, a domain from the endogenous lipoylated *E. coli* protein pyruvate dehydrogenase, for which the enzyme is expected to have low K_m , though this value has not been directly measured.[4] We reasoned that, because LplA has high k_{cat} for lipoylation of all three, LAP1, LAP2, and E2p, they

should all be labeled efficiently in the case of a PPI, which provides pseudo-zero-order conditions; however, we expected that, since transient transfection is likely to provide expressed fusion proteins at concentrations of at least μM inside cells, at or near the K_m of LplA for LAP2 and E2p, these fusions should also be rapidly labeled in the absence of an interaction, providing high background (whereas LAP1 should be labeled only very slowly in the absence of the interaction).

We co-expressed FRB-LplA with each of FKBP-LAP1, FKBP-LAP2, and FKBP-E2p in COS-7 cells; each of these constructs is over-expressed under control of the same strong promoter, the cytomegalovirus (CMV) promoter. We then treated the cells with rapamycin for one hour to promote the intracellular interaction of FRB and FKBP. Parallel cells received no rapamycin, in order to investigate the background when no interaction occurs. The cells were then labeled with lipoic acid, with 500 μM lipoic acid in DPBS for 1 minute, then immediately lysed and subjected to SDS-PAGE and immunoblotting analysis to detect the lipoic acid modification (Figure 4-7). (While lipoic acid has been demonstrated to enter intestinal cells by active transport through the action of a proton-linked monocarboxylic acid transporter,[19] it is likely that at these high concentrations the primary mechanism of cell entry is passive diffusion, as has been documented for biotin, which is also an amphiphilic monocarboxylic acid, at concentrations greater than 2 μM .[20]) The anti-lipoic acid signal was quantitated for FKBP-LAP1 by spot densitometry and reveals a signal-to-background ratio of approximately 15:1. Therefore LAP1 affords good labeling in the presence of rapamycin (i.e., when a PPI occurs) and low background in the absence of rapamycin (i.e., the absence of a PPI). Additionally, it is clear that FKBP-LAP2 and FKBP-E2p are equivalently labeled in the presence and absence of the PPI,

validating our premise that high-affinity peptides create high background and cannot be used in a PPI reporter.

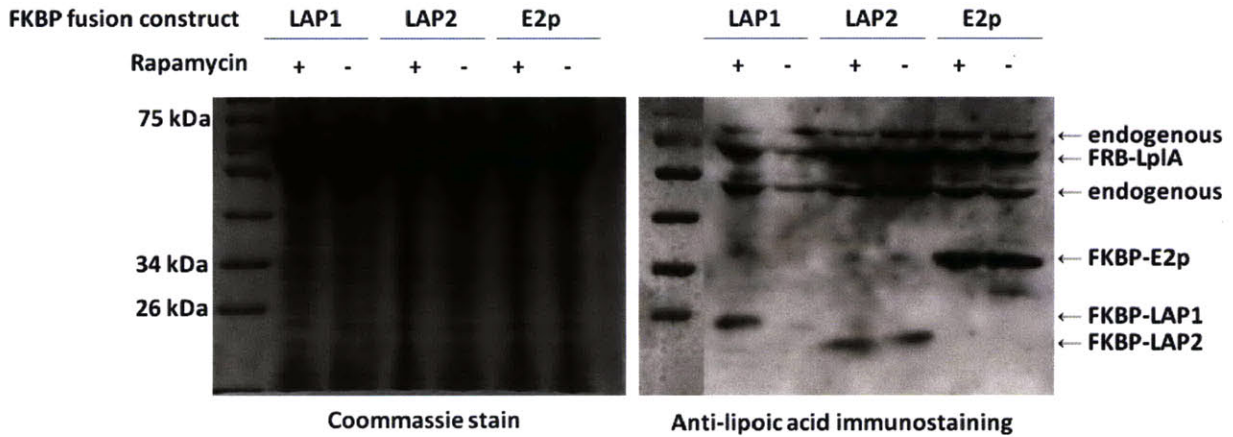


Figure 4-7. Intracellular proximity lipoylation analyzed by immunoblotting. FRB-LplA and FKBP-LAP1 were co-expressed in COS-7 cells; alternatively, FKBP-LAP1 was replaced with FKBP-LAP2 and FKBP-E2p to test the effect of enzyme-peptide affinity on proximity lipoylation background. Cells were labeled with 500 μ M lipoic acid in DPBS for 1 minute, then lysed immediately with direct application of SDS-PAGE loading buffer. Lipoylation was assessed by immunoblotting with anti-lipoic acid antibody (right). Parallel wells were run on an SDS-PAGE gel and Coomassie stained as a loading control. Only FKBP-LAP1 exhibits signal relative to background in the presence of rapamycin; high background is observed for FKBP-LAP2 and FKBP-E2p. FRB-LplA self-labeling is observed. Endogenous lipoylated proteins are indicated.

Proximity lipoylation in living cells with immunofluorescence detection

Having good evidence that proximity lipoylation works inside mammalian cells, we attempted to apply proximity lipoylation to cellular imaging. Because we had a good anti-lipoic acid antibody in hand, we expected that we would be able to detect lipoylation by immunofluorescence. Of course, because lipoylation can only be detected by antibody binding, imaging-based analysis of proximity lipoylation is restricted either to the cell surface or to the cytoplasm of fixed cells. Nevertheless, these are the same restrictions facing proximity biotinylation, and both methods can provide information about PPIs in mammalian cells not accessible to biochemical methods of PPI analysis.

We transfected COS-7 cells with FRB-LplA and FKBP-LAP1-NLS, where NLS is a nuclear localization signal that causes the FKBP-LAP1 to be trafficked into the nucleus. FRB-LplA is expressed throughout the cytoplasm, and immunostaining indicates that, upon rapamycin addition, it redistributes to the nucleus, probably due to its association with FKBP-LAP1-NLS (data not shown). We then treated the cells with rapamycin for one hour to promote the intracellular interaction of FRB and FKBP. The cells were then labeled with lipoic acid, as described in the Methods section, then immediately washed with ice-cold DPBS to stop the reaction and fixed. Immunostaining reveals that FKBP-LAP, which bears a *c-myc* epitope tag, is restricted to the nucleus of these cells (Figure 4-8). In the presence of rapamycin, lipoic acid signal is detected above background, and that signal coincides with the anti-*c-myc* immunostaining, indicating that the lipoic acid labeling occurs specifically on the LAP1 tag. Furthermore, no nuclear lipoylation signal is detected in the absence of rapamycin.

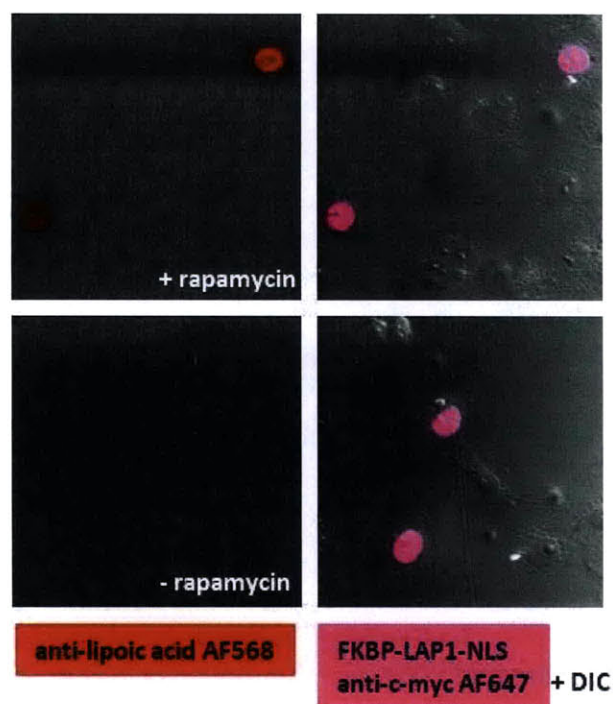


Figure 4-8. Immunofluorescence imaging of proximity lipoylation in the nucleus of COS-7 cells. COS-7 cells were transfected with FRB-LplA and FKBP-LAP1-NLS. Cells were incubated with 100 nM rapamycin for 1 hour, then labeled with lipoic acid for 1 minute, followed by formaldehyde fixation and methanol permeabilization. Anti-*c-myc* staining was performed to visualize *c-myc* epitope-tagged FKBP-LAP1-NLS, and anti-lipoic acid staining was performed to visualize the lipoic acid labeling. Anti-lipoic acid staining is shown on the left, in red. Anti-*c-myc* staining is shown on the right in pink, merged with the differential interference contrast (DIC) image. A negative control with rapamycin omitted demonstrates that lipoylation signal is interaction-specific.

Unfortunately, there are serious background problems with this immunofluorescence assay. Endogenous lipoylated proteins in mitochondria are brightly visualized with anti-lipoic acid antibody, and we could not identify any condition that significantly decreased this endogenous background, including 12-hour serum starvation (data not shown). Therefore, only when the recombinant fusions are strongly over-expressed, as is the case in our COS-7 cells, does the proximity lipoylation signal rise significantly above the endogenous background. In fact, we were never able to accomplish proximity lipoylation of a cytoplasmic FKBP-LAP1 construct above the mitochondrial background. Our nuclear construct is probably more brightly labeled because nuclear localization increases the local concentration of the labeled protein, thereby increasing the signal; the diffuse cytoplasmic construct does not achieve a sufficiently high local concentration to be detected above the background. Furthermore, even nuclear FKBP-LAP1 could not be reproducibly labeled above background in other cell lines, such as HEK cells, which do not over-express it as strongly in our hands. Finally, LplA self-labeling with high-affinity small-molecule substrates, such as lipoic acid, is a source of background that cannot be removed, as previously discussed, so good signal-to-background ratios are only obtained when substrate is expressed at higher levels than enzyme. It is therefore clear that proteins of interest expressed at near-endogenous levels probably cannot be detected with proximity lipoylation *via* an imaging readout.

In conclusion, proximity lipoylation provides a good response to PPIs in terms of signal-to-noise *in vitro*. However, its utility in the cellular context is limited to detection by immunoblotting applications because endogenous lipoylated proteins in the mitochondria contribute background to immunofluorescence detection.

Proximity coumarin ligation for live-cell PPI imaging

Method demonstration and optimization: Identification of the optimal enzyme and peptide, demonstration of labeling specificity

Due to the limited utility of proximity lipoylation for cellular imaging, we next sought a method for imaging PPIs, in particular inside living cells. We attempted to create a proximity coumarin ligation reporter for this purpose. Our first task was to determine whether proximity coumarin ligation works in living cells, and furthermore what enzyme and peptide constructs comprise the optimal reporter. That is, we needed to quantitatively determine which LplA coumarin ligase mutant (i.e., LplA(W37I) or LplA(W37V)) and which LAP1 peptide (i.e., 17-mer or 22-mer) give the best proximity coumarin ligation response when an interaction occurs and the lowest background in the absence of an interaction. All other things being equal, we also require that the enzyme and peptides are minimally perturbative in the context of the fusion constructs; in particular, the peptide substrate should be as small as possible.

We therefore compared the efficiency of labeling for each enzyme-substrate pair under very low expression conditions, to determine which provides the greatest labeling sensitivity. Furthermore, the FKBP-LAP1 substrate was targeted to the nucleus with the nuclear localization signal, allowing us to determine if either enzyme is more efficient for labeling substrates with specific sub-cellular localization. We fixed the cells in this experiment in order to query the expression level of our constructs by immunostaining, so that we could compare their labeling efficiency directly. Gratifyingly, we observed coumarin ligation in the presence of rapamycin for

all constructs, as well as extremely low background in the absence of rapamycin (representative images are shown in Figure 4-9).

Our single-cell quantitation of this dataset is shown in Figure 4-10. First we note that there is no significant difference in the labeling intensities for the 17-mer peptide as compared to the 22-mer peptide in the presence of rapamycin, and that background in the absence of rapamycin for both peptides is similarly low. Therefore we conclude that, given no difference in labeling efficiency, the 17-mer LAP1 construct is preferable because of its smaller size. Second, we observe a few brightly labeled cells using FRB-LpIA(W37I) in the presence of rapamycin, but that most cells are dimly labeled, invariant to expression level of the FKBP-LAP1 construct (as measured by anti-c-myc immunofluorescence). In contrast, FRB-LpIA(W37V) displays an essentially linear increase in coumarin labeling intensity with FKBP-LAP1 expression, reaching the same maximal intensity values attainable with LpIA(W37I). This data is consistent with a model in which LpIA(W37I) must be present in the cell above a certain threshold concentration in order to afford any labeling, whereas LpIA(W37V) provides labeling across a range of enzyme concentrations; since we have not detected the enzyme expression level in this experiment, this remains a hypothesis. We can conclude that LpIA(W37V) provides better sensitivity of labeling than LpIA(W37I). We note that LpIA(W37V) has been previously reported to produce higher background labeling at high expression levels than LpIA(W37I), [7] so it may be wise when applying proximity coumarin labeling to new proteins to test both enzymes and empirically determine which provides the best signal-to-background ratio for that particular PPI.

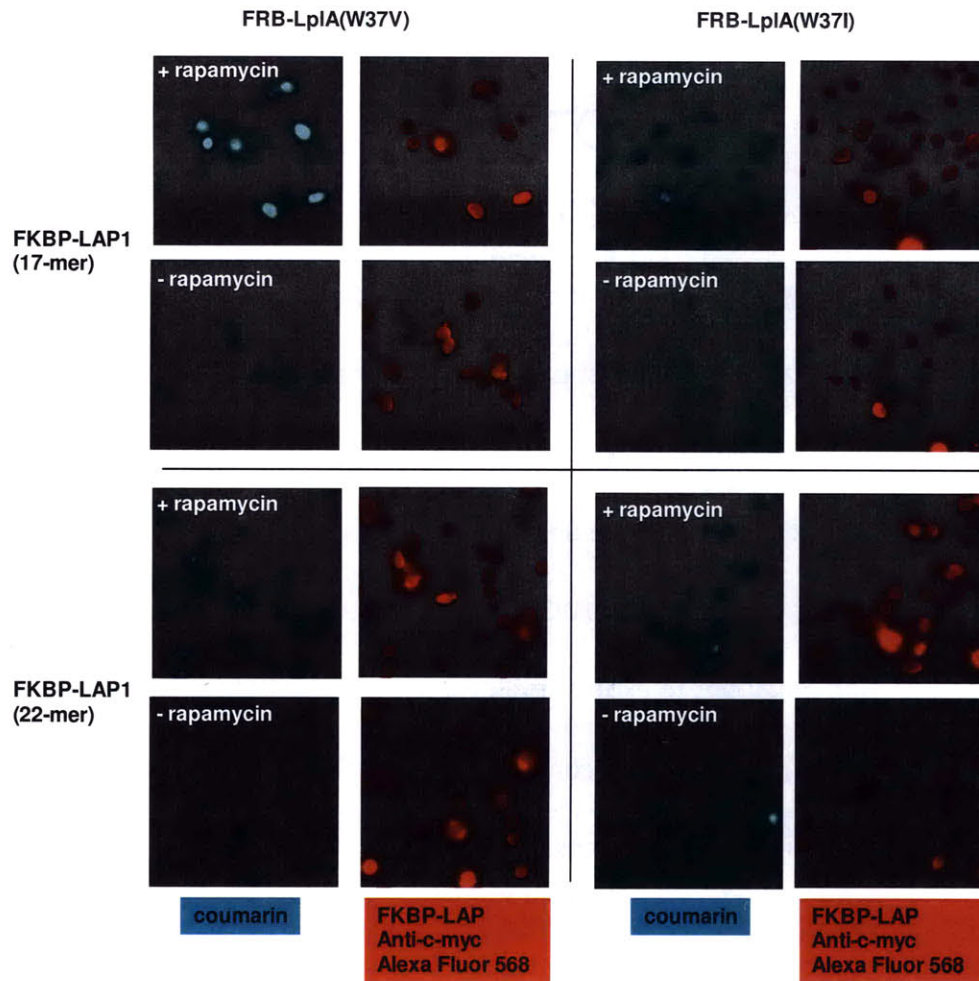


Figure 4-9. Determination of optimal enzyme and peptide for proximity coumarin labeling:

Representative images. HEK cells were transfected with FRB-LpIA(W37I) or FRB-LpIA(W37V) in combination with either FKBP-LAP1(17-mer)-NLS or FKBP-LAP1(22-mer)-NLS. Cells were treated with rapamycin, or rapamycin was omitted to assess background. Labeling was performed with 20 μ M (AM)₂-coumarin for 20 minutes. Cells were fixed and expression of FKBP-LAP1-NLS constructs was assessed by anti-c-myc immunostaining. Coumarin labeling is shown on the left in cyan. FKBP-LAP1-NLS anti-c-myc immunofluorescence is shown on the right in red.

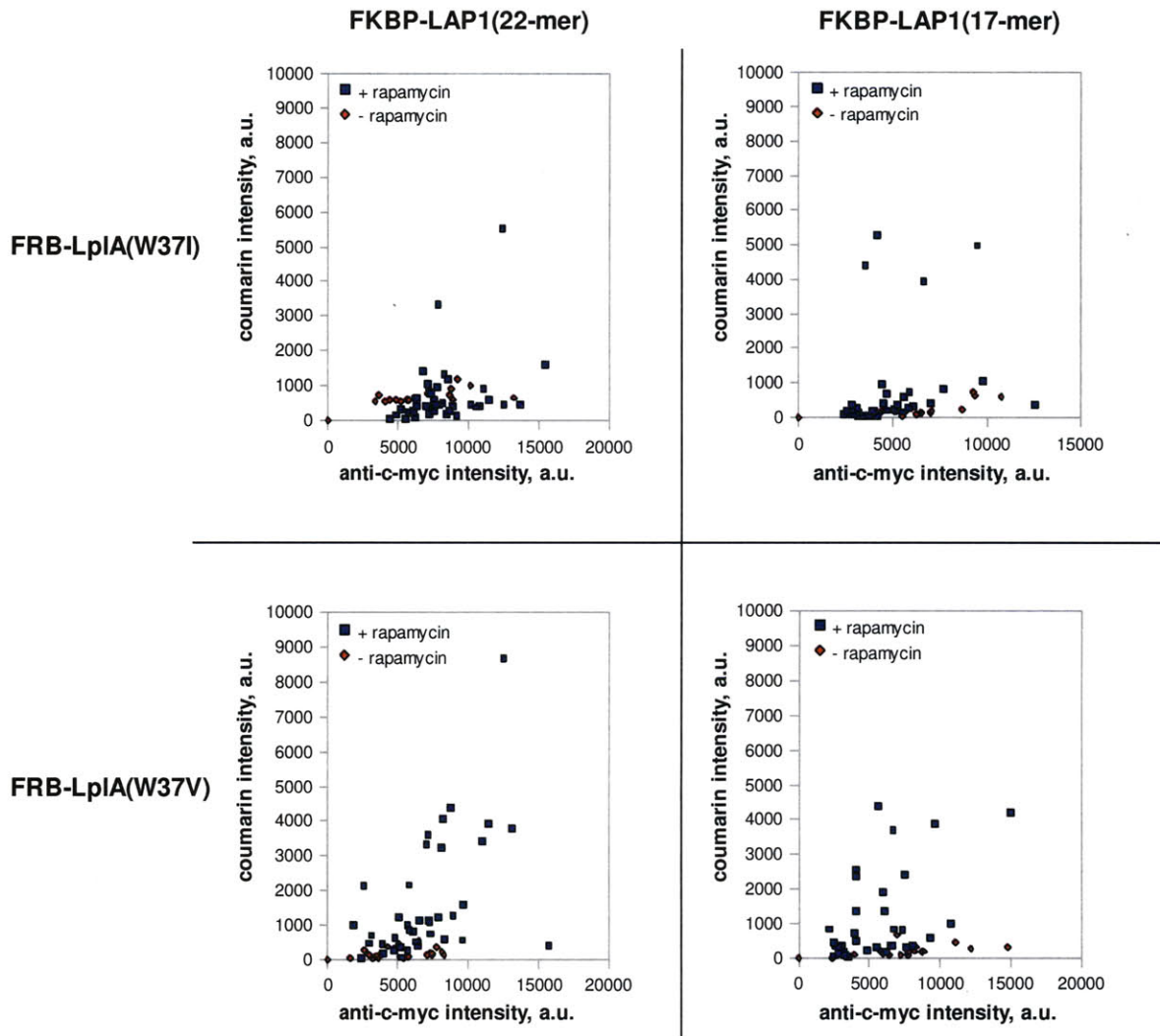


Figure 4-10. Determination of optimal enzyme and peptide for proximity coumarin labeling: Single-cell plots of coumarin intensity against anti-c-myc intensity. Single-cell plots of intensities from the dataset represented in Figure 4-8 are presented, wherein average per-cell coumarin intensity is plotted against the FKBP-LAP1-NLS expression level, as assessed by anti-c-myc immunofluorescence. No difference is observed between the 17-amino acid and 22-amino acid LAP1 sequences, indicating that they are equally competent for coumarin labeling. LpIA(W37V) gives better labeling at intermediate expression levels than does LpIA(W37I).

We then set out to demonstrate that coumarin proximity ligation works in living cells and that the labeling is site-specific and interaction-dependent, using LpIA(W37V) and the 17-amino acid LAP1 peptide. We also wanted to validate our hypothesis that LAP2 should give a poor response to PPIs in the case of coumarin ligation as well due to high background.

When FRB-LpIA(W37V) and FKBP-LAP1 are co-expressed in the cytoplasm of living HEK cells, addition of rapamycin followed by coumarin labeling for only ten minutes produces bright fluorescence in transfected cells (Figure 4-11). (In this image and all live-cell images that follow, the green fluorescent protein, GFP, is used as a co-transfection marker to identify living cells that are expressing our constructs.) In the absence of rapamycin, background is essentially zero (Figure 4-10). When LAP1 is replaced with LAP2, coumarin labeling is observed both in the presence and absence of rapamycin, as expected (Figure 4-11). Furthermore, the average labeling intensity on LAP1 and LAP2 are not significantly different, indicating that the sensitivity of labeling is not impaired by using the low-affinity substrate. Therefore, our proposed reporter design, requiring a high k_{cat} for labeling sensitivity, but a low K_m for minimal background, has been validated.

We further demonstrated the specificity of proximity coumarin labeling in three ways. First, we generated a point mutant of FKBP-LAP1, FKBP-LAP1(K→A), in which the single lysine residue of the LAP1 peptide that is modified by LpIA is mutated to an alanine, which we expect to eliminate labeling. When FRB-LpIA(W37V) is coexpressed with this construct, addition of rapamycin and coumarin fail to generate coumarin signal (Figure 4-11). This demonstrates that our coumarin proximity labeling method is site-specific. Second, we used FRB-LpIA in place of FRB-LpIA(W37V), as a control for enzymatic activity, since wild-type LpIA has been demonstrated not to accept the coumarin probe as a substrate. This construct fails to label FKBP-LAP1, demonstrating that enzymatic activity is necessary for coumarin proximity ligation (Figure 4-11).

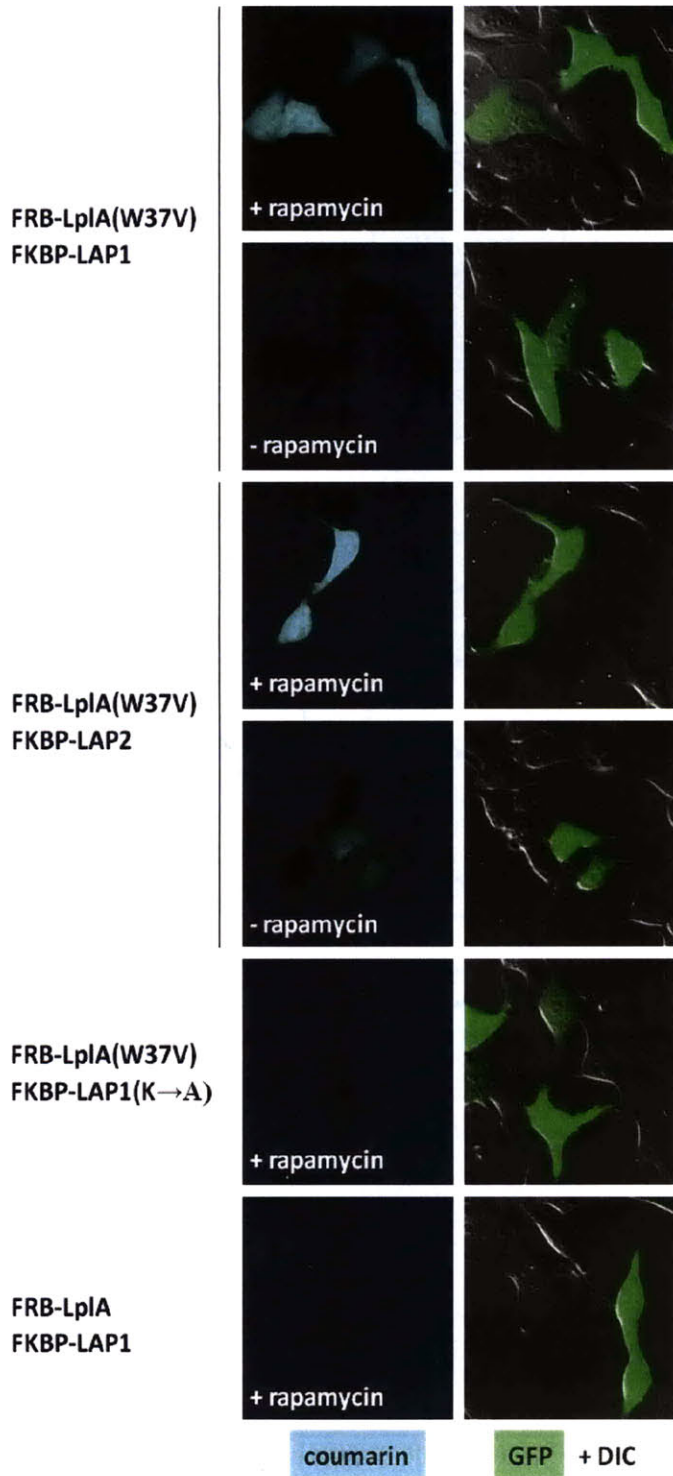


Figure 4-11. Imaging proximity coumarin ligation in the cytoplasm of living HEK cells. HEK cells were transfected with FRB-LplA(W37V), FKBP-LAP1, and GFP as a co-transfection marker. Cells were treated with rapamycin, then labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 60 minutes. Coumarin signal is shown to the left in blue. GFP signal, which denotes transfected cells, is superimposed on the DIC image on the right. Transfected cells treated with rapamycin display coumarin signal (top row). A negative control with rapamycin omitted shows no labeling (second row). If FKBP-LAP1 is replaced with FKBP-LAP2, coumarin labeling occurs independent of rapamycin addition (third and fourth rows) due to the high affinity of LplA(W37V) for this peptide. Additional negative controls with a lysine to alanine point mutation in LAP1 or replacement of LplA(W37V) with wild-type LplA are shown in rows five and six.

Third, we repeated the same experiment, but restricted all of our FKBP-LAP substrate fusions to the nucleus by appending an NLS sequence (Figure 4-12). In this case we see only nuclear coumarin signal; since FRB-LpIA(W37V) is not specifically localized in the cell, the nuclear labeling demonstrates that our enzyme is only labeling the LAP substrate and not other cytoplasmic proteins.

We further wanted to confirm that our labeling corresponds to the presence of both of our reporter constructs in the cells, and that the localization of the coumarin signal matches the localization of our FKBP-LAP constructs. We therefore labeled the cells as above, but performed fixation and immunofluorescence to visualize our FRB-enzyme and FKBP-substrate proteins (Figure 4-13). Our FRB fusion constructs all bear an HA epitope tag, and our FKBP fusion constructs all bear a *c-myc* tag, facilitating immunostaining. The results indicate that coumarin labeling requires both substrate and enzyme to be present in the cell; furthermore, coumarin localization is coincident with FKBP-LAP, as expected. Finally, we see that the protein expression levels in the absence of rapamycin and with our alanine and wild-type LpIA controls are essentially identical to the expression levels we see for FRB-LpIA(W37V) and FKBP-LAP1, indicating that our negative controls have low background because of the intrinsic properties of the system, not because of differing expression levels.

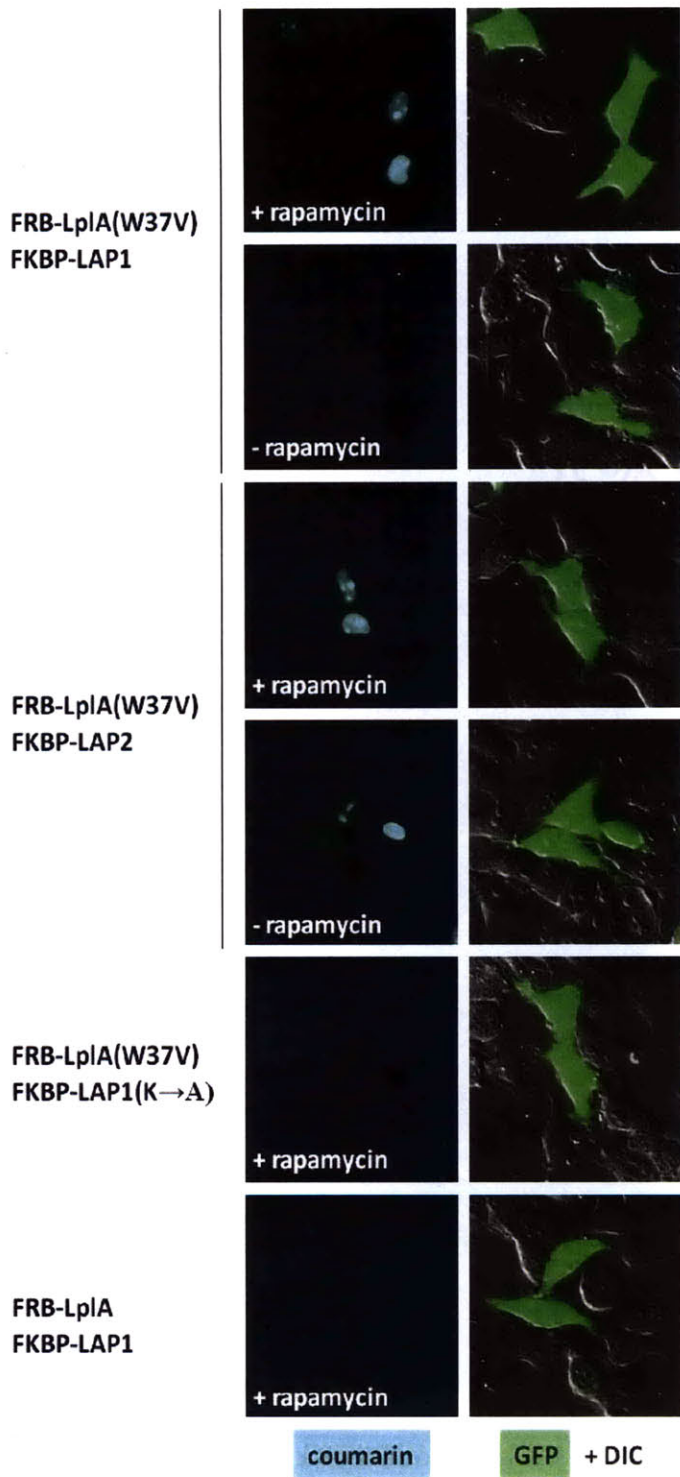


Figure 4-12. Imaging proximity coumarin ligation in the nucleus of living HEK cells. All experimental conditions and controls are exactly as described in the legend for Figure 6, except that FKBP-LAP1 is replaced with FKBP-LAP1-NLS, which is actively restricted to the nucleus. Similarly, FKBP-LAP2-NLS and FKBP-LAP1(K→A)-NLS are utilized.

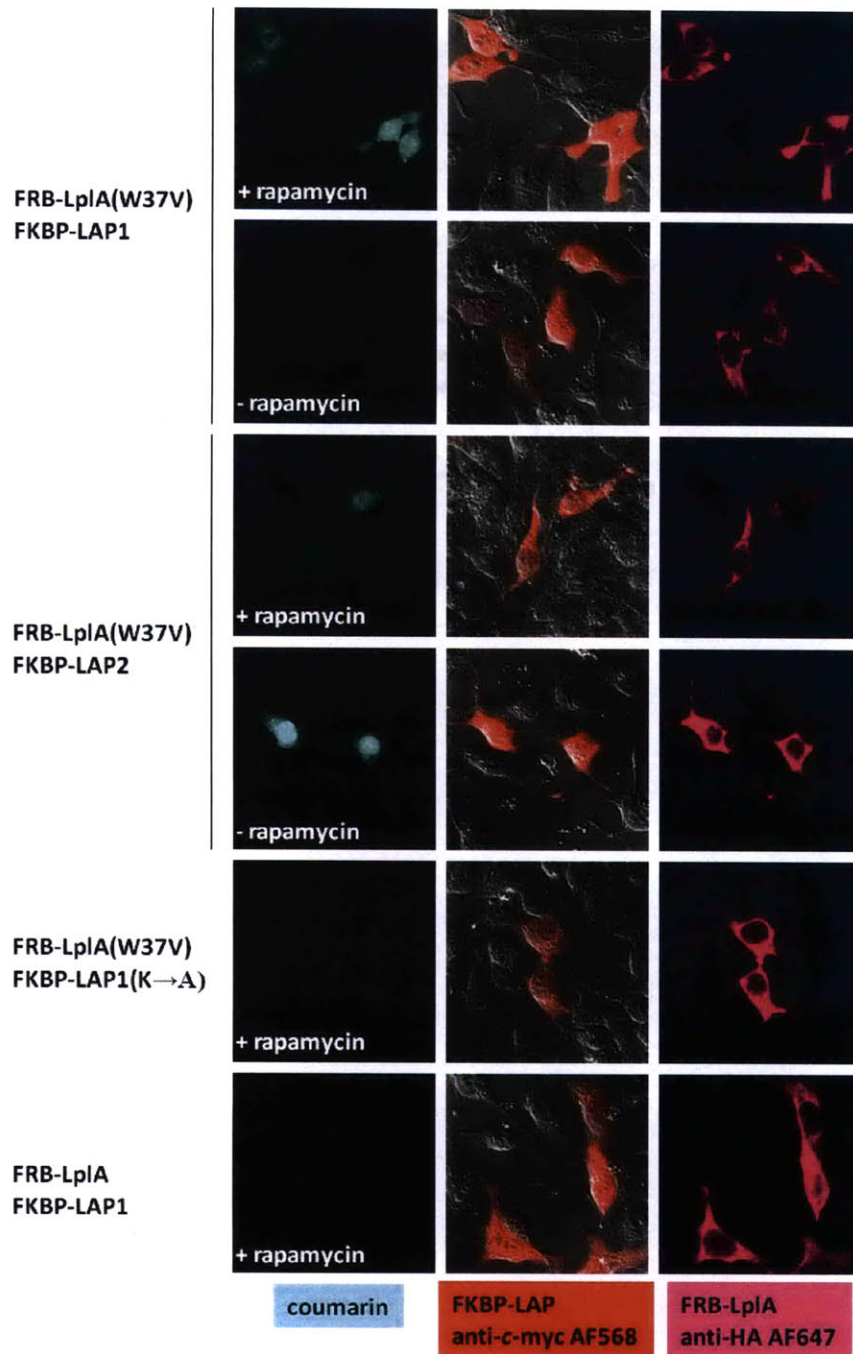


Figure 4-13. Investigation of FRB-LplA(W37V) and FKBP-LAP1 expression and localization during proximity coumarin ligation by immunofluorescence. HEK cells were transfected with FRB-LplA(W37V) and FKBP-LAP1. Controls are included with FKBP-LAP2 or a point mutant of FKBP-LAP1 replacing FKBP-LAP1, or FRB-LplA replacing FRB-LplA(W37V). Cells were treated with rapamycin, then labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 30 minutes, then formaldehyde fixation and methanol permeabilization. Anti-c-myc staining was performed to visualize FKBP-LAP constructs, and anti-HA staining was performed to visualize FRB-LplA constructs. Coumarin signal is shown to the left in blue. Anti-c-myc staining is shown in red superimposed on the DIC image in the center. Anti-HA staining is shown in pink on the right.

Generality of proximity coumarin ligation: investigating different fusion geometries and applying the method in other cell lines

Because many PPI detection methods, including FRET and PCAs, are sensitive to fusion geometry, we sought to determine if changing the orientation of our LpIA and LAP fusions to FRB and FKBP produced differences in proximity labeling. We cloned the “swapped” pair of FKBP-LpIA(W37V) and FRB-LAP1 and tested this pair side-by-side with our original set of constructs, FRB-LpIA(W37V) and FKBP-LAP1. The domain structures of these constructs are shown in Figure 4-14. The cells exhibiting the highest expression levels (that is, the brightest expression of a GFP co-transfection marker) of the transfected constructs are labeled to a similar extent in both cases (Figure 4-15). However, we observe more labeled cells, and labeling at intermediate expression levels, for the original pair. Both fusion pairs afford similarly low background in the absence of rapamycin. This tells us that, while the swapped pair does afford proximity coumarin labeling, the sensitivity under these labeling conditions is decreased.

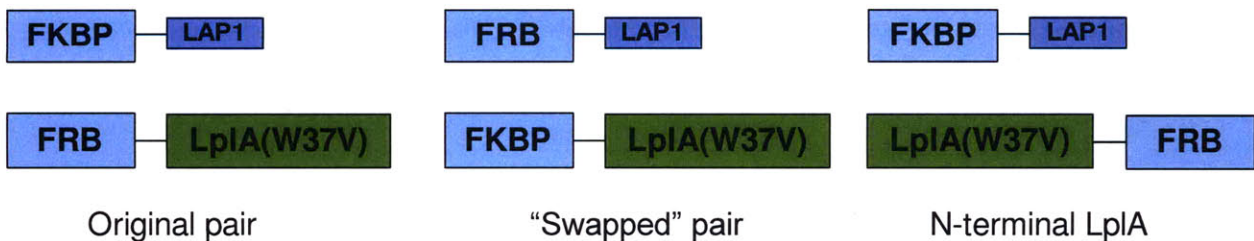


Figure 4-14. Domain structures of proximity coumarin ligation constructs. In the original pair, LAP1 is fused to the C-terminus of FKBP and LpIA(W37V) is fused to the C-terminus of FRB. In the “swapped” pair, LAP1 is fused to the C-terminus of FRB and LpIA(W37V) is fused to the C-terminus of FKBP. We also cloned a construct where LpIA(W37V) was fused to the N-terminus of FRB.

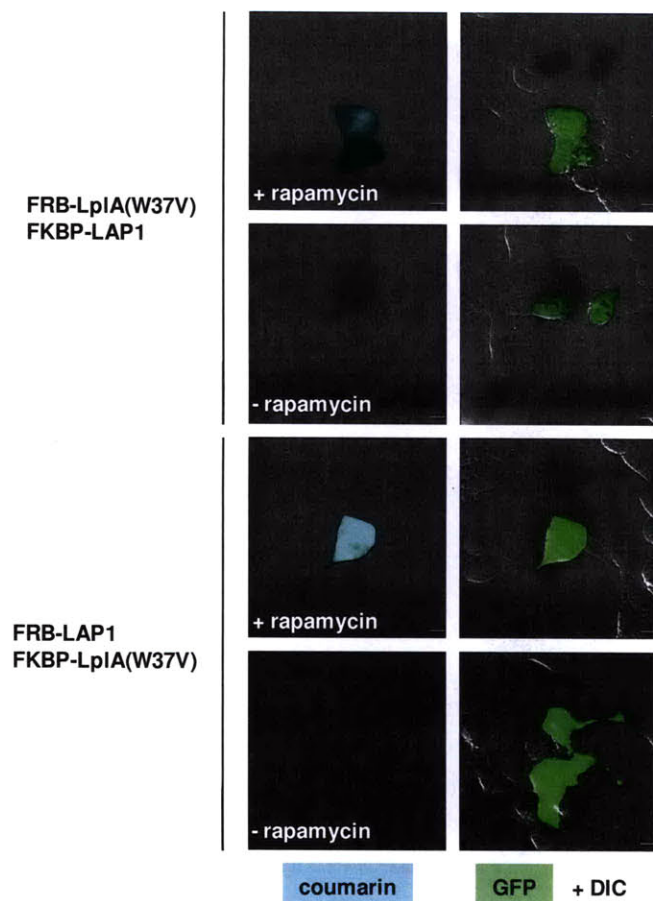


Figure 4-15. Investigation of geometric sensitivity of proximity coumarin ligation. HEK cells were transfected with FRB-LpIA(W37V) and FKBP-LAP1, or FKBP-LpIA(W37V) and FRB-LAP1. GFP is included as a co-transfection marker. Cells were treated with rapamycin, then labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 60 minutes. Coumarin signal is shown to the left in blue. GFP signal, which denotes transfected cells, is superimposed on the DIC image on the right. Only the brightest GFP-expressing cells in the “swapped pair” case are labeled. Scale bars, 10 μm.

We also cloned a construct where our enzyme is fused to the N-terminus of FRB rather than its C-terminus, which we denote LpIA(W37V)-FRB. Examination of the crystal structure reveals that, while the N-terminus of FRB is farther from the C-terminus of FKBP, it is still oriented in the correct direction to potentially afford enzyme-substrate accessibility. Under gentle labeling conditions on the cytoplasmic FKBP-LAP1 construct, which afford labeling using FRB-LpIA(W37V), we do not observe proximity coumarin labeling using LpIA(W37V)-FRB (data not shown). Therefore, proximity coumarin labeling works in two different orientations of enzyme and substrate peptide, but the labeling sensitivity does remain dependent on the fusion geometry. As a result, when investigating a new PPI where no crystal structure is available to guide construct design, every possible geometric fusion must be made.

All proximity coumarin ligation experiments to this point had been conducted in HEK cells, an immortal line of human embryonic kidney fibroblasts. It is, however, important to show that the method is sufficiently general to work in multiple cell lines, because different cell lines may show differences in expression levels, dye uptake, and wash-out efficiency. We therefore tested proximity coumarin labeling of our FRB-LpIA(W37V) and FKBP-LAP17 constructs in COS-7 and HeLa cells using the same labeling protocol developed for HEK cells (Figure 4-16). All three cell lines are brightly labeled in the presence of rapamycin, show very low background labeling in the absence of rapamycin, and exhibit no obvious changes in cell health or morphology after labeling. We therefore conclude that proximity labeling is generally applicable in immortalized cell lines.

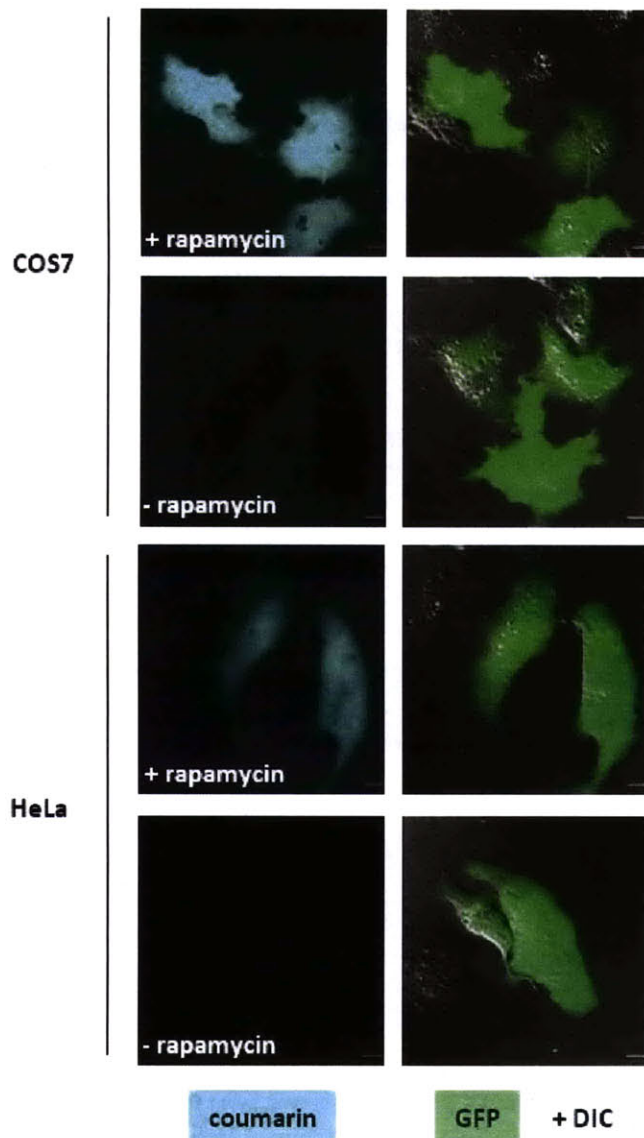


Figure 4-16. Imaging proximity coumarin ligation in the cytoplasm of living COS-7 and HeLa cells. COS-7 and HeLa cells were transfected with FRB-LpIA(W37V), FKBP-LAP1, and GFP as a co-transfection marker. Cells were treated with rapamycin, then labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 60 minutes. Coumarin signal is shown to the left in blue. GFP signal, which denotes transfected cells, is superimposed on the DIC image on the right. Bright coumarin signal is obtained in the presence of rapamycin and low background is observed in its absence for both cell lines, as observed for HEK cells. Scale bars, 10 μ m.

Kinetic analysis of proximity coumarin ligation

While kinetic analysis of coumarin ligation to LAP2 had been previously reported, our reaction of LpIA(W37V) with LAP1 has not been previously observed. Dr. Justin Cohen therefore measured the steady-state kinetic parameters for LpIA(W37V) ligation to FKBP-LAP1. The concentration of FKBP-LAP1 in these reactions was varied from 200 μ M to 1.5 mM, and the reaction products were analyzed by HPLC. The subsequent Michaelis-Menten curve-fitting

(Figure 4-17) revealed a K_m of LpIA(W37V) for LAP1 of $678 \pm 127 \mu\text{M}$, even higher than previously anticipated. This is excellent news, because even proteins overexpressed to the extent of low hundreds of micromolar in the cell should not be subject to false positive results from our proximity labeling method, because the reaction rate will still be in the bimolecular regime. Furthermore, the measured k_{cat} of $0.010 \pm 0.001 \text{ s}^{-1}$ is very similar to the previously reported k_{cat} of coumarin ligation onto LAP2 of 0.016 s^{-1} , [7] indicating that we have not sacrificed labeling sensitivity by using the low- K_m peptide. While this k_{cat} is not extremely fast, and is ten times slower than the rate of lipoic acid ligation by the wild-type enzyme, we still expect it to be sufficient for detection of PPIs with a half-life of approximately 1 minute, similar to the sensitivity of proximity biotinylation.

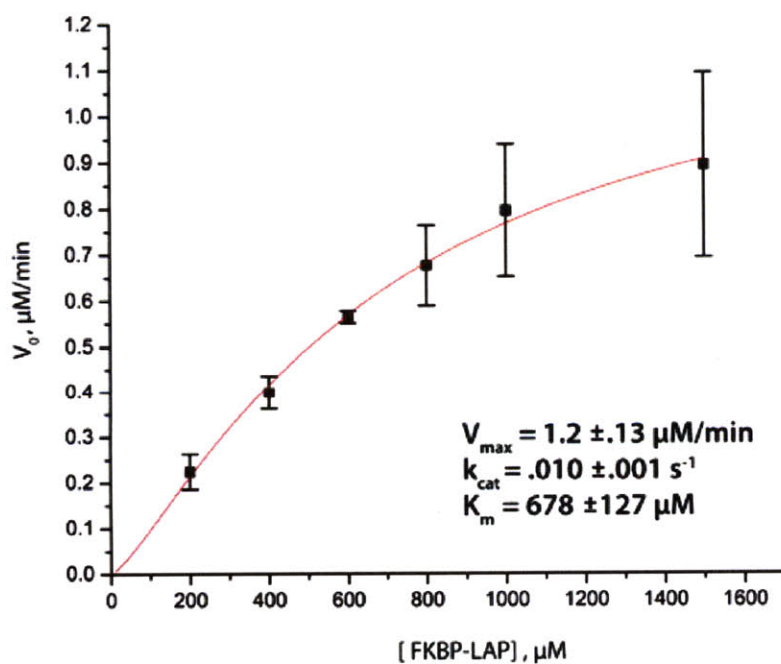


Figure 4-17. Michaelis-Menten curve for coumarin ligation to LAP1 by LpIA(W37V). The initial rate was measured in triplicate by HPLC and the error bars shown represent ± 1 s.d.

Rapamycin dose-response for proximity coumarin ligation

We wanted to determine if our method produces a quantitative response to the strength of a PPI. We therefore performed a rapamycin dose-response experiment to determine if we could quantitatively detect fractional association of the ternary complex at sub-saturating rapamycin concentrations. We used single-cell quantitative imaging for our assay. Analysis would have been simplified if we could perform coumarin labeling in cells, lyse the cells, then quantitate in-gel fluorescence by spot densitometry; however, no member of our laboratory has successfully detected coumarin signal by in-gel fluorescence after labeling in live cells. This is most likely a problem of detection sensitivity.

Cells were co-transfected with FRB-LplA(W37V) and FKBP-LAP1, then pre-treated with concentrations of rapamycin varying from 0.1 nM to 300 nM, extending well beyond both the upper and lower ends of the range of reported dissociation constants. Coumarin labeling was performed; the cells were then fixed and stained for both FRB and FKBP constructs (Figure 4-18). Visual inspection of the imaging data reveals that the coumarin labeling intensity generally increases as the rapamycin concentration increases, as expected, until reaching a maximal value.

In order to do the quantitation, transfected cells (as assessed by FKBP-LAP1 immunofluorescence) were selected as regions of interest. The intensities of coumarin, FKBP-LAP1, and FRB-LplA(W37V) were quantitated for each cell. The labeling yield is reported as coumarin intensity ratioed to FKBP-LAP1 intensity in order to normalize for the cell-to-cell variation in the expression levels of our constructs (a result of the transfection method). We then chose cells with FRB-LplA(W37V) intensities above a chosen cut-off value for our analysis.

The calculated single-cell coumarin labeling yield values were averaged for each concentration of rapamycin and plotted against the rapamycin concentration (Figure 4-19). Visual inspection of the data reveals that the labeling intensity is saturated at and above a concentration of 10 nM rapamycin, and that the labeling decays to our detection limit at and below 1 nM rapamycin. A single inflection point is present; data fitting to the Hill equation provides an apparent dissociation constant of 3.1 ± 0.6 nM, which agrees with the previously published value of 2.5 nM.[11] We therefore reproduce the previously reported single-site saturable binding of FRB to the FKBP-rapamycin complex within the range of reported dissociation constant values, demonstrating the quantitative response of proximity coumarin ligation.

[rapamycin], nM

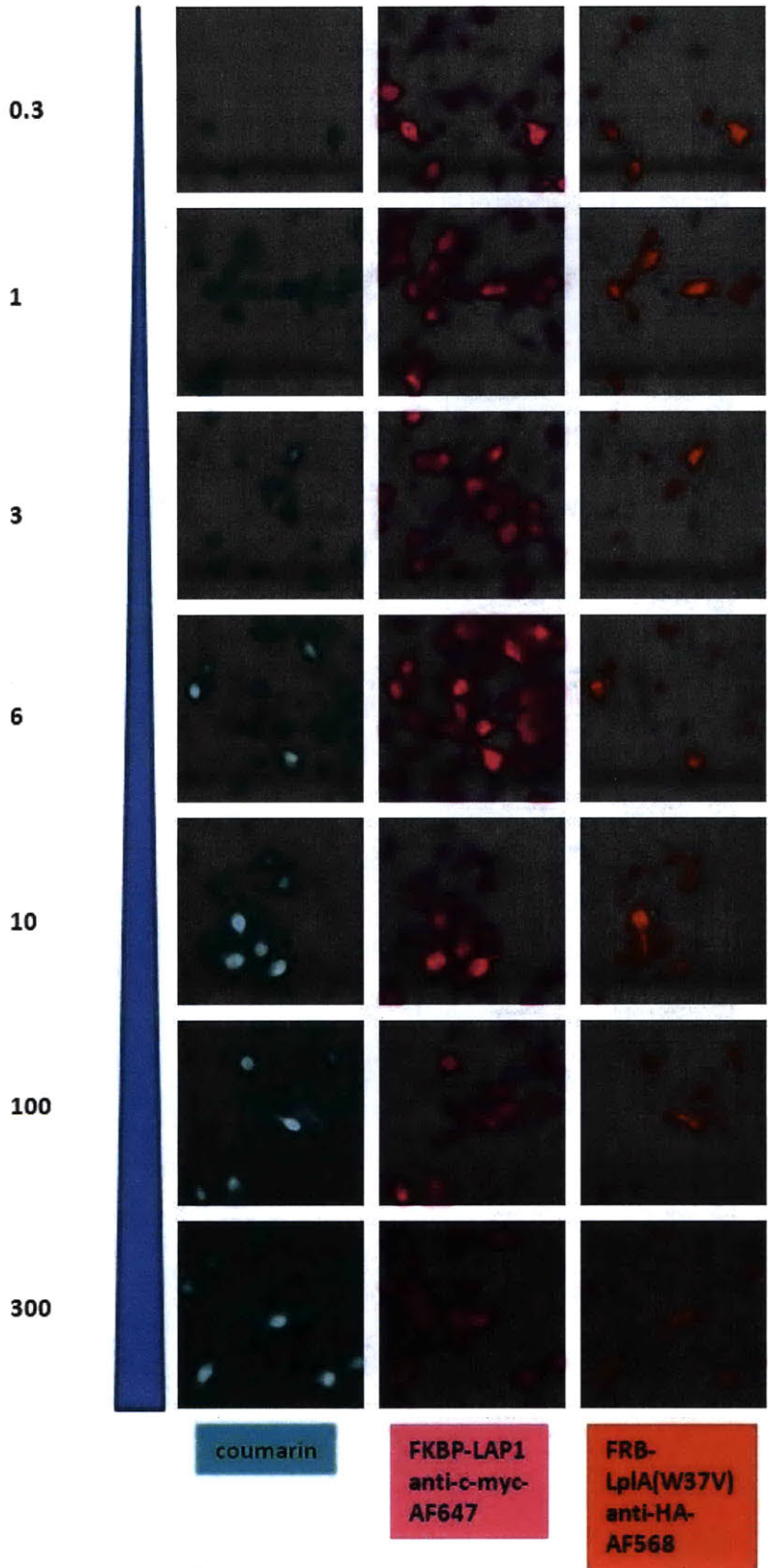


Figure 4-18. Rapamycin dose-response: representative images. HEK cells were transfected with FRB-LplA(W37V) and FKBP-LAP1. 24 hours after transfection, were treated with varying concentrations of rapamycin, then labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 30 minutes, then formaldehyde fixation and methanol permeabilization. Anti-c-myc staining was performed to visualize FKBP-LAP constructs, and anti-HA staining was performed to visualize FRB-LplA constructs. Coumarin signal is shown to the left in blue. Anti-c-myc staining is shown in pink in the center. Anti-HA staining is shown in red on the right. Representative images are shown.

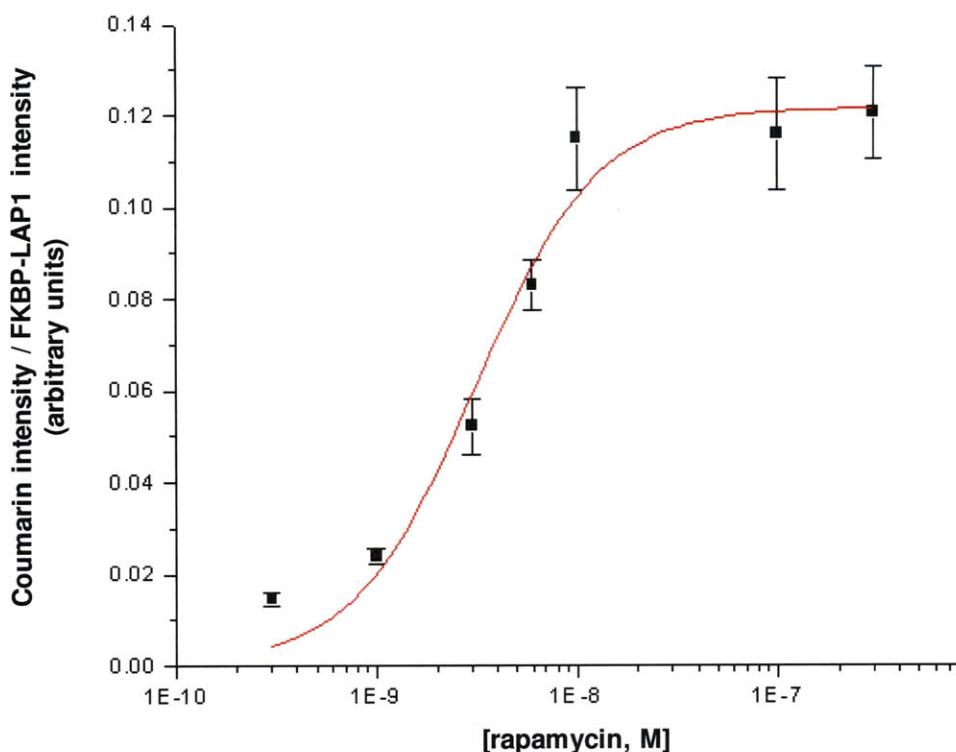


Figure 4-19. Rapamycin dose-response: quantitation. Each rapamycin concentration dataset shown in Figure 14 was utilized to generate a rapamycin dose-response curve. Coumarin labeling as a function of both FRB-LpIA(W37V) and FKBP-LAP1 expression levels was quantitated for individual cells as described in the methods. The calculated intensities were averaged, then plotted as a function of rapamycin concentration. Error bars, \pm standard error of the mean.

Wedge method to determine labeling yield and sensitivity of proximity coumarin ligation

We proceeded to estimate the sensitivity (that is, the dynamic range of protein concentrations over which we have a signal-to-background ratio greater than 2:1) and labeling yield of our method inside living cells. The wedge method has been developed to provide estimates of the concentrations of fluorescent molecules in cells.[21] In general, the intensity of a fluorophore inside a cell can be measured by imaging. The average thickness of common cell lines has been reported. Therefore, by measuring the intensity of a standard of that fluorophore at the same

thickness, the concentration of the intracellular molecule can be estimated. To do so, a wedge-shaped chamber of known size is constructed from glass coverslips, and the location in the focal plane at which it reaches desired thickness in the z-direction is calculated by triangulation from the known wedge dimensions; the intensity of the standard is measured at that point.

In order to make these measurements, we cloned a fusion of mCherry, a red fluorescent protein, to FKBP-LAP1. This provided a fluorescent readout of the FKBP-LAP1 concentration in the cells; of course, we have not compared the expression level of mCherry-FKBP-LAP1 to the untagged FKBP-LAP1 construct, so these labeling yields specifically and only apply to the yield attained for the interaction of these two constructs, FRB-LpIA(W37V) with mCherry-FKBP-LAP1. We expressed mCherry-FKBP-LAP1 in cells with FRB-LpIA(W37V), then pre-treated the cells with rapamycin to induce the interaction, or omitted rapamycin to quantitate background labeling in the absence of the interaction. We then performed coumarin labeling for either 10 or 20 minutes and imaged the cells live (Figure 4-20). For our standards, we utilized purified mCherry protein and a purified conjugate of coumarin to FKBP-LAP1, prepared as described in the Methods section. We used this standard rather than free coumarin because coumarin quantum yield has been observed to decrease upon ligation to target proteins (data not shown). After converting the cellular intensities to concentrations, coumarin concentration is plotted against mCherry-FKBP-LAP1 concentration (Figure 4-21).

Visual inspection of the data reveals the expected trends. The signal-to-background ratio is similar (on average approximately 5:1) at both labeling times. At very low mCherry-FKBP-LAP1 concentrations, the coumarin signal decreases below our detection limit. As the concentration

of mCherry-FKBP-LAP1 increases, coumarin signal increases essentially linearly. However, at very high mCherry-FKBP-LAP1 concentrations, the background coumarin signal also begins to increase. We expect that if we could access even higher protein concentrations that approach the K_m of LplA and LAP, the background would continue to increase to values equaling the signal. It is clear from this experiment that even at the very high intracellular concentrations of up to 200 μ M, our signal-to-background ratio is still very high, indicating that our method should not be prone to generating false positive results even for over-expressed proteins, as expected.

By calculating the ratio of average coumarin concentration to average mCherry-FKBP-LAP1 concentration in all analyzed cells, we estimate the labeling yield to be approximately 8% for a 10-minute coumarin incubation, and 14% after a 20-minute coumarin incubation. However, there is high cell-to-cell variation, probably due to variations in FRB-LplA(W37V) expression level, which we have not measured in this experiment (because a YFP fusion to FRB-LplA(W37V) significantly impaired coumarin labeling, data not shown). This variation tells us something important: if the two proteins of interest are expressed at similar levels in the cell, and if we are looking at a single enzymatic turnover during our labeling time (which we expect to be the case for stable PPIs), the labeling yield can depend on either the enzyme or the peptide concentration in individual cells, depending on which is limiting.

More importantly, we calculate the minimum concentration of mCherry-FKBP-LAP1 required in the cell to produce a measurable signal, that is, a signal-to-background ratio of 2:1. We fit linear trend lines to the 20-minute labeling data in the presence of rapamycin (signal) and the

absence of rapamycin (background), and from a comparison of their slopes we conclude that the signal-to-background ratio remains greater than 2 at mCherry-FKBP-LAP1 concentrations greater than 6 μM . Similar trendline fitting to the 10-minute data provided a signal-to-noise ratio of greater than 2 at mCherry-FKBP-LAP1 concentrations greater than 13 μM .

We note that even if a protein is expressed at total concentrations below our estimated detection limit, it may still be detectable using proximity coumarin ligation under certain conditions. For example, proteins targeted to specific organelles (or sub-cellular locations) and proteins that oligomerize may be present in sufficiently high local concentrations to increase the coumarin signal above the detection limit.

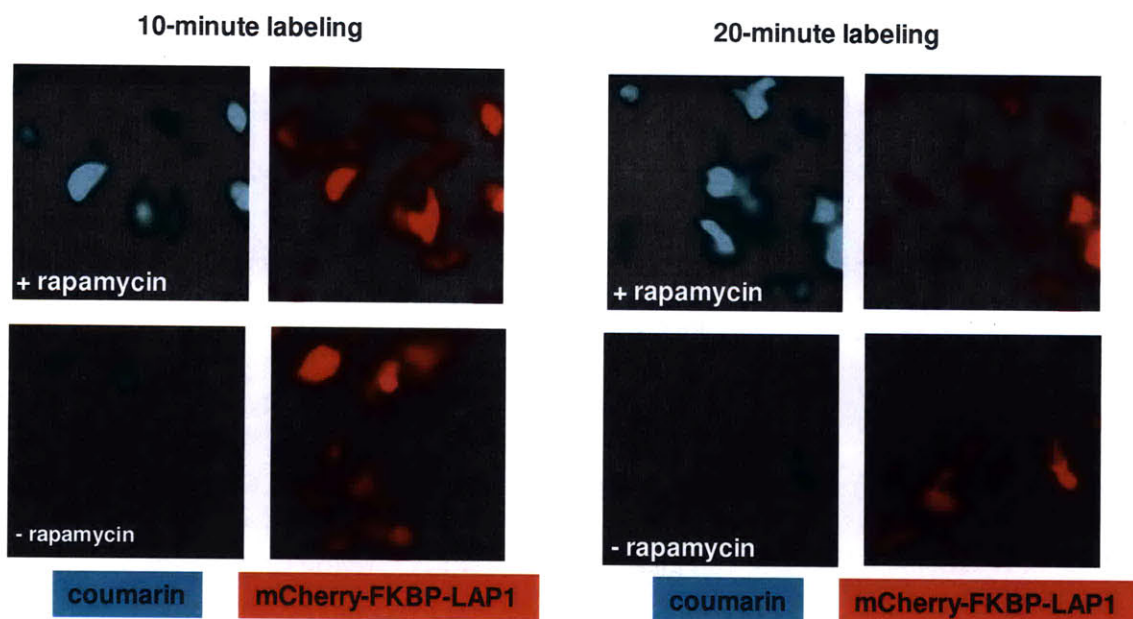


Figure 4-20. Determination of proximity coumarin ligation labeling yield: representative images. HEK cells expressing FRB-LpIA(W37V) and mCherry-FKBP-LAP1 were treated with rapamycin, or rapamycin was omitted to determine background. Cells were then incubated with 20 μM (AM)₂coumarin for either 10 or 20 minutes. Coumarin labeling is shown on the left in cyan, and mCherry-FKBP-LAP1 is shown on the right in red.

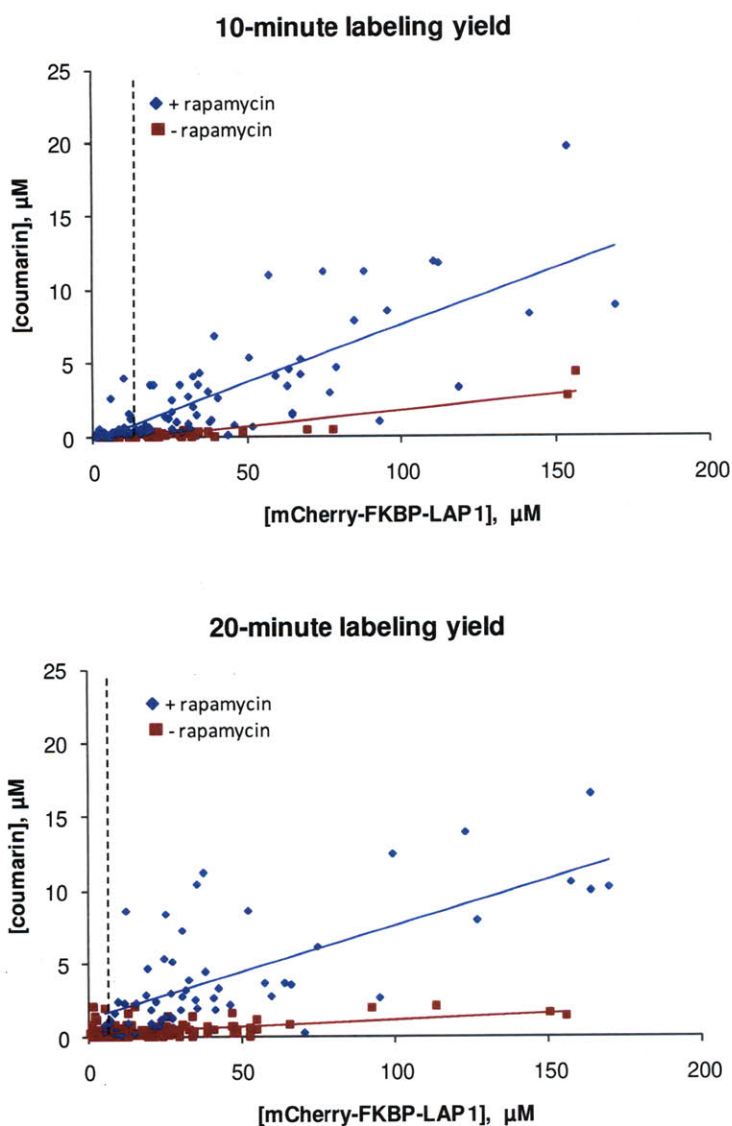


Figure 4-21. Determination of proximity coumarin ligation labeling yield. For the data set containing the representative images shown in Figure , average coumarin and mCherry intensities were quantitated. At least 45 cells were analyzed for each condition. Coumarin and mCherry intensities were converted into concentrations by comparison to intensities of purified coumarin-FKBP-LAP1 and mCherry standards of known concentration. Sensitivity and labeling yield were estimated from single-cell plots of measured concentrations. Linear fits to the data were generated and used to estimate labeling yield and sensitivity. A dashed line indicates the points on each plot at which a signal-to-noise ratio of 2:1 (the lower limit of detection) is achieved, 13 μM after 10 minutes' labeling and 6 μM after 20 minutes' labeling.

Comparing proximity coumarin ligation to bimolecular fluorescence complementation

Given that so many PPI detection methods already exist, we bear the burden of demonstrating that our method offers some advantage over comparable methods. In particular, bimolecular fluorescence complementation (BiFC) with the yellow fluorescent protein, YFP, has been utilized for fluorescence imaging of hundreds of PPIs inside living cells (see Chapter 1 for a complete discussion of this method as well as other reported PPI detection methods).[22] In

this method, YFP is divided into two non-fluorescent fragments, and each fragment is fused to one member of an interacting pair of proteins. If an interaction occurs, the fragments of YFP associate and fold to produce fluorescence (Figure 4-22). Because YFP BiFC has become the standard method in the field of cellular PPI imaging, it is the benchmark against which we must evaluate coumarin proximity ligation.

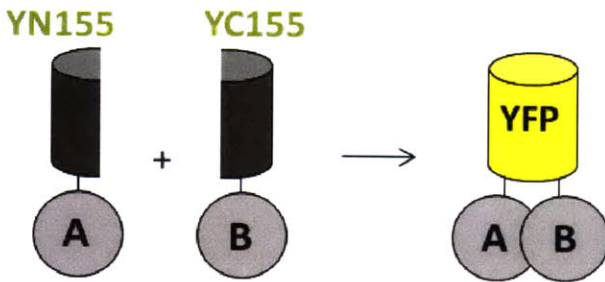


Figure 4-22. Principle of bimolecular fluorescence complementation (BiFC). Two non-fluorescent fragments of enhanced YFP (EYFP) are fused to interacting proteins A and B; upon interaction, the fragments of EYFP associate, fold, and oxidation by molecular oxygen generates the EYFP fluorophore, producing fluorescence.

We began with design of our split YFP reporter constructs. The literature converges on enhanced YFP (S65G, S72A, T203Y), or EYFP, as the most favorable construct for PPI detection, because its fragments offer extremely low background in the absence of a PPI, and good signal upon reconstitution.[23] Other enhanced YFPs, such as Venus, produce higher background.[24] We therefore proceeded with EYFP for our comparison.

For complementation experiments, EYFP is divided into a large N-terminal fragment, YN155, which comprises residues 1-154, and a smaller C-terminal fragment, YC155, comprises residues 155-238.[23] We created the fusions FRB-YN155 and FKBP-YC155. We hoped that these constructs would be sterically comparable to FRB-LpIA(W37V) and FKBP-LAP1, since the larger “piece” of the reporter in both cases (i.e., YN155 or LpIA(W37V)) is fused to the C-terminus of

FRB, and the smaller “piece” (i.e., YC155 or LAP1) is fused to the C-terminus of FKBP. The domain structures of the constructs used for our YFP BiFC comparison are depicted in Figure 4-23.

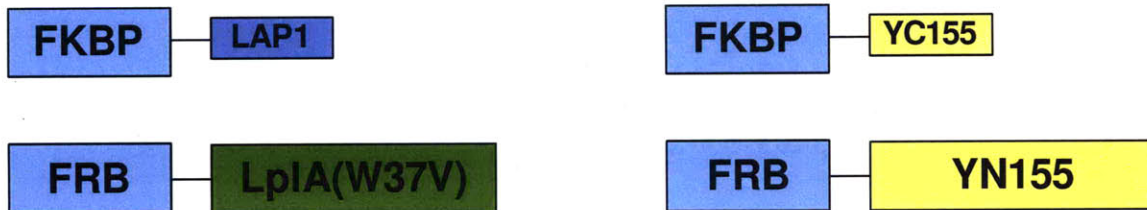


Figure 4-23. Domain structures of proximity coumarin ligation and YFP BiFC constructs. The “smaller piece” of each reporter, LAP1 for proximity coumarin ligation and YC155 (comprising residues 155-238 of EYFP) for BiFC, is fused to the C-terminus of FKBP. The “larger piece” of each reporter, LplA(W37V) for proximity coumarin ligation and YN155 (comprising residues 1-154 of EYFP) for BiFC, is fused to the C-terminus of FRB.

We first assayed YFP BiFC against proximity coumarin ligation at low expression levels of our constructs; we empirically determined that this condition is obtained 36 hours after transfection. We attempted to label both coumarin proximity ligation and YFP BiFC cells using similar experimental conditions, and fixed the cells in order to directly compare expression levels of these constructs by immunofluorescence. In order to achieve significant coumarin proximity labeling at these low protein concentrations, we extended our coumarin incubation time to 40 minutes; the good signal-to-noise obtained here demonstrate that for low-abundance proteins, longer labeling times can provide increased sensitivity, while the low concentrations mean the background remains low. YFP BiFC cells were grown either at 37°C after transfection, or 37°C followed by incubation at 30°C for 12 hours. This 30°C incubation has been reported to increase signal, possibly by enhancing fluorophore maturation.[25] Rapamycin was added to the BiFC cells during the final 12 hours before labeling to induce complex formation. All cells were then fixed and stained for the *c-myc* epitope tag on the FKBP

construct, and labeling intensity quantified as a function of FKBP expression level (Figure 4-24A). First, we note that at low expression levels, both coumarin proximity labeling and BiFC produce signal that linearly correlates with expression level of the constructs. Second, it is clear that the response of coumarin proximity labeling (with extended labeling times) and BiFC under normal growth conditions (that is, growth at 37°C) produces a nearly equivalent response in terms of signal-to-noise. Finally, incubation at 30°C clearly increases the signal-to-noise response of BiFC by approximately two-fold, as previously reported; under these conditions, BiFC is superior to coumarin proximity labeling. However we note that growth at lower temperatures may alter expression levels, aggregation, and the likelihood of an interaction occurring.

However, an important problem of BiFC becomes clear at high expression levels, which we empirically determined can be accessed 24 hours after transfection (Figure 4-24B). Under these conditions, coumarin proximity labeling can be performed for only 10 minutes, generating a bright response that increases linearly with expression level of the FRB construct. However, the YFP BiFC cells surprisingly show an inverse correlation of YFP signal with expression level. (We attempted to perform this experiment under the 30°C growth condition for the BiFC cells, but were unable to identify any immunostaining in cells in the absence of rapamycin despite several replicates; since we could not detect expression of the constructs, we were unable to assess the background under that condition and eliminated it from our analysis.)

Visual inspection of the primary imaging data makes clear the reason for this signal decay (Figure 4-25). Both in the presence of rapamycin and in its absence, FRB-YN155 displays

extensive aggregation. YFP fluorescence, which arises when the reporter folds properly, is diffuse and localized evenly across the cytoplasm; the aggregates and plaques visualized by immunofluorescence do not co-localize with the reconstituted YFP signal and therefore probably do not contribute to the signal. In contrast, FRB-LpIA(W37V) does not display aggregation under any condition. Therefore we conclude that the BiFC reporter fragments are not folded and tend to aggregate; at high expression levels, the aggregation competes with folding and signal generation. Therefore there is an expression level “ceiling” above which BiFC signal decreases. Furthermore, we hypothesize that the low background observed with BiFC in the absence of an interaction occurs because the fusion proteins are entirely misfolded and aggregated. The tendency of YFP BiFC reporters to aggregate has been previously reported, supporting our proposed mechanism.[23]

We conclude from these studies that, at low expression levels, the signal-to-noise ratios attainable by coumarin proximity labeling and BiFC are similar, though BiFC signal may be increased approximately two-fold through growth at decreased temperatures, and at high expression levels, our labeling method is superior because BiFC signal is limited at the upper range of protein concentrations by aggregation. We note that this is the case for the FRB and FKBP fusion constructs utilized in this study; the aggregation phenotype may be specific to these constructs, and may not occur for other proteins.

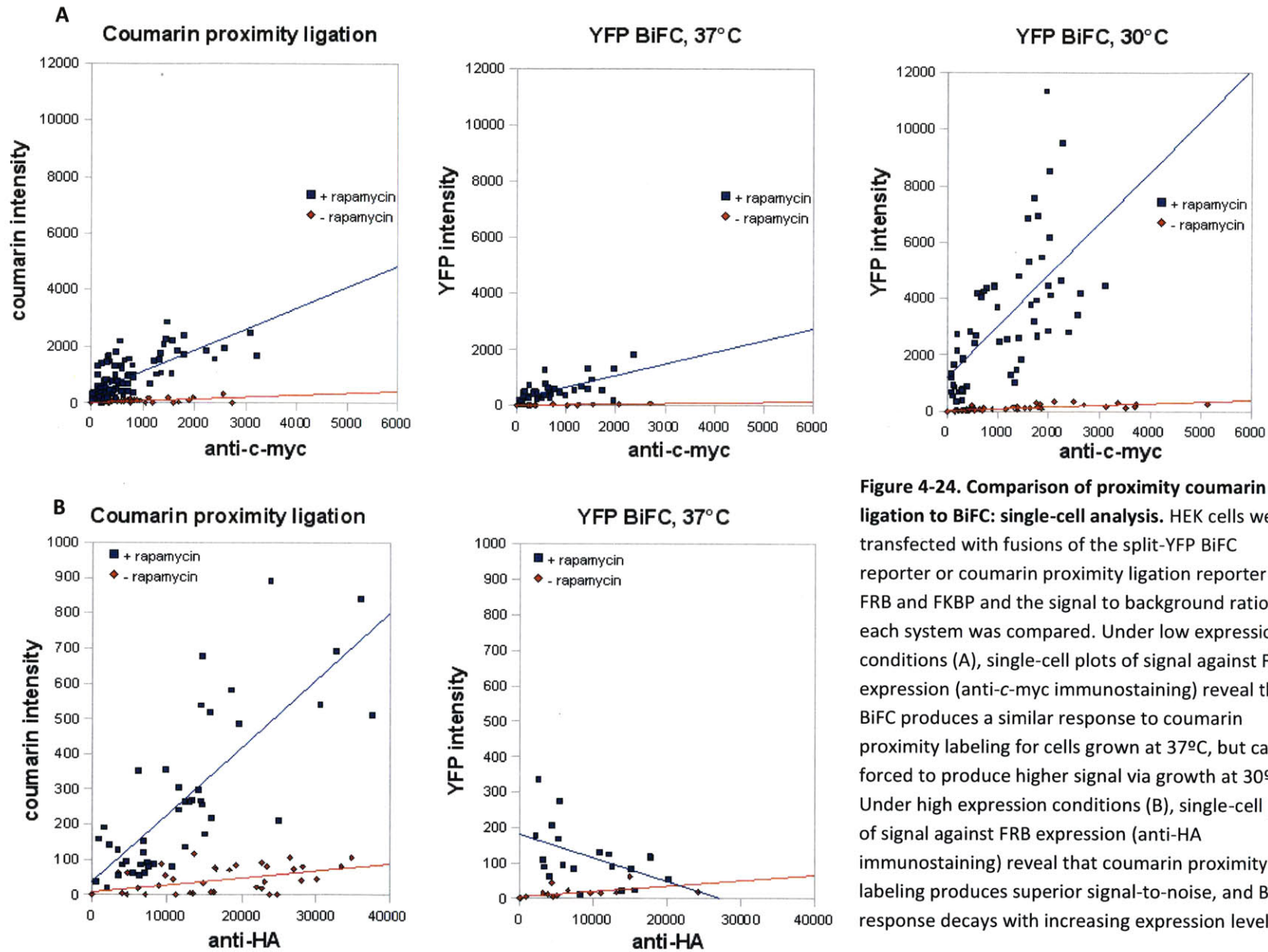


Figure 4-24. Comparison of proximity coumarin ligation to BiFC: single-cell analysis. HEK cells were transfected with fusions of the split-YFP BiFC reporter or coumarin proximity ligation reporter to FRB and FKBP and the signal to background ratio for each system was compared. Under low expression conditions (A), single-cell plots of signal against FKBP expression (anti-*c-myc* immunostaining) reveal that BiFC produces a similar response to coumarin proximity labeling for cells grown at 37°C, but can be forced to produce higher signal via growth at 30°C. Under high expression conditions (B), single-cell plots of signal against FRB expression (anti-HA immunostaining) reveal that coumarin proximity labeling produces superior signal-to-noise, and BiFC response decays with increasing expression level.

YFP BiFC, 37°C

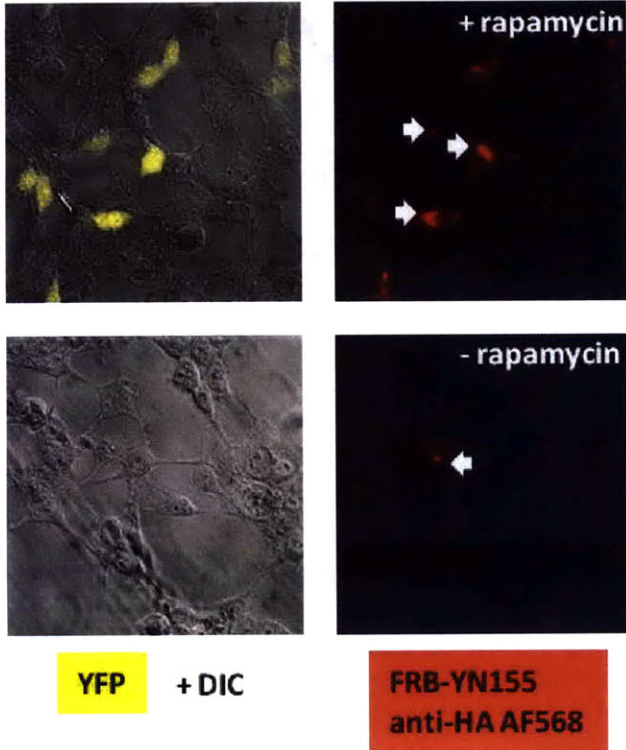
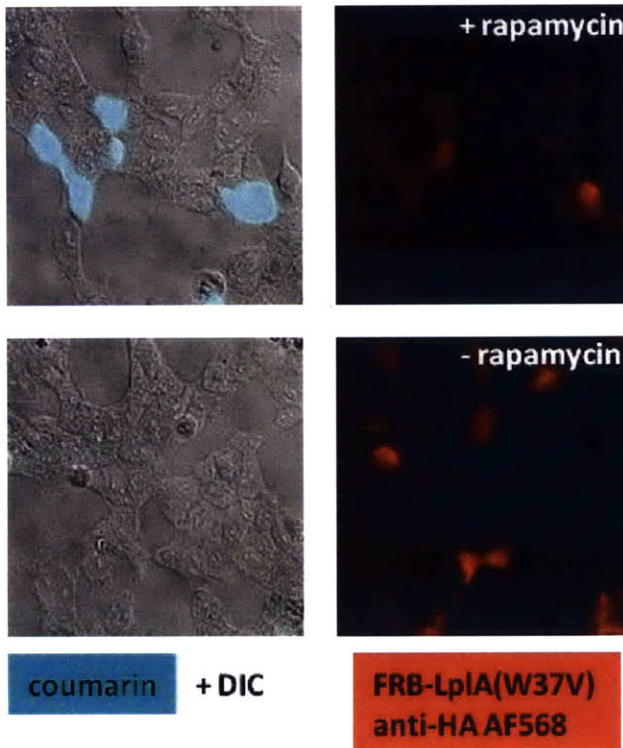


Figure 4-25. Comparison of proximity coumarin ligation to BiFC: representative images. The BiFC and coumarin proximity ligation reporters were compared under high expression conditions. These representative images are from the dataset used to generate the plots shown in Figure 4-20B. Top, YFP BiFC images. The YFP signal is shown in yellow on the left superimposed on the DIC image. FRB-YN155 immunofluorescence is shown on the right in red. White arrows indicate aggregates of protein. Bottom, coumarin proximity ligation images. Coumarin signal is shown in cyan, superimposed on the DIC image on the left. FRB-LplA(W37V) immunofluorescence is shown in red on the right. All immunofluorescence images are normalized on the same intensity scale.

Coumarin proximity ligation



In addition to the aggregation problem we have identified here, we note that BiFC has additional limitations relative to proximity coumarin ligation. First, maximal maturation of the EYFP fluorophore requires several hours after the protein folds, so detection of inducible PPIs is limited with BiFC.[24] Second, the folding of EYFP after the PPI occurs is demonstrably not reversible, so the complex is trapped by this method.[22] Because coumarin proximity labeling requires a relatively short labeling time for maximal signal, and does not trap the interaction partners, it should be the method of choice for investigating any system that would be perturbed by any of those factors, independent of aggregation concerns. Furthermore, we established that longer labeling times can increase the sensitivity of our detection of low-abundance proteins, as long as background is independently verified to remain low; we speculate that this result may apply to the detection of transient PPIs as well.

In conclusion, we utilized the rapamycin-dependent interaction of FRB and FKBP to validate our proximity coumarin ligation method, to optimize our reporter, and to demonstrate that the method is at best superior to and at worst comparable to the well-established YFP BiFC method. We then sought to apply our method to a biologically relevant PPI.

Part II: Application of proximity coumarin ligation to study the interaction of PSD-95 and neuroligin-1 in neurons

Introduction

While the rapamycin-dependent interaction of FRB and FKBP is an excellent system for developing and validating a method, it is also a simple system of easy-to-manipulate small proteins that are amenable to fusions. Furthermore our fusion construct design was guided by a crystal structure of the proteins in complex with one another, which is certainly not available for most PPIs. We therefore sought to apply our method to “real” proteins, intact proteins that have stringent requirements for the permissible insertion points for fusion constructs (that is, which have been demonstrated to be perturbed by fusions unless the inserted sequence is placed in a specific location in the protein of interest), and for which no crystal structure of the interaction complex exists. Furthermore, we wanted to push the envelope to see if we could accomplish labeling in delicate primary cells such as neurons. We found exactly the system we sought in the interaction of PSD-95 with neuroligin-1 in the postsynaptic density of excitatory neuronal synapses.

An interesting feature of PPIs involved in intracellular signaling is their modularity; that is, many PPIs are mediated by canonical interaction domains. Examples are SH2 domains, which bind phosphotyrosine residues,[26] and SH3 domains, which bind proline-containing consensus sequences.[27] For application of proximity coumarin ligation in a real system, we chose a PDZ domain-mediated PPI. PDZ domains bind a consensus motif at the C-terminus of their target proteins. Several classes of PDZ domains have been characterized based on their sequence

specificity. In particular, class I PDZ domains recognize the consensus sequence S/T-X-φ, where S/T is serine or threonine, X is any amino acid, and φ is a hydrophobic residue.[28] Interactions of cell-surface protein substituents of the post-synaptic terminus, including cell adhesion molecules, receptors, and ion channels, with PDZ domain-containing scaffolding proteins are responsible for anchoring and organizing these cell-surface molecules at the appropriate locations.[29]

Neuroigin-1 is a transmembrane cell adhesion protein present at the post-synaptic terminus of excitatory synapses.[30] It forms a stable *trans*-synaptic complex with neuexin, a pre-synaptic cell adhesion molecule.[31] Neuroigin family members are necessary and sufficient for synaptic differentiation,[32] and mutations in neuroigin-1 have been shown to be linked to autism spectrum disorders.[33] The interaction of neuroigin-1 with neuexin is thought to be important for synapse maturation and development.[34]

Another important interaction of neuroigin-1, with PSD-95, occurs inside the cell, and is mediated by a PDZ domain. Neuroigin-1 contains a C-terminal domain that extends into the cytosol and bears a canonical recognition sequence for class I PDZ domains, the tripeptide TRV, at its C-terminus.[35] PSD-95 is a cytosolic scaffolding protein present in the post-synaptic density of excitatory synapses that bears three PDZ domains. The first two PDZ domains of PSD-95 have been shown (*via* multiple techniques including yeast two-hybrid screens, bioluminescence resonant energy transfer, co-localization imaging and co-immunoprecipitation) to associate with the C-termini of potassium channels,[35] NMDA receptors,[36] and stargazin,[37] a scaffolding protein involved in recruitment of AMPA

receptors. The third PDZ domain of PSD-95 has been demonstrated by yeast two-hybrid screening, co-affinity purification, and co-localization imaging to specifically interact with the C-terminus of neuroligin-1.[35] The interaction of PSD-95 with neuroligin-1 has been demonstrated to be involved in excitatory synapse maturation through over-expression studies.[38, 39] A diagrammatic representation of the interaction of neuroligin-1 with PSD-95 is presented in Figure 4-26.

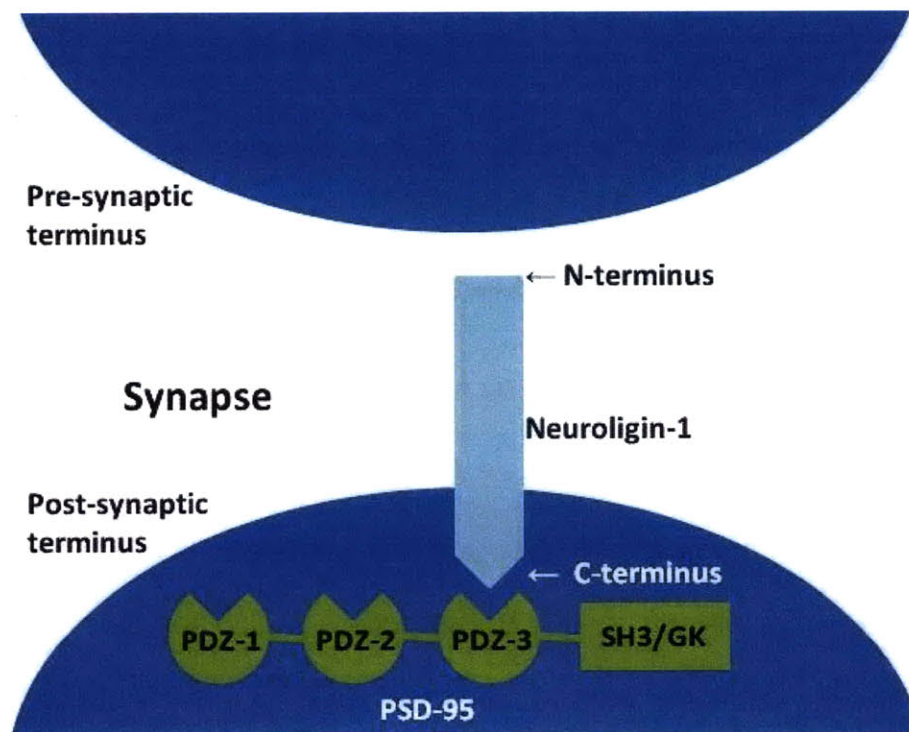


Figure 4-26. Diagrammatic representation of the interaction of neuroligin-1 and PSD-95 in the post-synaptic density. Neuroligin-1 is a transmembrane synaptic adhesion protein presented at the post-synaptic terminus. Its N-terminal domain participates in extracellular protein-protein interactions. The intracellular C-terminus of neuroligin-1 interacts with the third PDZ domain of PSD-95, a scaffolding protein of the post-synaptic density containing 3 PDZ domains, an SH3 domain, and a guanylate kinase domain. Figure adapted from Irie, M., et al., *Binding of neuroligins to PSD-95*. Science, 1997. **277**(5331): p. 1511-1515.

Many questions remain about the spatial and temporal characteristics of the PSD-95-neuroigin-1 interaction in neurons. It has been demonstrated that PSD-95 is responsible for localizing neuroigin-1 to excitatory synapses.[40] Two mechanisms are possible for this regulation. Either PSD-95 can traffic neuroigin-1 to the appropriate synapses, which would imply that PSD-95 and neuroigin-1 interactions occur at both synaptic and extra-synaptic sites; alternatively, neuroigin-1 can be present at excitatory synapses, and subsequent association of PSD-95 can stabilize its localization, which would imply that these proteins only interact at synapses. We suspect that our method can address this question, and others, in the future due to its excellent spatial and temporal resolution.

Results and discussion

Specific labeling of the interaction of PSD-95 and neuroigin-1 in HEK cells

Our first task was to design fusions of LpIA(W37V) and LAP1 to our proteins of interest. Since PDZ domains bind the C-terminus of their interaction partners, C-terminal fusions (such as GFP) to proteins bearing PDZ-binding motifs act as dominant negative mutations that block the interaction.[41] Internal GFP fusions to neuroigin-1 after threonine residue 776 that do not impair its interactions, trafficking, and localization had been previously reported.[42] We reasoned that the sensitivity of neuroigin-1 to fusions made LAP1, rather than LpIA, the better choice to fuse to it. Therefore, Dr. Justin Cohen inserted LAP1 after T776 of neuroigin-1 using overlap extension PCR. The LAP1 is, again, internal in the neuroigin-1 sequence, with the final 28 amino acids of neuroigin-1 following the LAP1 in the recombinant protein. This construct does afford proximity coumarin labeling (vide infra), but it was not obvious *a priori* that this

internal LAP1 peptide would be sterically accessible or kinetically competent for labeling. The construct was additionally constructed with an N-terminal AP tag to allow cell-surface biotinylation and streptavidin labeling, so the entire construct is called AP-neuroigin-1-LAP1 (Figure 4-27).

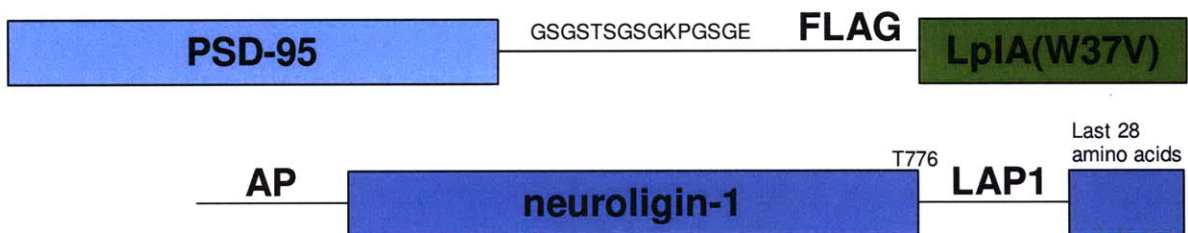


Figure 4-27. Domain structures of neuroigin-1 and PSD-95 constructs. PSD-95 was fused to a C-terminal 15-amino acid linker, followed by the FLAG epitope, then LpIA(W37V). Neuroigin-1 was N-terminally tagged with the biotin ligase acceptor peptide, AP, to facilitate streptavidin labeling. LAP1 peptide was inserted after threonine 776 of neuroigin-1, and the final 28 amino acids of neuroigin-1 follow the peptide and comprise the C-terminus of the construct.

We had access to a fusion of PSD-95 to mCherry wherein mCherry is placed at the C-terminus of PSD-95. This construct has previously been demonstrated to localize properly to excitatory synapses and to be correctly recruited to developing excitatory synapses (a functional assay), [43] as well as to interact normally with its binding partners in recruitment assays,[44] so we construed that this fusion geometry does not affect the localization and function of PSD-95. We therefore reasoned that placing LpIA(W37V) at the C-terminus of PSD-95 would probably afford a construct with similar behavior to PSD-95-mCherry.

However, the design of the PSD-95-LpIA(W37V) fusion presented an unanticipated experimental difficulty: LpIA enzymatic activity can be affected in the context of fusion to

another protein. Our first design utilized the 10-amino acid FLAG epitope tag as a linker between PSD-95 and LplA(W37V), and though this construct can be expressed in HEK cells, it did not label either the neuroligin-1-LAP1 construct or a positive control construct, LAP2-YFP, which is a kinetically competent substrate that expresses at a high level and is robustly labeled by enzymatically active LplA(W37V) constructs (data not shown). We subsequently re-designed our fusion construct with an additional flexible 15-amino acid linker between PSD-95 and FLAG-LplA(W37V) (Figure 4-23), theorizing that the longer linker could allow LplA to fold properly and/or access its substrate. The linker we chose has been successfully employed in the Ting lab in the past for construction of flexible multi-domain fluorescence resonant energy transfer (FRET) reporters.[45, 46] We confirmed enzymatic activity of this construct in an imaging assay by coumarin labeling of co-expressed LAP2-YFP in HEK cells (data not shown). Therefore, multiple validation steps are necessary when utilizing proximity coumarin ligation for detection of new PPIs; first, it must be determined that the proteins of interest are not perturbed by fusion to LplA and LAP; and second, it must be established that the LplA is enzymatically active and that the LAP tag is accessible using positive control experiments.

We then sought non-interacting mutants of our constructs to utilize as negative controls to determine if our labeling is real and interaction-dependent. We constructed a negative control construct wherein the three C-terminal amino acids of our AP-neuroligin-1-LAP1 construct are deleted. Recalling that the LAP1 is internal, it should be clear that these three amino acids, TRV, constitute the C-terminus of AP-neuroligin-1. This deletion is expected to eliminate the interaction, as these amino acids constitute the most strongly conserved part of the PDZ domain recognition sequence; furthermore the crystal structure of PSD-95 in complex with a

peptide derived from the protein CRIPT-1 shows that the final three amino acids of the bound peptide are deep in the binding pocket and are responsible for a majority of the interactions with the PDZ-domain.[47] In fact, previous reports bear this prediction out; this three-amino acid deletion has previously been reported to eliminate the interaction of neuroligin family members with PDZ domains in yeast two-hybrid assays.[35, 48] We designate this deletion mutant as AP-neuroligin-1-LAP1- Δ PDZ to indicate that the PDZ binding motif has been deleted.

No point mutant of PSD-95 that abrogates its interaction with neuroligin-1 has, to our knowledge, been reported. Close inspection of the PSD-95 crystal structure reveals why this might be the case; the target peptide interacts with many residues in the PDZ domain *via* both amino acid side chains and also amide bond interactions in the PSD-95 peptide backbone, so no single mutation (or obvious small number of mutations) would obviously abrogate the binding interaction.[47] Rather than deleting entire domains of PSD-95, as previously reported,[35] we simply chose to utilize LplA(W37V) as a non-interacting negative control.

We tested coumarin proximity labeling of these constructs in HEK cells. We began in heterologous cells to simplify the optimization process, since we already knew that coumarin labeling works well in these cells in general. Furthermore, this interaction was first characterized in heterologous cells, so we knew that PSD-95 and neuroligin-1 can interact in this context.[35]

When PSD-95-LplA(W37V) is co-expressed with AP-neuroligin-1-LAP1 in living cells, ten minutes of coumarin labeling affords bright coumarin labeling proximal to the plasma membrane, as expected for labeling of a membrane protein. When AP-neuroligin-1-LAP1 is replaced with the

non-interacting mutant AP-neurologin-1-LAP1- Δ PDZ, and also when PSD-95-LplA(W37V) is replaced with LplA(W37V), the interaction does not occur and labeling is abolished (Figure 4-28).

With longer labeling times (i.e., a coumarin incubation of 40 minutes) and high expression levels, we do observe non-specific labeling in both negative controls equivalent to the level of labeling in the presence of the interaction; we have never observed this level of background in the case of FRB and FKBP labeling, even after one hour of labeling (data not shown). This is probably a consequence of extremely high over-expression of these particular constructs in the heterologous context, and demonstrates that, while our method exhibits extremely good dynamic range, the need remains to carefully control expression levels when investigating PPIs, as well as to optimize labeling conditions for each new pair of proteins under investigation, confirming that, for the protein pair utilized and the labeling conditions employed, background is low.

We conclude that we can specifically label the interaction of neurologin-1 and PSD-95 in HEK cells. Our next task was to establish whether we could carry this labeling into neurons.

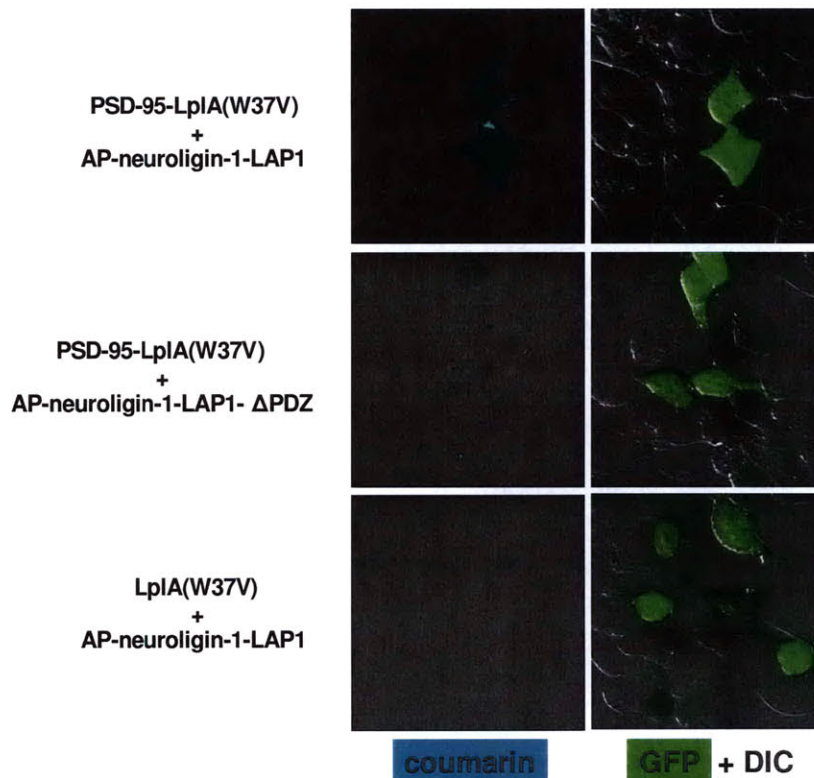


Figure 4-28. Imaging the interaction of PSD-95 and neuroigin-1 in living HEK cells. HEK cells were transfected with PSD-95-LpIA(W37V), AP-neuroigin-1-LAP1, and GFP as a co-transfection marker. Cells were labeled with (AM)₁ coumarin for 10 minutes, followed by wash-out of excess coumarin for 60 minutes. Coumarin signal is shown to the left in cyan. GFP signal, which denotes transfected cells, is superimposed on the DIC image on the right. Replacing AP-neuroigin-1-LAP1 with the non-interacting mutant AP-neuroigin-1-LAP1-ΔPDZ, or replacing PSD-95-LpIA(W37V) with LpIA(W37V) eliminates labeling, indicating that coumarin signal is interaction-dependent.

Coumarin proximity labeling of the interaction of PSD-95 and neuroigin-1 in neurons

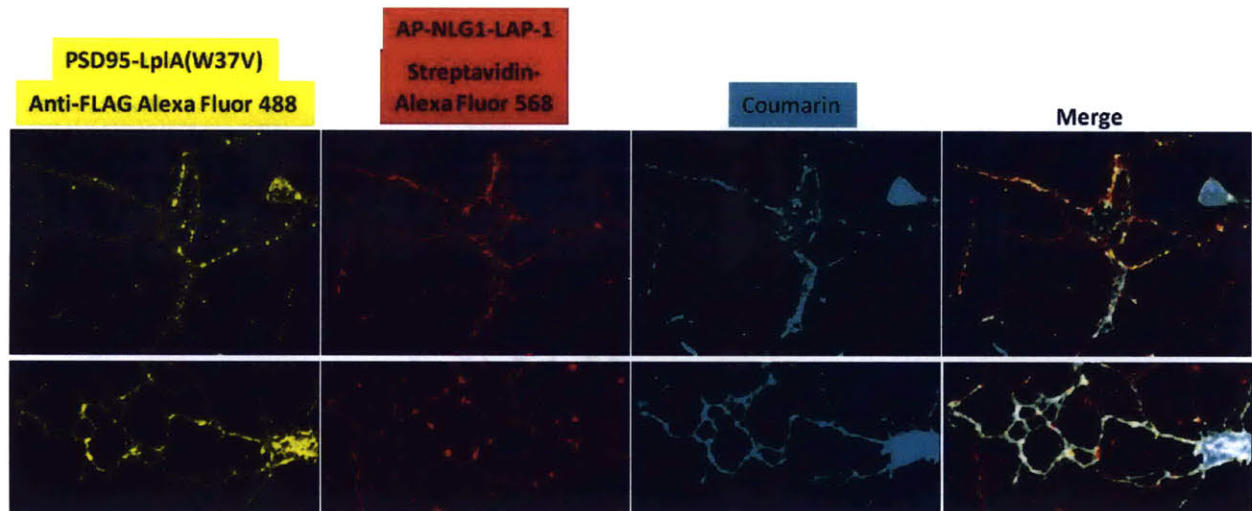
Working together with Dr. Amar Thyagarajan, we determined that PSD-95-LpIA(W37V) and AP-neuroigin-1-LAP can be expressed in cultured dissociated rat hippocampal neurons, with optimal expression of both constructs obtained 24 hours after co-transfection. We note that under our experimental conditions, we are purposely over-expressing our fusion constructs in order to attain good coumarin signal. PSD-95-LpIA(W37V) was observed to localize to puncta in neuronal processes, as expected for a synaptic protein, providing good evidence that the fusion

of PSD-95 to LplA(W37V) did not significantly perturb its trafficking and localization. Dr. Thyagarajan extended these experiments to include coumarin labeling (Figure 4-29). Neurons were transfected on DIV4 (5 days *in vitro*) with PSD-95-LplA(W37V), AP-neuroigin-1-LAP1, and BirA-ER, a previously reported construct[49] that localizes BirA to the endoplasmic reticulum and allows biotinylation of cell-surface proteins as they traverse the secretory pathway. Cells were incubated overnight with 10 μ M biotin to give biotinylation of AP-neuroigin-1-LAP1. On DIV5, the neurons were labeled with coumarin, then labeled with streptavidin-Alexa Fluor 568 conjugate to visualize AP-neuroigin-1-LAP1. After fixation with formaldehyde and methanol permeabilization, immunostaining against the FLAG epitope on PSD-95-LplA(W37V) was performed. The results (Figure 4-29A) show bright coumarin labeling both in cell bodies and neuronal processes, which co-localizes to a large extent with the PSD-95-LplA(W37V) immunofluorescence signal. If AP-neuroigin-1-LAP1 is replaced with the non-interacting mutant AP-neuroigin-1-LAP1- Δ PDZ (Figure 4-29B), the coumarin signal in the processes is eliminated. Low coumarin background remains in the cell body; this is probably because the local concentration of both proteins is very high in the cell body. Reassuringly, immunostaining and streptavidin staining demonstrate that both the wild-type and non-interacting mutant neuroigin constructs express at similar levels and have grossly similar localizations. This provides support for the idea that the non-interacting neuroigin-1 mutant does not abrogate labeling simply because it is not functional or mistrafficked, but rather that the elimination of coumarin signal is simply due to the lack of interaction with our PSD-95 fusion construct.

We therefore conclude that proximity coumarin labeling accurately reports on the interactions of real proteins, and furthermore that it is amenable to labeling delicate primary cells such as

neurons. We are currently characterizing the interaction of neuroligin-1 and PSD-95 in neurons further. We plan to determine if the locations of coumarin labeling coincide with genuine synaptic markers in order to confirm that our constructs localize, as expected, to excitatory synapses; furthermore we are investigating whether there is an interaction-dependent increase in the interaction of neuroligin-1 and PSD-95.

A



B

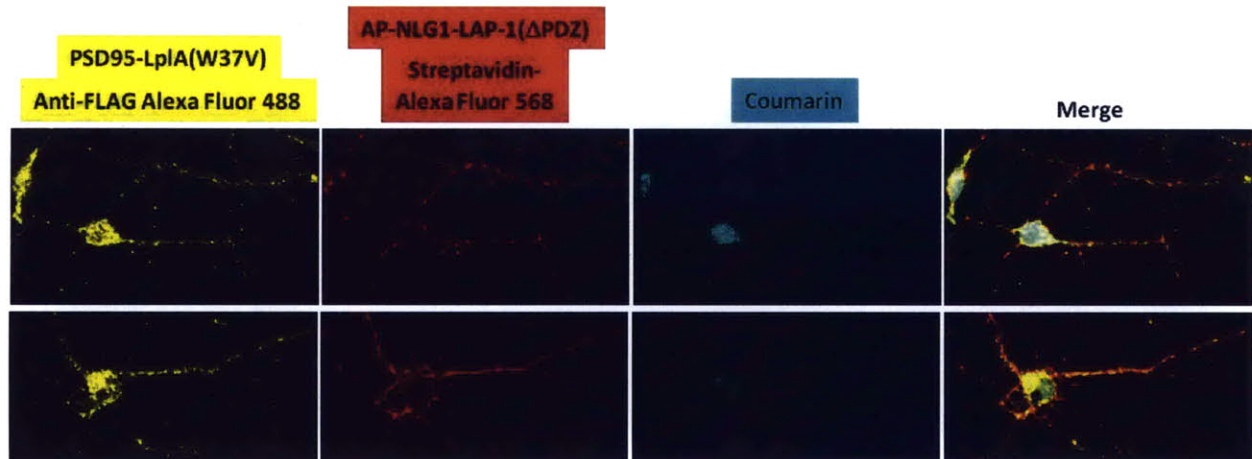


Figure 4-29. Imaging the interaction of PSD-95 and neuroligin-1 in rat hippocampal neurons. Neurons were transfected on DIV4 with PSD-95-LplA(W37V), AP-NLG1-LAP1, and BirA-ER. Neurons were incubated with 10 μ M biotin overnight to afford biotinylation of AP-neuroligin-LAP1. On DIV5, neurons were labeled with 20 μ M (AM)₂-coumarin for 15 minutes, washed for 45 minutes, then fixed with paraformaldehyde and permeabilized with methanol. Anti-FLAG epitope immunostaining against PSD-95-LplA(W37V) was performed to visualize enzyme localization and expression level. (A) Coumarin labeling is seen in neuronal processes and cell bodies. (B) Replacing AP-neuroligin-1-LAP1 with the non-interacting mutant neuroligin-1-LAP1- Δ PDZ eliminates coumarin labeling, except for low background in the cell body, probably due to high local concentrations.

Part III: Conclusion

In conclusion, we have developed new methods for detecting PPIs in mammalian cells. While proximity lipoylation does produce a good response in the presence of a PPI and low background labeling in the absence of an interaction, its utility for imaging assays is limited by the presence of endogenous lipoylated proteins in cells that cannot be removed. It may yet prove useful in gel-based assays and for proteomic analysis.

Proximity coumarin ligation, in contrast, has proved to be an effective method for imaging PPIs in living mammalian cells. In contrast to previously reported methods, it offers several distinct advantages. Unlike PCAs,[22] proximity coumarin ligation does not trap the interaction complex, leaving the interaction partners free to dissociate after labeling. Background is extremely low due to the poor affinity of LplA for LAP1, reducing false positive results. Unlike the traditional yeast two hybrid assay,[50] it is applicable in mammalian cells and works for membrane proteins (such as neuroligin-1). Furthermore, it compares favorably to BiFC in terms of signal-to-background ratio without being subject to aggregation.

Several fundamental limitations to this technology remain. First, the coumarin fluorophore is not yet fully optimal for biological imaging. The pKa of 7-hydroxycoumarin is 7.5, so at physiological pH only about half of the fluorophore is deprotonated.[51] Unfortunately, only the deprotonated form of 7-hydroxycoumarin is fluorescent, so the maximal signal we can attain is limited not only by labeling yield but also by coumarin ionization. Second, the 7-hydroxycoumarin photobleaches quickly. Third, significant cellular autofluorescence exists in blue wavelengths, and the short excitation light used for coumarin imaging cannot penetrate tissues. These factors limit the utility of coumarin labeling for biological imaging. In the future,

the photophysical properties of proximity ligation can be improved by engineering LplA to accept brighter coumarins, such as Pacific Blue, a fluorinated coumarin with a pKa of 3.7,[51] and red fluorophores, such as resorufin. These efforts are already underway in the Ting research group.

A second limitation to our sensitivity is the slow catalytic rate of LplA(W37V) of 0.010 s^{-1} . This rate is 10 to 30 times slower than the catalytic rates determined for lipoic acid and other engineered unnatural substrates for LplA.[7] This limits the sensitivity of the method and prevents us from using labeling times of less than about ten minutes. While *in vitro* selections could produce a faster LplA mutant, this will require significant engineering effort.

A third limitation is our total labeling time of approximately one hour. Most of this time is required for excess coumarin wash-out prior to imaging. A shorter labeling time, on the order of the lifetime of the PPI of interest, would be ideal, but it is difficult to envision a solution to this problem using enzyme-catalyzed labeling, unless a fluorogenic probe that fluoresces upon ligation could be developed.

A fourth problem is that, while our peptide is very small and therefore likely to minimally perturb its fusion partner, LplA is a large enzyme, at 42 kilodaltons, and could potentially cause misfolding or mistrafficking of proteins to which it is fused. There is no simple solution to this problem, so every new protein under investigation must be rigorously characterized to ensure that it folds, traffics, localizes, and functions properly upon fusion to LplA.

Another limitation is the low activity of LplA in the secretory pathway that we and other members of our laboratory have observed, which renders it useless for labeling in these

compartments and at the cell surface. This may be due to the oxidizing nature of these compartments; disulfide formation between surface cysteine residues of LpIA has been previously reported to inactivate the enzyme.[52] While proximity coumarin ligation has been demonstrated to work in the cytosol and nucleus, a truly general method requires engineering LpIA to be active in all subcellular and extracellular environments.

A final problem is that once the coumarin has been ligated to the LAP, it remains covalently bound, even if the PPI is no longer occurring. Therefore our coumarin signal does not directly correspond to a complex of interacting proteins, but rather reports on protein that underwent interaction during the labeling time. While this does provide useful information about the PPI, an ideal method would turn on during the interaction, and turn off upon dissociation, as is the case with fluorescence resonant energy transfer, or FRET. Unfortunately such a system is not easily achievable using a ligase enzyme-based reporter.

Despite these limitations, we expect that coumarin proximity ligation should prove generally useful for PPI detection due to its robust response to both model and real PPIs in both cell lines and primary cells. In the future, we will apply coumarin proximity ligation to analysis of the dynamics of the PSD-95-neuroigin-1 interaction during synapse development as well as the localization of this interaction complex in neurons. We additionally plan to extend live-cell proximity labeling to fluorophores with improved photophysical properties.

In the future, we hope to extend proximity-dependent coumarin labeling to trans-synaptic PPI labeling. Dr. Amar Thyagarajan has recently reported a new methodology called BLINC, or biotin labeling of intercellular contacts (A. Thyagarajan and A. Y. Ting, *Cell* 2010, in press). BLINC is an

extension of proximity biotinylation to the imaging of cell-surface PPIs in living cells. In this case, each cell-surface interaction partner is fused to either biotin ligase (BirA) or the BirA acceptor peptide, AP; when the proteins interact, in particular across cell-cell junctions, biotinylation occurs and streptavidin staining reports on the interaction. BLINC has been utilized to investigate the trans-synaptic interaction of neuroligin-1 and neurexin, which, as previously mentioned, is a synaptic adhesion complex important for synapse formation and maturation. However, this method is only applicable to dissociated cultures of neurons, because streptavidin is too large to diffusively penetrate tissue sections or tissues of living animals. Because our coumarin probe is small, we anticipate that proximity coumarin labeling with LpIA could address this permeability issue, as well as decreasing labeling times due to the single-step labeling protocol. Once LpIA variants that are active on the cell surface have been developed through current Ting lab engineering efforts, this extension of coumarin proximity labeling to “L-LINC,” or LpIA labeling of intercellular contacts, will be an exciting future direction.

Methods

Cloning and mutagenesis

Nucleotide sequences of all constructs utilized in this study are available at <http://stellar.mit.edu/S/project/tinglabreagents/r02/materials.html>. Constructs were prepared either by standard restriction cloning methods or QuikChange mutagenesis (Stratagene) as described by the manufacturer.

Mammalian cell culture

HEK, HeLa, and COS-7 cells were cultured in growth media, consisting of Dulbecco's modification of Eagle's medium (DMEM, Cellgro) supplemented with 10% fetal bovine serum (FBS, PAA Laboratories), 50 units/mL penicillin, and 50 µg/mL streptomycin (Cellgro). Cells were maintained at 37°C under an atmosphere of 5% CO₂ unless otherwise noted. For cellular imaging, cells were grown on glass coverslips. HeLa and COS-7 cells were grown directly on the glass substrate. HEK cells were grown on glass pre-treated with 50 µg/ml fibronectin (Millipore).

Fluorescence imaging

Cells were imaged in Dulbecco's phosphate buffered saline (DPBS) on glass coverslips. A Zeiss Axiovert 200M inverted microscope with a 40x oil-immersion objective was used for epifluorescence imaging. Coumarin (400/20 excitation, 425 dichroic, 435/30 emission), YFP/Alexa Fluor 488 (493/16 excitation, 506 dichroic, 525/30 emission), mCherry/Alexa Fluor 568 (570/20 excitation, 585 dichroic, 605/30 emission), Alexa Fluor 647 (630/20 excitation, 660

dichroic, 680/30 emission) and differential interference contrast (DIC) images were collected. For confocal imaging, we utilized a Zeiss AxioObserver inverted microscope with a 60x oil-immersion objective, outfitted with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), 561 (DPSS), and 640 nm (diode) lasers (all 50 mW). Coumarin (405 laser excitation, 445/40 emission), GFP/Alexa Fluor 488 (491 laser excitation, 528/38 emission), Alexa Fluor 568 (561 laser excitation, 617/73 emission), Alexa Fluor 647 (640 laser excitation, 700/75 emission), and DIC images were collected. All image analysis was with Slide Book software (Intelligent Imaging Innovations). Fluorophore intensities in each experiment were normalized to the same intensity ranges. Acquisition times ranged from 20 milliseconds to 5 seconds.

Protein expression and purification

E. coli BL21 cells transformed with pET expression plasmids for either FRB-LpIA or FKBP-LAP1 were grown in Luria broth supplemented with 100 µg/mL ampicillin at 37°C until reaching OD 0.6. Protein expression was induced with 420 µM IPTG at 30°C for 5 hours. Cells were harvested by centrifugation (6,000 x g, 10 minutes, 4°C) then lysed with B-PER (Thermo Scientific) containing 2.5 mM PMSF, protease inhibitor cocktail (Calbiochem), and DNase I (New England Biolabs). The lysate was cleared by centrifugation (17,700 x g, 15 minutes, 4°C). The hexahistidine-tagged proteins were purified by affinity chromatography on Ni-NTA agarose (Qiagen). Clarified lysate was loaded onto 1 mL nickel-nitrilotriacetic acid agarose resin (Qiagen) by gravity flow, then washed with binding buffer (50 mM Tris-HCl pH 7.8, 300 mM NaCl) followed by washing buffer (binding buffer + 30 mM imidazole). Fractions were analyzed by

SDS-PAGE, and fractions containing pure protein were dialyzed three times against 4 liters of phosphate buffered saline (PBS), pH 7.4. Protein concentration was determined by BCA assay (Thermo Scientific) with bovine serum albumin (BSA) as a standard. Purified proteins were aliquoted, flash frozen, and stored at -80°C.

In vitro proximity lipoylation

FRB-LplA , FKBP-LAP1, and rapamycin were combined at 10 μ M in PBS, pH 7.4, with 5 mM $MgCl_2$, and incubated at room temperature for 30 minutes to pre-induce the interaction. 1 mM lipoic acid was added, then reactions were initiated with 5 mM ATP. Identical reactions with rapamycin omitted served as negative controls. The reactions were allowed to proceed for either 1 minute or 4 minutes, then quenched with a final concentration of 45 mM EDTA pH 8.0.

In vitro reaction mixtures were combined with SDS-PAGE loading buffer, then boiled for 5 minutes to denature all proteins. 30 μ L of each reaction mixture were loaded per lane on a 16% acrylamide SDS-PAGE gel, electrophoresed, then analyzed by Western blotting, as described below.

Immunoblotting detection of proximity lipoylation in cells

COS-7 cells were grown to 50% confluency in a 24-well plate, then transfected with 600 ng FRB-LplA -pcDNA3 and 600 ng FKBP-LAP1 per well using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. For comparison of signal-to-noise, FKBP-LAP1-pcDNA3 was replaced with FKBP-LAP2-pcDNA3 or FKBP-E2p-pcDNA3. 24 hours after transfection, growth media was removed and fresh growth media containing 100 nM rapamycin was applied to the cells for one hour at 37°C. Rapamycin was omitted from parallel wells as a negative control. The

media was then removed, and pre-warmed DPBS containing 500 μ M lipoic acid was applied to the cells for one minute. The labeling solution was removed and the cells were immediately lysed (and the reaction therefore quenched) with direct application of SDS-PAGE loading buffer (40 μ L per well). All proteins were denatured by boiling for 5 minutes. 30 μ L of this material was loaded per well on a 14% acrylamide SDS-PAGE gel, electrophoresed, then analyzed by Western blotting, as described below.

Western blotting

Samples were run on SDS-PAGE gels, then transferred to nitrocellulose for 120 minutes at 500 mA. (Parallel reactions or wells were run on an identical SDS-PAGE gel, then stained with Coomassie brilliant blue, as loading controls.) After transfer, membranes were blocked with 3% BSA in tris-buffered saline with 0.05% Tween-20 (TBS-T) for 1 hour at room temperature. For lipoic acid detection, the membrane was treated with rabbit polyclonal anti-lipoic acid antibody (Calbiochem) at a 1:300 dilution in 3% BSA in TBS-T at room temperature for one hour, then washed three times for 5 minutes with TBS-T. The membrane was then incubated with goat anti-rabbit horseradish peroxidase conjugate (Bio-Rad) in 3% BSA in TBS-T at a 1:3000 dilution for one hour at room temperature, then again washed three times for 5 minutes with TBS-T. Chemiluminescence detection was performed with SuperSignal West Pico reagent (Pierce) for *in vitro* reactions, or SuperSignal West Femto reagent (Pierce) for cell lysates, and imaged on an Alpha Innotech Chemilmager 5500. Spot densitometry was performed using AlphaEase FC version 3.2.2 software (Alpha Innotech).

Immunofluorescence detection of proximity lipoylation in cells

COS-7 cells were grown to 50% confluency on glass coverslips, then transfected with 200 ng FRB-LpIA -pcDNA3 and 400 ng FKBP-LAP1-NLS-pcDNA3 per 0.95 cm² surface area using Lipofectamine 2000 according to the manufacturer's instructions. 24 hours after transfection, growth media was removed and fresh growth media containing 100 nM rapamycin was applied to the cells for one hour at 37°C. Rapamycin was omitted from parallel wells as a negative control. The media was then removed, and pre-warmed DPBS containing 500 µM lipoic acid was applied to the cells for one minute. The cells were washed once with ice-cold DPBS, then fixed with 3.7% paraformaldehyde in DPBS at 4°C for 10 minutes, then permeabilized with methanol at -20°C for 10 minutes. Fixed cells were washed with DPBS, then blocked overnight with blocking buffer (3% BSA in DPBS with 0.1% Tween-20). Lipoic acid and the c-myc epitope of the FKBP-LAP-NLS construct were detected with immunostaining. Rabbit anti-lipoic acid antibody and mouse anti-c-myc antibody (Calbiochem) were added together, both at a 1:300 dilution, in blocking buffer for one hour at room temperature. Cells were washed three times for five minutes each with DPBS. Cells were then treated with 1:1000 dilutions of goat anti-rabbit antibody conjugated to Alexa Fluor 568 (Invitrogen) and goat anti-mouse antibody conjugated to Alexa Fluor 657 (Invitrogen) in blocking buffer for one hour at room temperature. Cells were again washed three times for five minutes each with DPBS prior to imaging. Confocal images were acquired at 60x magnification.

Proximity coumarin ligation in the cytoplasm of live HEK cells

HEK cells were grown to 70% confluency on fibronectin-coated glass coverslips, then transfected with 400 ng FRB-LpIA(W37V)-pcDNA3, 400 ng FKBP-LAP1-pcDNA3, and 20 ng GFP as

a co-transfection marker per 0.95 cm² using Neofectin (Mid-Atlantic Biolabs) according to the manufacturer's instructions. The GFP expression plasmid was a kind gift from Dr. Joshua Sanes (Harvard). For the LAP2 peptide comparison, FKBP-LAP1-pcDNA3 was replaced with FKBP-LAP2-pcDNA3. For negative control experiments, FKBP-LAP1-pcDNA3 was replaced with FKBP-LAP1(K→A)-pcDNA3, or FRB-LpIA(W37V)-pcDNA3 was replaced with FRB-LpIA-pcDNA3. 24 hours after transfection, 100 nM rapamycin was added in growth media for 1 hour at 37°C, or omitted as a negative control. Growth media was then removed and the cells were labeled by applying 20 μM (AM)₂-coumarin in serum-free DMEM at 37°C for 10 minutes. Excess coumarin was washed out with three changes of fresh DMEM over 60 minutes at 37°C. Cells were imaged in DPBS. Confocal images were acquired at 60x magnification.

Proximity coumarin ligation in the nucleus of live cells

This experiment was conducted exactly as for the cytoplasmic labeling, except that FKBP-LAP1-pcDNA3 was replaced with FKBP-LAP1-NLS-pcDNA3, FKBP-LAP2-NLS-pcDNA3, or FKBP-LAP1(K→A)-NLS-pcDNA3, as appropriate. Confocal images were acquired at 60x magnification.

Proximity coumarin ligation in the cytoplasm of fixed cells

Cells were grown and transfected exactly as for live cytoplasmic labeling, except that after transfection, cells were grown at 30°C under an atmosphere of 5% CO₂ in order to control cell division and expression level. 24 hours after transfection, 100 nM rapamycin was added in growth media for 1 hour at 37°C, or was omitted as a negative control. Growth media was then removed and the cells were labeled by applying 20 μM (AM)₂-coumarin in serum-free DMEM at 37°C for 10 minutes. Excess coumarin was washed out with one application of fresh DMEM for

30 minutes, then cells were fixed with 3.7% paraformaldehyde in DPBS at 4°C for 10 minutes, then permeabilized with methanol at -20°C for 10 minutes. Fixed cells were washed with DPBS, then blocked for one hour at room temperature with blocking buffer. Cells were stained with a 1:1000 dilution of rabbit anti-HA antibody (Rockland Immunochemicals) in blocking buffer at 4°C overnight. Cells were then washed three times with DPBS for five minutes each. Cells were then treated serially for 1 hour at room temperature with the following antibodies, each diluted 1:1000 in blocking buffer, in the following order: goat anti-rabbit Alexa Fluor 647 conjugate, followed by mouse anti-c-myc, followed by goat anti-mouse Alexa Fluor 568 conjugate (Invitrogen). Three five-minute DPBS washes were applied between each antibody incubation step. After staining, cells were imaged in DPBS. Confocal images were acquired at 60x magnification.

Comparison of LAP1 17-mer and 22-mer, LplA(W37V) and LplA(W37I)

HEK cells were grown to 70% confluency on fibronectin-coated glass coverslips, then transfected with 600 ng FRB-LplA-pcDNA3 (either the W37I or W37V mutant) and FKBP-LAP1-NLS-pcDNA3 (either the 17-mer or 22-mer LAP) per 0.95 cm² using Lipofectamine 2000 according to the manufacturer's instructions. 36 hours after transfection, 100 nM rapamycin was added in growth media for 1 hour at 37°C, or was omitted as a negative control. Growth media was then removed and the cells were labeled by applying 20 μM (AM)₂-coumarin in serum-free DMEM at 37°C for 20 minutes. Excess coumarin was washed out with one application of fresh DMEM for 30 minutes, then cells were fixed with 3.7% paraformaldehyde in DPBS at 4°C for 10 minutes, then permeabilized with methanol at -20°C for 10 minutes. Fixed

cells were washed with DPBS, then blocked for one hour at room temperature with blocking buffer. Cells were then stained with a 1:1000 dilution of chicken anti-c-myc antibody (Aves) in blocking buffer at room temperature for one hour. Cells were then washed three times with DPBS for five minutes each, then stained with a 1:1000 dilution of goat anti-chicken Alexa Fluor 568 conjugate (Invitrogen) for one hour at room temperature. Cells were washed three times with DPBS for five minutes each prior to imaging in DPBS. Epifluorescence images were acquired at 40x magnification.

For quantitation, regions of interest (ROIs) were manually drawn on transfected nuclei by visually inspecting the anti-c-myc immunofluorescence images. Average intensities of coumarin and anti-c-myc immunofluorescence were computed. Background correction was applied by drawing a ROI on an untransfected cell in each field of view and subtracting these background intensities from all values generated from that particular field of view. Coumarin intensity was then plotted against anti-c-myc intensity for each cell.

Proximity coumarin ligation in other cell lines

COS-7 and HeLa cells were grown to 50% confluency on glass coverslips, then transfected with 400 ng FRB-LpIA(W37V)-pcDNA3, 400 ng FKBP-LAP1-pcDNA3, and 20 ng GFP (a co-transfection marker) per 0.95 cm² using Lipofectamine 2000 according to the manufacturer's instructions. Cells were then labeled and imaged exactly as described for cytoplasmic labeling of HEK cells. Confocal images were acquired at 60x magnification.

Kinetics

The Michaelis-Menten curve for determining the values of k_{cat} and K_{m} for coumarin ligation by LplA W37V onto FKBP-LAP was generated from HPLC experiments used to determine the initial reaction rate (V_0). The conditions used were as follows: 2 μM enzyme, 1 mM 7-OH coumarin, 2 mM magnesium acetate, and 25 mM sodium phosphate buffer, pH 7.2. The amount of FKBP-LAP was varied from 1.5 mM to 200 μM . Each initial rate was measured in triplicate and the error bars shown represent ± 1 s.d.

Rapamycin dose-response

HEK cells were grown to 70% confluency on fibronectin-coated glass coverslips, then transfected with 400 ng FRB-LplA(W37V)-pcDNA3 and 400 ng FKBP-LAP1-pcDNA3 per 0.95 cm^2 using Lipofectamine 2000 according to the manufacturer's instructions. 24 hours after transfection, concentrations of rapamycin ranging from 0.3 nM to 300 nM were added in growth media for 1 hour at 37°C. Growth media was then removed and the cells were labeled by applying 20 μM (AM)₂-coumarin in serum-free DMEM at 37°C for 10 minutes. Excess coumarin was washed out with one application of fresh DMEM for 30 minutes, then cells were fixed with 3.7% paraformaldehyde in DPBS at 4°C for 10 minutes, then permeabilized with methanol at -20°C for 10 minutes. Fixed cells were washed with DPBS, then blocked overnight in blocking buffer at 4°C. Cells were then immunostained serially with 1:1000 dilutions in blocking buffer of the following antibodies in the following order, for one hour each at room temperature: mouse anti-c-myc, goat anti-mouse Alexa Fluor 647 conjugate, rabbit anti-HA, goat anti-rabbit Alexa Fluor 568 conjugate. Three five-minute DPBS washes were applied

between each antibody incubation step. After staining, cells were imaged in DPBS.

Epifluorescence images were acquired at 40x magnification.

For quantiation, ROIs were manually drawn on transfected cells by visually inspecting the anti-c-myc immunofluorescence images. Average intensities of coumarin, anti-c-myc immunofluorescence, and anti-HA immunofluorescence were computed. Background correction was applied by drawing a ROI on an untransfected cell in each field of view and subtracting these background intensities from all values generated from that particular field of view. ROIs with anti-HA intensities greater than 3000 were kept for analysis, leaving at least 8 data points for each rapamycin concentration and as many as 25. The coumarin intensity was ratioed to the anti-c-myc intensity for each ROI, then multiplied by 100 for ease of data display; these values were averaged for each rapamycin concentration. Error is reported as standard error of the mean.

Wedge method for labeling yield estimation

Cellular labeling and imaging: HEK cells were grown to 70% confluency on fibronectin-coated glass coverslips, then transfected with either 800 ng each or 100 ng each of FRB-LpIA(W37V)-pcDNA3 and mCherry-FKBP-LAP1-pcDNA3 per 0.95 cm² using Lipofectamine 2000 according to the manufacturer's instructions. 36 hours after transfection, 100 nM rapamycin was added in growth media for 1 hour at 37°C, or omitted to assess background. Growth media was then removed and the cells were labeled by applying 20 μM (AM)₂-coumarin in serum-free DMEM at 37°C for 10 or 20 minutes. Excess coumarin was washed out with three changes of fresh DMEM over 60 minutes at 37°C. Cells were imaged in DPBS. For image analysis, ROIs were manually

drawn on transfected cells by manually inspecting the mCherry channel, and average coumarin and mCherry intensities were acquired for each ROI. Epifluorescence images were acquired at 40x magnification.

Preparation of mCherry standard: mCherry in pRSETB was a gift from Roger Tsien. This protein was expressed and purified from *E. coli* exactly as described above for FRB-LplA and FKBP-LAP1 except that concentration was measured by UV-visible absorbance using the reported extinction coefficient of mCherry of $72,000 \text{ M}^{-1}\text{cm}^{-1}$ at 587 nm.[53]

Preparation of FKBP-LAP1-coumarin conjugate standard: Coumarin was enzymatically conjugated to FKBP-LAP1 in an *in vitro* reaction containing 25 μM purified LplA(W37V), 1 mM FKBP-LAP1, 1 mM coumarin probe, 5 mM ATP, 1 mM $\text{Mg}(\text{OAc})_1$ in PBS. The reaction was incubated overnight at 30°C, then subjected to HPLC purification. The product peak was manually collected. HPLC eluate was dried down on a rotavap followed by lyophilization to dryness, and the resulting material was dissolved in 1 mL PBS. The final concentration was determined by BCA assay.

The identity of the product was confirmed by mass spectrometry. The collected HPLC eluate was injected onto an Applied Biosystems 200 QTRAP mass spectrometer. The flow rate was 10 $\mu\text{L}/\text{min}$ and detection was in positive ionization mode. The calculated mass of $17546 \pm 5 \text{ Da}$ agrees within the error of the instrument with the expected mass of 17540 Da.

For quantitation we use the wedge method. We constructed a wedge-shaped chamber from three glass coverslips. The wedge for our mCherry standard was 5 mm long and the wedge for our FKBP-LAP1-coumarin standard was 5.5 mm long. Each increased monotonically from 0 to

150 μm in the z dimension. The wedge was filled with approximately 10 μL standard solutions (40 μM mCherry, 1 μM FKBP-LAP1-coumarin), then imaged under conditions identical to those used for cellular imaging. We assumed that our cells are, on average, 5 μm thick, so calculated the region of the wedge that represented 5 μm in height and measured the intensity of mCherry and FKBP-LAP1-coumarin at those locations as a standard for the concentration of these molecules inside single cells. Because free coumarin is completely washed out of cells by our protocol, the ratio of coumarin intensity to mCherry intensity represents the labeling yield inside cells.

Comparison of YFP BiFC with proximity coumarin ligation

Low expression level comparison: For YFP BiFC analysis, HEK cells were grown to 70% confluency, then transfected with 1 μg each of FRB-YN155-pcDNA3 and FKBP-YC155-pcDNA3 per 0.95 cm^2 using Lipofectamine 2000 according to the manufacturer's instructions. One set of BiFC cells was grown at 37°C after transfection for 36 hours. Another set of BiFC cells was grown for 24 hours after transfection at 37°C, then moved to 30°C for 12 hours. 24 hours after transfection, 100 nM rapamycin was applied to all BiFC cells. 36 hours after transfection, cells were fixed with paraformaldehyde and methanol as previously described.

For coumarin proximity ligation, cells were transfected with 1 μg each of FRB-LpIA(W37V)-pcDNA3 and FKBP-LAP1-pcDNA3 per 0.95 cm^2 using Lipofectamine 2000 according to the manufacturer's instructions. 36 hours after transfection, cells were treated with rapamycin, labeled, and fixed as described for proximity coumarin ligation in the cytoplasm of fixed cells.

All cells were blocked for one hour at room temperature in blocking buffer, then stained with a 1:1000 dilution of chicken anti-c-myc antibody in blocking buffer at room temperature for one hour. Cells were then washed three times with DPBS for five minutes each, then stained with a 1:1000 dilution of goat anti-chicken Alexa Fluor 568 conjugate for one hour at room temperature. Cells were washed three times with DPBS for five minutes each prior to imaging in DPBS. Epifluorescence images were acquired at 40x magnification.

For image analysis, ROIs were manually drawn on the anti-c-myc immunofluorescence channel, and average coumarin and anti-c-myc intensities were acquired.

High expression level experiment: For YFP BiFC analysis, HEK cells were grown to 70% confluency, then transfected with 100 ng each of FRB-YN155-pcDNA3 and FKBP-YC155-pcDNA3 per 0.95 cm² using Lipofectamine 2000 according to the manufacturer's instructions. Cells were grown at 37°C for 24 hours after transfection in the presence of 100 nM rapamycin. Cells were then fixed with paraformaldehyde and methanol as described.

For coumarin proximity labeling, 70% confluent HEK cells were transfected with 100 ng each FRB-LpIA(W37V)-pcDNA3 and FKBP-LAP1-pcDNA3 per 0.95 cm² using Lipofectamine 2000 according to the manufacturer's instructions. 24 hours after transfection, cells were treated with rapamycin, coumarin labeled, and fixed as described for proximity coumarin ligation in the cytoplasm of fixed cells.

All cells were blocked with blocking buffer at room temperature for one hour. Cells were stained with rabbit anti-HA antibody at a 1:1000 dilution in blocking buffer at room temperature for one hour, washed three times with DPBS, then stained with goat anti-rabbit

Alexa Fluor 568 conjugate in blocking buffer at room temperature for one hour prior to washing and imaging. Epifluorescence images were acquired at 40x magnification.

For image analysis, transfected cells were identified in the anti-HA immunofluorescence channel and ROIs were drawn by hand; in some cases of intracellular aggregation, the ROI had to be drawn on the DIC image in order to define the shape of the entire cell. Average YFP and anti-HA intensities were acquired for each ROI.

Proximity coumarin labeling of the interaction of PSD-95 and neuroligin-1 in HEK cells

HEK cells were grown to 70% confluency on fibronectin-coated glass coverslips, then transfected with 100 ng PSD-95-LpIA(W37V)-pNICE, 500 ng neuroligin-1-LAP1-pNICE, and 20 ng GFP per cm² using Neofectin according to the manufacturer's instructions. For negative controls, neuroligin-1-LAP1-pNICE was replaced with an equal amount of neuroligin-1-LAP1-pNICE Δ PDZ-pNICE, or PSD-95-LpIA(W37V) -pNICE was replaced with 20 ng FLAG-LpIA(W37V)-pcDNA3. 24 hours after transfection, the cells were labeled by applying 20 μ M (AM)₂-coumarin in serum-free DMEM at 37°C for 10 minutes. Excess coumarin was washed out with three changes of fresh DMEM over 60 minutes at 37°C. Cells were imaged in DPBS. Confocal images were acquired at 60x magnification.

Proximity coumarin ligation in neurons

Hippocampal neurons were dissociated from E18 Sprague Dawley rat embryos and cultured as previously described[54, 55] and suspended in MEM supplemented with 10% FBS and 2% B-27 supplement. Neurons were plated onto poly-L-lysine coated glass coverslips (12 mm diameter)

at a density of 120,000 per well. 12 hours later media was replaced with NeuroBasal media supplemented with B-27 and 0.5 mM GlutaMax™ (Invitrogen). Neurons were co-transfected with 1 µg of AP-neuroigin-1-LAP1, 350 ng of PSD95-LpLA(W37V) and 450 ng of BirA-ER plasmids at days *in vitro* (DIV) 5 using Lipofectamine 2000 (Invitrogen) in 500 µl of MEM. After transfection for 4 hours, the MEM was replaced with the pre-conditioned media in which the cells were cultured initially with 10 µM biotin added to afford overnight labeling of AP-tagged protein. 1 µg of AP-neuroigin-1-LAP1-ΔPDZ was used as a negative control in place of AP-NLG1-LAP1. Coumarin labeling was performed 24 hours after transfection by incubating cells with 20 µM (AM)₂-coumarin in serum-free Modified Eagle's Medium (MEM) (Sigma) at 37°C for 15 minutes, followed by a 45-minute wash in MEM at 37°C. Biotinylated AP-neuroigin-1-LAP1 and AP-neuroigin-1-LAP1-ΔPDZ were visualized by staining with streptavidin-Alexa Fluor 568 conjugate in Tyrode's buffer containing 1% dialyzed bovine serum albumin for 5 minutes at room temperature. Cells were rinsed, then fixed with formaldehyde and methanol as described above. The FLAG epitope on the PSD-95-LpLA(W37V) construct was visualized by immunostaining with monoclonal anti-FLAG M2 antibody (Sigma) followed by goat anti-mouse Alexa Fluor 488 (Invitrogen).

References

1. Fernandez-Suarez, M., T.S. Chen, and A.Y. Ting, *Protein-protein interaction detection in vitro and in cells by proximity biotinylation*. J.Am.Chem.Soc., 2008. **130**(29): p. 9251-9253.
2. Beckett, D., E. Kovaleva, and P.J. Schatz, *A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation*. Protein Sci., 1999. **8**(4): p. 921-929.
3. Fernandez-Suarez, M., et al., *Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes*. Nat.Biotechnol., 2007. **25**(12): p. 1483-1487.
4. Puthenveetil, S., et al., *Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase*. J.Am.Chem.Soc., 2009. **131**(45): p. 16430-16438.
5. Konrad, T., et al., *alpha-Lipoic acid treatment decreases serum lactate and pyruvate concentrations and improves glucose effectiveness in lean and obese patients with type 2 diabetes*. Diabetes Care, 1999. **22**(2): p. 280-7.
6. Zhao, X., J.R. Miller, and J.E. Cronan, *The reaction of LipB, the octanoyl-[acyl carrier protein]:protein N-octanoyltransferase of lipoic acid synthesis, proceeds through an acyl-enzyme intermediate*. Biochemistry, 2005. **44**(50): p. 16737-46.
7. Uttamapinant, C., et al., *A fluorophore ligase for site-specific protein labeling inside living cells*. Proc Natl Acad Sci U S A, 2010. **107**(24): p. 10914-9.
8. Tsien, R.Y., *A non-disruptive technique for loading calcium buffers and indicators into cells*. Nature, 1981. **290**(5806): p. 527-8.
9. Roe, M.W., J.J. Lemasters, and B. Herman, *Assessment of Fura-2 for measurements of cytosolic free calcium*. Cell Calcium, 1990. **11**(2-3): p. 63-73.
10. Vezina, C., A. Kudelski, and S.N. Sehgal, *Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle*. J Antibiot (Tokyo), 1975. **28**(10): p. 721-6.
11. Chen, J., et al., *Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12-rapamycin-associated protein and characterization of a critical serine residue*. Proc Natl Acad Sci U S A, 1995. **92**(11): p. 4947-51.
12. Banaszynski, L.A., C.W. Liu, and T.J. Wandless, *Characterization of the FKBP.rapamycin.FRB ternary complex*. J.Am.Chem.Soc., 2005. **127**(13): p. 4715-4721.
13. Choi, J., et al., *Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP*. Science, 1996. **273**(5272): p. 239-42.
14. Baruah, H., et al., *An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking*. Angew.Chem.Int.Ed Engl., 2008. **47**(37): p. 7018-7021.
15. Choi-Rhee, E., H. Schulman, and J.E. Cronan, *Promiscuous protein biotinylation by Escherichia coli biotin protein ligase*. Protein Sci., 2004. **13**(11): p. 3043-3050.
16. Cronan, J.E., *Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase*. J.Nutr.Biochem., 2005. **16**(7): p. 416-418.
17. Streaker, E.D. and D. Beckett, *Nonenzymatic biotinylation of a biotin carboxyl carrier protein: unusual reactivity of the physiological target lysine*. Protein Sci., 2006. **15**(8): p. 1928-1935.
18. Kwon, K. and D. Beckett, *Function of a conserved sequence motif in biotin holoenzyme synthetases*. Protein Sci., 2000. **9**(8): p. 1530-1539.
19. Takaishi, N., et al., *Transepithelial transport of alpha-lipoic acid across human intestinal Caco-2 cell monolayers*. J Agric Food Chem, 2007. **55**(13): p. 5253-9.
20. Gore, J., C. Hoinard, and P. Maingault, *Biotin uptake by isolated rat intestinal cells*. Biochim Biophys Acta, 1986. **856**(2): p. 357-61.
21. Adams, S.R., et al., *New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications*. J Am Chem Soc, 2002. **124**(21): p. 6063-76.

22. Kerppola, T.K., *Visualization of molecular interactions by fluorescence complementation*. Nature Reviews Molecular Cell Biology, 2006. **7**(6): p. 449-456.
23. Hu, C.D., Y. Chinenov, and T.K. Kerppola, *Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation*. Mol.Cell, 2002. **9**(4): p. 789-798.
24. Robida, A.M. and T.K. Kerppola, *Bimolecular fluorescence complementation analysis of inducible protein interactions: effects of factors affecting protein folding on fluorescent protein fragment association*. J.Mol.Biol., 2009. **394**(3): p. 391-409.
25. Kerppola, T.K., *Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells*. Nat Protoc, 2006. **1**(3): p. 1278-86.
26. Filippakopoulos, P., S. Muller, and S. Knapp, *SH2 domains: modulators of nonreceptor tyrosine kinase activity*. Curr Opin Struct Biol, 2009. **19**(6): p. 643-9.
27. Li, S.S., *Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction*. Biochem J, 2005. **390**(Pt 3): p. 641-53.
28. Stiffler, M.A., et al., *PDZ domain binding selectivity is optimized across the mouse proteome*. Science, 2007. **317**(5836): p. 364-9.
29. Yamagata, M., J.R. Sanes, and J.A. Weiner, *Synaptic adhesion molecules*. Curr.Opin.Cell Biol., 2003. **15**(5): p. 621-632.
30. Song, J.Y., et al., *Neuroigin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses*. Proc.Natl.Acad.Sci.U.S.A, 1999. **96**(3): p. 1100-1105.
31. Ichtchenko, K., et al., *Neuroigin 1: a splice site-specific ligand for beta-neurexins*. Cell, 1995. **81**(3): p. 435-443.
32. Chih, B., H. Engelman, and P. Scheiffele, *Control of excitatory and inhibitory synapse formation by neuroligins*. Science, 2005. **307**(5713): p. 1324-1328.
33. Chih, B., et al., *Disorder-associated mutations lead to functional inactivation of neuroligins*. Hum.Mol.Genet., 2004. **13**(14): p. 1471-1477.
34. Lise, M.F. and A. El Husseini, *The neuroligin and neurexin families: from structure to function at the synapse*. Cell Mol.Life Sci., 2006. **63**(16): p. 1833-1849.
35. Irie, M., et al., *Binding of neuroligins to PSD-95*. Science, 1997. **277**(5331): p. 1511-1515.
36. Gottschalk, M., et al., *Detecting protein-protein interactions in living cells: development of a bioluminescence resonance energy transfer assay to evaluate the PSD-95/NMDA receptor interaction*. Neurochem.Res., 2009. **34**(10): p. 1729-1737.
37. Chen, L., et al., *Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms*. Nature, 2000. **408**(6815): p. 936-943.
38. El Husseini, A.E., et al., *PSD-95 involvement in maturation of excitatory synapses*. Science, 2000. **290**(5495): p. 1364-1368.
39. Levinson, J.N., et al., *Neuroligins mediate excitatory and inhibitory synapse formation: involvement of PSD-95 and neurexin-1beta in neuroligin-induced synaptic specificity*. J.Biol.Chem., 2005. **280**(17): p. 17312-17319.
40. Prange, O., et al., *A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin*. Proc.Natl.Acad.Sci.U.S.A, 2004. **101**(38): p. 13915-13920.
41. Dresbach, T., et al., *Synaptic targeting of neuroligin is independent of neurexin and SAP90/PSD95 binding*. Molecular and Cellular Neuroscience, 2004. **27**(3): p. 227-235.
42. Sara, Y., et al., *Selective capability of SynCAM and neuroligin for functional synapse assembly*. J Neurosci, 2005. **25**(1): p. 260-70.

43. Heine, M., et al., *Activity-independent and subunit-specific recruitment of functional AMPA receptors at neurexin/neuroigin contacts*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(52): p. 20947-20952.
44. de Wit, J., et al., *LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation*. Neuron, 2009. **64**(6): p. 799-806.
45. Lin, C.W., C.Y. Jao, and A.Y. Ting, *Genetically encoded fluorescent reporters of histone methylation in living cells*. J Am Chem Soc, 2004. **126**(19): p. 5982-3.
46. Lin, C.W. and A.Y. Ting, *A genetically encoded fluorescent reporter of histone phosphorylation in living cells*. Angew Chem Int Ed Engl, 2004. **43**(22): p. 2940-3.
47. Doyle, D.A., et al., *Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ*. Cell, 1996. **85**(7): p. 1067-76.
48. Pouloupoulos, A., et al., *Neuroigin 2 drives postsynaptic assembly at perisomatic inhibitory synapses through gephyrin and collybistin*. Neuron, 2009. **63**(5): p. 628-642.
49. Howarth, M. and A.Y. Ting, *Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin*. Nat.Protoc., 2008. **3**(3): p. 534-545.
50. Luban, J. and S.P. Goff, *The yeast two-hybrid system for studying protein-protein interactions*. Curr Opin Biotechnol, 1995. **6**(1): p. 59-64.
51. Sun, W.C., K.R. Gee, and R.P. Haugland, *Synthesis of novel fluorinated coumarins: excellent UV-light excitable fluorescent dyes*. Bioorg Med Chem Lett, 1998. **8**(22): p. 3107-10.
52. Green, D.E., et al., *Purification and properties of the lipoate protein ligase of Escherichia coli*. Biochem.J., 1995. **309** (Pt 3): p. 853-862.
53. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein*. Nat Biotechnol, 2004. **22**(12): p. 1567-72.
54. Banker, G. and K. Goslin, *Culturing nerve cells*. 2nd ed. 1998, Cambridge, Mass.: MIT Press. xii, 666 p., 11 p. of plates.
55. Gerrow, K., et al., *A preformed complex of postsynaptic proteins is involved in excitatory synapse development*. Neuron, 2006. **49**(4): p. 547-562.

Sarah A. Slavoff

Phone 617-800-6562

sarah.slavoff@gmail.com

Education

- 2005-2010** **Ph.D., Chemistry**
Massachusetts Institute of Technology Cambridge, MA
- 2001-2005** **B.S., Biochemistry, Summa Cum Laude**
University of Maryland, College Park College Park, MD

Research Experience

- 2005-2010** **Graduate Research Assistant, Massachusetts Institute of Technology**
Research Supervisor: Alice Y. Ting
Developed new reporters for protein-protein interaction imaging in living mammalian cells and a new method for site-specific protein labeling *in vitro* using bio-orthogonal chemistry. Developed expertise in HPLC, ESI-MS, protein biochemistry, mammalian cell culture, fluorescence imaging, molecular cloning and organic chemistry.
- 2004-2005** **Undergraduate Research Assistant, University of Maryland Department of Chemistry and Biochemistry**
Research Supervisor: Neil V. Blough
Performed spectroscopic and chromatographic analysis of the reactions of thiyl radicals and airborne particulate matter with fluorogenic radical sensors. Established and maintained mammalian cell cultures.
- 2003-2004** **Research Intern, Joint Institute for Food Safety and Applied Nutrition of the University of Maryland and Food and Drug Administration**
Participated in long-term research study developing analytical methods to detect and quantify hepatotoxic compounds from *Teucrium* species present as adulterants in dietary supplements. Prepared plant extracts and performed HPLC quantitation of diterpenoid compounds.

Fellowships and Awards

- 2009** **David A. Johnson Graduate Student Summer Fellowship**
- 2008-2009** **American Chemical Society Division of Organic Chemistry Graduate Fellowship**
- 2006-2007** **Whitaker Health Sciences Fund Graduate Fellowship**
- 2006** **Outstanding Teaching Assistant Award, Massachusetts Institute of Technology**
- 2005** **Phi Beta Kappa**
- 2001-2005** **Banneker/Key Scholarship, University of Maryland, College Park**

Publications

S. A. Slavoff, A. Thyagarajan, J. D. Cohen, A. Y. Ting. "Protein-protein interaction imaging in living cells and neurons via proximity-dependent fluorophore ligation." Manuscript in preparation.

S. A. Slavoff, I. Chen, Y.-A. Choi, A. Y. Ting. "Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species." *Journal of the American Chemical Society*, **2008**, *130*, 1160-1162.

P. R. Sundaresan, S. A. Slavoff, E. Grundel, K. D. White, J. I. Rader. "Isolation and characterisation of selected germander diterpenoids from authenticated *Teucrium chamaedrys* and *T. canadense* by HPLC, HPLC-MS and NMR." *Phytochemical Analysis*, **2006**, *17*, 243-250.

Invited Talks

Nov. 2007 MIT Departments of Chemistry and Biology Supergroup Meeting (Cambridge, MA)
"New strategies for targeting chemical probes using biotin ligase."

Sept. 2010 MIT Biological Chemistry division seminar series (Cambridge, MA)
"Protein-protein interaction imaging in living cells and neurons via proximity-dependent fluorophore ligation"

Poster Presentations

Sept. 2009 Fifth Annual NIH Director's Pioneer Award Symposium (Bethesda, MD)
S. Slavoff, J. D. Cohen, M. Fernández-Suárez, T. S. Chen, A. Thyagarajan, A. Y. Ting. "Imaging protein-protein interactions in cells via proximity-dependent fluorophore ligation."

June 2009 41st Annual National Organic Symposium (Boulder, CO)
S. Slavoff and A. Y. Ting. "Imaging protein-protein interactions in living cells using proximity-dependent fluorophore ligation."

Apr. 2005 11th Annual FDA Science Forum (Washington, DC)
P. R. Sundaresan, S. A. Slavoff, E. Grundel, K. D. White, J. I. Rader. "Differences in the constituents of two subspecies of *Teucrium*, *Teucrium chamaedrys* L. and *T. canadense* determined by analytical reversed-phase method."