Rational design and directed evolution of probe ligases for site-specific protein labeling and live-cell imaging

by

Katharine Alice White

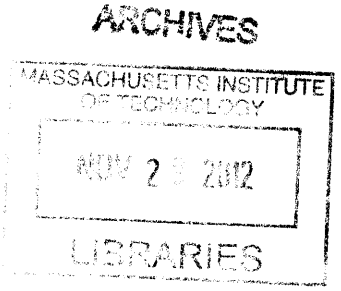
B.S. Chemistry (2003)
Saint Mary's College, Notre Dame, IN

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 2012



Signature of the Author:

Department of Chemistry
June 6, 2012

Certified by:

Alice Y. Ting
Ellen Swallow Richards Associate Professor of Chemistry
Department of Chemistry

Accepted by:

Robert W. Field
Chairman, Departmental Committee on Graduate Students

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

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

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

Alexander M. Klibanov
Novartis Professor of Chemistry and Bioengineering

Rational design and directed evolution of probe ligases for site-specific protein labeling and live-cell imaging

by

Katharine Alice White

Submitted to the Department of Chemistry
on June 6, 2012 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy

ABSTRACT

Chemical fluorophores have superior photophysical properties to fluorescent proteins and are much smaller. However, in order to use these probes for live-cell protein imaging, highly specific labeling methods are required. Here, we will describe three efforts to re-engineer the *E. coli* enzyme, lipoic acid ligase (LplA), to catalyze the ligation of small-molecule probes onto recombinant proteins. We call this collection of methods the PRIME (PRobe Incorporation Mediated by Enzymes) methodologies. First, we describe the structure-guided mutagenesis of LplA and the identification of an LplA variant that can ligate a blue coumarin fluorophore onto a 13-amino acid LplA acceptor peptide (LAP2). This "coumarin ligase" can be used to image cellular proteins with high specificity, sensitivity, and minimal perturbation of the biology of the protein of interest. We also demonstrate how subpopulations of a protein of interest can be labeled using genetically targeted coumarin ligase. Second, we describe our attempts to use yeast display evolution and fluorescence activated cell sorting (FACS) to evolve a truncated LplA enzyme. The original truncated enzyme had severely decreased activity for LplA's natural substrate, lipoic acid. We created a 10^7 library of LplA mutants and, after four rounds of selection, produced a truncated LplA mutant with lipoylation activity equivalent to full-length LplA. We next sought to evolve activity for an unnatural small molecule probe, but found that this strategy was limited by both increased hydrophobic probe sticking when using the truncated enzyme and some enzyme-dependent non-specificity. Finally, from a library of 10^7 LplA mutants, we evolved a full-length LplA capable of ligating an unnatural picolyl azide (pAz) substrate. We demonstrated improved activity of the "pAz ligase" in the secretory pathway and cell surface, two regions where coumarin ligase is inactive. This enzyme can also be used to image cell surface protein-protein interactions as well as label proteins as they are trafficked through the endoplasmic reticulum. These probe ligases will be useful tools for cell biologists interested in studying protein function or protein-protein interactions in the context of living cells.

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

Acknowledgements

I first acknowledge my advisor and mentor, Professor Alice Ting. Her passionate approach to science was a continual source of inspiration to me. Over my five years in the lab, Alice has relentlessly challenged me to think critically, work carefully, and push myself to produce the best science possible. Her drive and ambition are truly admirable.

I also want to recognize the members of my committee, Professor Alexander Klibanov and Professor Catherine Drennan, for being sources of support through my graduate work. My annual meetings with Cathy never failed to rejuvenate and inspire me. I particularly thank her for her support over the last six months; even while on sabbatical Cathy checked in with me, encouraging me when things never seemed to be going right.

I extend my heartfelt thanks to Dr. Dorothy Feigl. Her organic chemistry class changed the course of my life, pulling me from a major in mathematics into the bright and beautiful world of chemistry. She suggested that I try a summer research internship, encouraged me to apply to graduate schools, and convinced me to accept the offer from MIT. I would not be here today if not for her early and continual support. Saint Mary's College is lucky to have her and I can only hope to one day have such a profound and positive impact on the lives of my students.

I am deeply grateful to all the members of the Ting Lab. I particularly thank Dan, Tao, and Peng. I feel fortunate to have been able to work with these three fantastic scientists. They are hard working, methodical, intuitive, and brilliant. I learned more from these three than I imagined possible. To the women of the Ting Lab, past and present, I extend my warmest thanks. Marta, Sarah, Yoon-Aa all welcomed me into the lab and I am so happy that I have been able to work with Jen, Steph, Vicky, Wenjing, and Amy. Knowing that there is a group of colleagues who will never meet your questions or ideas with derision, condescension, or ridicule is a very valuable thing. To Phil, Ken, and Justin, you were the best bay-mates ever and I'm fortunate to have worked and laughed with you. To all Ting Labbers, I know that years from now the stress of minimeetings and failed experiments will have faded from our memories and we will have only the camaraderie and the bright, blinding moments of joy that make graduate school a beautiful thing.

To the Trivia Gang-Thomas, Dan, Arturo, Christina, Eric, and Becca-I owe you everything. I would have quit graduate school, but for you. You pulled me from the brink. You have shared in my successes, comforted me in my disappointments, and have been the best friends I have ever known. You gave laughter to my days and life to my life. I couldn't have done it without you. You carried me through, more than you'll ever know. Thomas, I feel lucky every day that you are in my life. Your love and support mean so much to me. I cannot wait for our next adventure together.

I dedicate this thesis to my family: Mom, Dad, Molly, Paddy, Eileen, and Norah. The more I see of the world the more I come to realize how extremely rare it is for a family to be filled with such overwhelming love. Your unconditional support made this 'hypothesis-driven research' seem worthy and valuable, which, by extension, helped me feel worthy and valuable. Thank you for that.

Table of Contents

List of Figures

Chapter 1: Current methodologies for site-specific protein labeling

Chapter 2: Engineering *E. coli* lipoic acid ligase for site-specific intracellular coumarin labeling

Chapter 3: Application of coumarin PRIME to compartmentalized labeling of intracellular proteins

Chapter 4: Considerations for the directed evolution of LplA and introduction to evolution methods

Chapter 5: Yeast-display evolution of a truncated LplA for site-specific protein labeling

Chapter 6: *In vitro* evolution of full-length LplA for improved activity in the endoplasmic reticulum and cell surface

Chapter 7: Combining yeast-display evolution and *in vitro* compartmentalization for the directed evolution of LplA

List of Tables

List of Abbreviations

2OGDH	2-oxogluterate dehydrogenase complex
11-Br	11-bromoundecanoic acid
10-Br-AMP	bromodecanoic adenylate ester
AF	AlexaFluor
AM	Acetoxymethyl ester
AMP	Adenosine monophosphate
AP	*E. coli* biotin ligase acceptor peptide
ATP	Adenosine triphosphate
BALK	Biotin alkyne
BCA	bicinchoninic acid
BFP	Blue fluorescent protein
BSA	Bovine serum albumin
CAAX	Farnesylation sequence
C_f	Final concentration
CFP	Cyan fluorescent protein
COS-7	African green monkey kidney cells
DCM	Dichloromethane
DIC	Differential interference contrast
DMEM	Dulbecco's modified Eagle's Medium
DMF	Dimenthyl formamide
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
DPBS	Dulbecco's phosphate-buffered saline
DPBS-B	Dulbecco's phosphate-buffered saline + 1% BSA
DTT	Dithiothreitol
E2p	Hybrid lipoyl domain (9 kDa) derived from second subunit of *E. coli* PDH complex
EBFP	Enhanced blue fluorescent protein
E. coli	Escherichia coli
EDTA	Ethylenediamine tetraacetic acid
ER	Endoplasmic reticulum
ESI-MS	Electrospray ionization mass spectrometry
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FP	Fluorescent protein
GSH/GSSG	Buffered with 10:1 ratio of reduced glutathione: oxidized glutathione
GFP	Green fluorescent protein
HA	Hemagglutanin
HEK	Human embryonic kidney cells
hLplA	LplA; "humanized" gene: codons optimized for mammalian expression
HPLC	High-performance liquid chromatography
HT	HaloTag protein
IPTG	isopropylthiogalactoside

IVC	*In vitro* compartmentalization
LAP1	Lipoic acid ligase acceptor peptide made through rational design
LAP2	Lipoic acid ligase acceptor peptide made through yeast-display
LAP2(ala)	LAP2 where the lysine modification site is mutated to alanine
LAP2-YIP	Yeast constitutively expressing LAP2 as a fusion to Aga1p
LMB	Leptomycin B
LPL	lipoate-protein ligase
MEM	Modified Eagle's medium
$Mg(OAC)_2$	Magnesium acetate
mRNA	messenger RNA
NES	Nuclear export sequence
NLS	Nuclear localization sequence
NHS	N-hydroxysuccinimide
NLG	Neuroligin
NRX	Neurexin
OD	optical density
PRIME	PRobe Incorporation Mediated by Enzymes
pAz	picolyl azide
PBS	phosphate buffered saline
PBS-B	phosphate buffered saline + 1% BSA
PCR	polymerase chain reaction
PE	phycoerythrin
PMSF	phenyl-methyl-sulfonyl fluoride
SA-PE	Streptavidin=phycoerythrin
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sfGFP	super-folder GFP
THPTA	click ligand
TEV	Tobacco etch virus
TBS	Tris-buffered saline
TBS-B	Tris-buffered saline + 1% BSA
TLC	Thin layer chromatography
TM	Transmembrane domain of PDGFR
Tris	Trishydroxymehtylaminomethane
WT	wild-type
YFP	Yellow fluorescent protein

Chapter 1. Introduction: Current methods for site-specific labeling of cell-surface and intracellular proteins.

Introduction

Over the past decade, fluorescence microscopy has become a leading tool for probing the function, localization, and interactions of cellular proteins. Although fluorescence microscopy allows for precise visualization of proteins with high temporal and spatial resolution, highly specific and minimally perturbative labeling methods are required in order to study biological function. The most common method for the fluorescent imaging of proteins utilizes recombinant fusions to fluorescent proteins (FPs). Fluorescent proteins are genetically encodable, and thus easy to target and perfectly specific. FPs are available in a wide range of colors and new developments in photoswitchable FPs and split-FPs allow for superresolution imaging and more sophisticated experiments investigating protein-protein interactions[1-3]. However, there are several limitations when using FPs to probe biological function. The most obvious limitation is that FPs are large tags, which can interfere with target protein localization, trafficking, and function[4-6]. FP fusions have been reported to increase receptor aggregation[7,8], alter target protein localization[4,5,9] as well as interfere with protein function[10]. In addition, the FP chromophores are not as bright as smaller organic dyes, have a lag time in maturation, and are more susceptible to photobleaching[1]. Small molecule probes, on the other hand, are desirable for their brightness, small size, and the variety of reporting functions they can perform. Unfortunately, since these probes must be incorporated into the target protein post-translationally, there is a challenge of achieving specific targeting. This challenge is particularly daunting when working in the complex and heterogeneous environment of the living cell.

Many labs have been working to develop new protein labeling methods with the goal of capitalizing on the superior photophysical properties of small, organic fluorophores and probes while relying on genetically encoded sequences for probe targeting. While there are a few tagless protein labeling methods such as unnatural amino acid mutagenesis[11-13] and ligand-directed tosyl chemistry[14] (Figure 1-1A), most protein labeling methods use a genetically encoded protein or peptide fusion to the protein of interest. Several labeling methods use a protein or peptide recognition sequence to directly recruit the small chemical probe. The final recruitment can result in a tight non-

covalent interaction (Figure 1-1B e.g. FKBP[15], DHFR[16,17], FlAsH[18,19], polyhistidine[20,21], TexasRed binding peptide[22]) or a covalent conjugation of the probe to the recognition sequence (Figure 1-1C, e.g. cutinase[23], SNAP/CLIP[24-26], and HaloTag[27]). Recently, several labs have used proximity-induced reactivity to successfully transform existing non-covalent tags to covalent tags (e.g. DHFR(L28C)[28] and covalent polyhistidine[29]). These methods use a ligand-directed reactive group and a nucleophilic residue on either the tag or the protein of interest to achieve covalent attachment of the entire ligand-probe conjugate (Figure 1-1D). In the previously outlined labeling schemes, labeling specificity is largely dependent on the bimolecular interaction between probe and the protein or peptide recognition sequence. With the idea of achieving exquisite labeling specificity while using a small tag, our lab and others have developed enzyme-mediated labeling techniques[30-34] (Figure 1-1E). The enzyme-mediated labeling methods rely on the intrinsic specificity enzymes have for their peptide substrates to achieve specific labeling of a very small tag. In this chapter, we will briefly describe the various protein labeling strategies that have been demonstrated on cell-surface and intracellular proteins. The information presented herein is summarized in Table 1-1 at the end of the chapter.

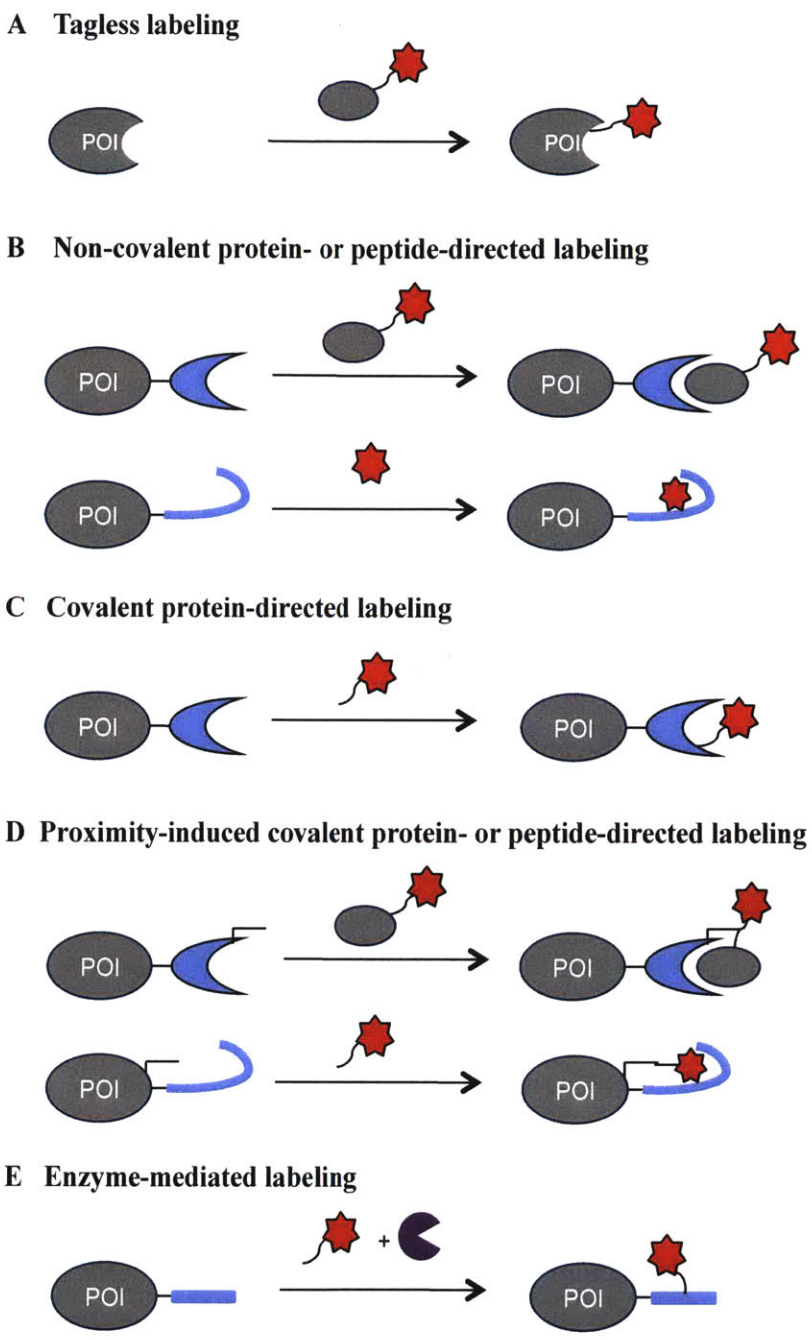


Figure 1-1: Strategies for labeling proteins with small molecule probes. (A) Tagless labeling. Ligand-directed tosyl chemistry uses a high-affinity ligand modified with an introduced reactive group. After ligand binding, the tosyl group reacts with a nucleophillic residue on the protein of interest (POI), producing covalent conjugation of the probe. **(B)** Non-covalent protein- or peptide-directed labeling. The protein or peptide tag directly binds probe or probe-ligand conjugate. **(C)** Covalent protein-directed labeling. A probe-conjugated suicide substrate is used for covalent probe attachment to the protein tag. **(D)** Proximity-induced covalent protein- or peptide-directed labeling. A probe or ligand is modified with a reactive group. After binding of probe or ligand, the reactive group covalently attaches the probe or probe-ligand to an introduced cysteine residue on either the protein tag or protein of interest. **(E)** Enzyme-mediated labeling. An enzyme catalyzes the covalent ligation of the probe to a peptide tag fused to the protein of interest.

Protein-based labeling methods

It is important to note that when using probe-recruiting tags there are significant trade-offs between labeling specificity and tag size. In general, protein-based recognition sequences give better labeling specificity than peptide-based recognition sequences. For non-covalent interactions, the increased specificity with protein recognition sequences is primarily the result of increased interaction surface area for ligand or probe binding. However, some protein-based labeling methods produce covalent modifications, which are by nature stronger than binding interactions.

Cell-surface protein labeling

Cutinase is a covalent protein-based labeling method that has been applied to the labeling of cell-surface proteins. Cutinase is a fungal serine esterase that reacts irreversibly with para-nitrophenyl phosphonate (p-NPP). For labeling, cutinase is genetically fused to the protein of interest and then can react with fluorophore p-NPP conjugates, producing covalently labeled recombinant protein fusions[23]. Cutinase is a large tag (22 kDa), comparable in size to many FPs, so it may disrupt the biology of the target protein. This labeling method is limited to the cell surface because p-NPP derivatives are highly charged and membrane-impermeant. Moreover, cutinase itself is a secreted enzyme that contains two disulfide bridges which are required for activity[35]; even if a membrane-permeable substrate were created, it is likely that recovering cutinase activity in the reducing environment of the cytosol would be a challenge.

Intracellular protein labeling

Several protein-based labeling methods use genetic fusions to proteins that are capable of binding small molecules with high affinities. One example of this type of labeling strategy developed by the Cornish lab uses *E. coli* dihydrofolatereductase (DHFR)[17]. The DHFR method relies on the ability of two DHFR inhibitors: methotrexate and trimethoprim to bind the enzyme with picomolar affinities[16,17,36]. By making fluorophore conjugates to either methotrexate or trimethoprim, specific fluorescent labeling can be achieved. Since methotrexate and trimethoprim are themselves cell-permeant, this method can be extended to intracellular labeling simply by utilizing cell-permeant fluorophores. Unfortunately, because methotrexate is also an inhibitor of

mammalian DHFR, labeling with methotrexate-based fluorophores requires using DHFR-deficient cell lines to eliminate background labeling of endogenous DHFR[17]. In order to eliminate the need for DHFR null cell lines, the Cornish lab developed the trimethoprim-based probes. Since trimethoprim has a higher affinity for *E. coli* DHFR over the mammalian DHFR, DHFR-deficient cell-lines are no longer required[16,37]. However, trimethoprim conjugates must be tested to ensure minimal background labeling of endogenous DHFR in each new cell line used[37].

In a conceptually similar approach, the Nolan lab developed a labeling method that uses a mutated protein and synthetic ligand pair to achieve specific probe targeting[15]. The technique uses an F36V point mutant of the FK506 binding protein (FKBP) which has picomolar affinity for a "bumped" synthetic ligand for FKBP (SLF') (K_d = 0.094 nM)[15]. Since SLF' is a designed ligand, it has high affinity for the modified F36V FKBP, but very low affinity for endogenous mammalian FKBP. This technique of using a protein mutant and specifically designed synthetic ligand reduces the background from endogenous protein labeling that initially plagued the DHFR technique. However, both the FKBP and DHFR methods have a few drawbacks that are inherent to the techniques. First, both methods use large tags that can interfere with protein function, trafficking and interactions. Second, since both methods are non-covalent, labeling signal can deteriorate over time as the ligand or inhibitor dissociates from the protein, which prevents long-term imaging. This problem is particularly obvious for DHFR where methotrexate conjugates have been reported to dissociate from the protein in just 1 hour[17]. To address this issue of quick probe-dissociation, a covalent form of DHFR labeling was recently developed[28]. By using the high affinity trimethoprim binding to bring a reactive acrylimide electrophile in close proximity to an introduced cysteine residue (L28C), Gallagher and colleagues were able to achieve covalent labeling with DHFR-tagged proteins.

The problem of probe-dissociation from a protein tag can alternatively be addressed by using suicide substrates that can covalently modify enzyme active sites. This technique is used in two recently developed covalent protein-based labeling methods. Both the HaloTag[27] and the SNAP/CLIP[24-26] methods rely on modified enzymes that are capable of self-labeling with fluorophore conjugated suicide-substrates. SNAP-

tag was the first such method introduced by Johnsson and colleagues. This method uses the human DNA repair enzyme O^6-alkylguanine-DNA alkyltransferase (hAGT). In the natural reaction, hAGT repairs O^6-alkylated guanine residues by covalent transfer of the alkyl group to an active site cysteine residue[25]. However, by using fluorophore-conjugated derivatives of O^6-benzylguanine, labeling of recombinant proteins fused to hAGT can be achieved. While SNAP tag was originally limited to AGT-deficient cell lines, the evolution of an orthogonal mutant eliminated this restriction and opened the method to labeling in any human cell line[26]. CLIP is an orthogonal labeling method that uses a mutant of hAGT that can only recognize O^6-benzylcytosine derivatives[24]. Using SNAP/CLIP in tandem allows for orthogonal two-color labeling of intracellular protein pairs[24]. In a similar approach, the HaloTag method uses a modified bacterial haloalkane dehalogenase to achieve covalent labeling using fluorophore-haloalkane conjugates[27]. In the natural reaction, the dehalogenase removes halides from aliphatic substrates by nucleophilic displacement of the halide. In the wild-type enzyme, an active-site aspartate forms an ester bond with the alkane substrate. Subsequent hydrolysis of the bond is catalyzed by a proximal histidine residue functioning as a general base, allowing release of the dehalogenated product and regenerates the aspartate residue. In the modified dehalogenase, that histidine residue has been mutated allowing covalent trapping of the enzyme-substrate complex[27]. Developed by Promega, this method has been used to attach not only organic fluorophores to cell surface and intracellular proteins, but also has been used to target halo-alkane coated quantum dots to cell-surface proteins[38]. These three covalent protein-based labeling methods have very high intracellular specificity and fast labeling reactions (under 30 minutes). However, all of these methods require large tag fusions (182-213 amino acids), which can interfere with target protein function, localization, and trafficking.

Peptide-based labeling methods

Peptide-based labeling methods modify the protein of interest via a genetically encoded peptide tag that is capable of binding small molecule probes with high affinity. While peptides are less perturbative than large protein tags, it is more difficult to achieve high binding affinity with peptides because of the reduced number of binding interactions between the peptide tag and the probe. In all of the methods introduced in this section,

specific peptide-probe interactions must be weighed against the potential for low binding affinities, fast probe off-rates, and non-specific background interactions between the probe and endogenous cellular components.

Cell-surface protein labeling

Ion-binding peptides have been used to site-specifically label proteins on the cell surface. Polyhistidine peptides have been used to specifically target nickel(II) nitrilotriacetic acid (Ni-NTA) complexes to proteins on the cell surface[20,21]. However, the binding affinity of this interaction is low compared to similar peptide-based targeting methods described below. With a K_d in the low micromolar range, the signal deterioration is very rapid and target protein expression must be very high in order to see strong labeling. Ni^{2+} toxicity is also an issue with this labeling method. The development of a dizinc probe with improved binding affinity and lower toxicity corrected some of these issues, but the method is still non-covalent and limited to cell-surface proteins[39]. A similar approach utilizes a polyaspartate peptide to recruit and bind a dinuclear Zn(II) complex[40]. As with the polyhistidine peptide, the polyaspartate peptide has low affinity and exhibits rapid dissociation of zinc-based probes. To correct the issue of rapid probe dissociation, Nonaka and colleagues developed a probe that included a chloroacetamide moiety which allows for covalent conjugation of the probe to the protein of interest; covalent labeling requires an accessible cysteine residue near the site of labeling which generally must be engineered into the protein of interest[41]. While the covalent modification prevents probe dissociation, this method is still limited to the cell surface due to the properties of the dinuclear zinc probes.

Peptides that bind the proteins streptavidin or bungarotoxin have also been used for cell surface protein labeling[42]. However, the large size of streptavidin and bungarotoxin makes the overall label large and partially defeats the purpose of using a small peptide tag. While the fusion protein will likely traffic properly and function properly prior to labeling, any analysis of labeled protein is complicated by a large overall tag size. Moreover, since the large size of streptavidin and bungarotoxin prevents intracellular delivery, this method is limited to the cell surface.

Intracellular protein labeling

Arguably the most successful peptide-based labeling method is the FlAsH method developed by the Tsien laboratory[19,43]. This method uses a 6-12 amino acid peptide that contains a tetracysteine motif (CCPGCC) that is capable of binding a biarsenical dye with picomolar affinity (K_d = 4-10 pM). The high affinity of this interaction means that the labeling signal does not deteriorate over time. Another feature of this method is that the biarsenical dyes become fluorescent upon binding to the peptide, which helps to reduce background fluorescence from unbound probe inside living cells. The small biarsenical dyes have been made in multiple colors along with singlet oxygen generators that are applicable to electron microscopy[44]. The combination of small tag size and cell permeable dyes with tight binding affinity has led to quick commercialization of this technology and repeated use in multiple cell biology studies[45-52]. However, even with all of these advantages, the FlAsH technology does have significant drawbacks that limit its utility. First, the biarsenical probes have weak affinity for mono- and di-cysteines which results in high background[53]. To address this issue, competing chelating agents such as 1, 2-ethanedithiol can be used to reduce background. However, this extends the labeling time significantly[53]. A higher affinity peptide (FLNCCPGCCMEP) was engineered that increases labeling signal and thus improves signal-to-noise ratios when performing labeling, but background is still a significant issue when labeling lower-expressing targets. Unwanted palmitoylation and oxidation of the tetracysteine motif can also decrease specific labeling signal[43]. Finally, the requirement for reduced cysteines prevents labeling in the secretory pathway or cell surface.

Peptides that bind small organic molecules have also been developed. The only one that has been demonstrated inside living cells is a peptide that specifically binds the fluorophore Texas Red, developed by Nolan and colleagues[15]. The peptide binds with picomolar affinity (K_d ~25pM) and has been used on the cell surface and inside living cells. However, accumulation of the TexasRed fluorophore in mitochondria gives high intracellular background and restricts labeling to only the most highly expressed targets[15]. Moreover, since the peptide is specific for the Texas Red fluorophore, labeling with other dyes is not possible using this method.

Tagless labeling methods

Intracellular protein labeling

Unnatural amino acid mutagenesis is one of the most versatile methods for labeling intracellular proteins[11,12,54-57]. This method labels proteins by using engineered tRNA synthetases to charge tRNAs with unnatural amino acids. The engineered tRNAs can recognize the amber stop codon (UAG) and incorporate the unnatural amino acid co-translationally and site-specifically into the protein of interest. Unnatural amino acid mutagenesis has been used to directly incorporate probes such as benzophenone[58] and fluorescent coumarin[59] derivatives as well as functional groups like ketones[57], azides[56,60], and hydrazines[55]. While this method is minimally perturbative to the target protein, allows incorporation of a variety of reporters, and is, in theory, perfectly specific, there are significant drawbacks. The first issue is the requirement to evolve new aminoacyl-tRNA synthetases for each new unnatural amino acid target. This can be time-consuming and technologically challenging for labs that don't already have the required bacterial strains or experience with the selection protocol. Other significant problems include competing termination at amber stop codons with accumulation of truncated target protein and concomitant read-through of endogenous amber stop codons[11]. The latter becomes particularly problematic when using this method in mammalian cells, as amber stop codons are more frequent in the mammalian genome than in the *E. coli* genome.

Another recent tagless protein labeling method was developed by Tsukiji and colleagues[14]. The strategy behind this method is to use a high affinity ligand to target a reactive probe specifically to a native protein. The first example of this method used a ligand-directed tosyl group to label carboxyanhydrase. Similar to the covalent Asp tag and covalent DHFR modifications, this method requires tight binding of a ligand-fluorophore conjugate with an introduced reactive group for proximity-dependent labeling of the target protein. Unlike the previously described methods, however, this chemistry is an SN2-type reaction wherein the tosyl moiety forms a covalent bond to the nucleophilic protein residue with the original inhibitor structure as the leaving group[14]. This means that the inhibitor itself is not covalently attached to the protein, so function is not inherently disrupted. Examples of this method have been demonstrated in living cells,

in tissue, and in mice[14]. Unfortunately, one drawback to this method is that it is only applicable to proteins that have known high affinity ligands or inhibitors[61]. Even if such ligands exist, the chemistry to produce the reactive probe-tosyl-ligand compounds can be quite challenging. Finally, the nature of the chemistry requires that the probe be conjugated to a residue proximal to the native ligand binding site; it is possible that the added bulk of the probe in a biologically critical portion of the protein of interest may interfere with subsequent ligand binding events and the biological activity of the protein of interest.

While tagless methods are minimally perturbative, there are drawbacks that limit the utility of these methods for intracellular protein labeling. There is significant demand for less technologically challenging methods that allow for specific incorporation of a wider range of probes. The methods discussed in the final section attempt to address this unmet need.

Enzyme-mediated labeling methods

As mentioned previously, vast improvements have been made to peptide-based labeling methods by utilizing enzymes to improve labeling specificity. These methods hijack endogenous post-translational modification enzymes and engineer them to perform site-specific ligation of small molecule probes onto an introduced peptide tag. These methods combine the high specificity of enzyme-substrate interactions with the benefits small peptide tags. The crux of these labeling schemes is that the enzyme must be highly sequence specific but flexible in terms of small molecule substrate.

Cell-surface protein labeling

The first reported example of an enzyme-mediated labeling method exploits phosphopantetheine transferases (PPTases, AcpS and Sfp) to perform site-specific protein labeling[62-64]. PPTases in their natural reaction catalyze the transfer of 4'-phosphopantetheine from coenzyme A (CoA) to a serine residue of an acyl or peptidyl carrier protein (ACP and PCP, respectively), which are involved in the biosynthesis of polyketides and nonribosomal polypeptides. PPTases are quite permissive in the CoA derivatives that they are able to use as substrates, making this method quite versatile. While the original method was demonstrated on ACP and PCP proteins, evolved 12-mer

peptides (A1 and S6 for AcpS and Sfp, respectively) are now frequently used for this labeling method[65]. The ligation is fast, specific, covalent, and the method has been used successfully to probe biological function[34,66]; however, since the CoA derivatives are highly charged, this technique is limited to the cell surface.

A recently developed method, SorTag, utilizes a bacterial transpeptidase to covalently label cell-surface proteins[67]. The method relies on Sortase A from *Staphylococcus aureus*, which in its natural reaction functions as both a protease and transpeptidase to attach pilli and perform additional decoration of proteins anchored in the cell wall. Sortase A specifically recognizes LPXTG peptide sequences and cleaves between the T and G residues while forming a new peptide bond with a polyglycine-protein. Specific protein labeling can be achieved by using a LPXTG tag to the protein of interest and supplying polyglycine-probe conjugates. This method uses a very small tag (5aa) and is fast and specific. Recently, Chen and colleagues have used yeast-display evolution to develop a Sortase A mutant with improved coupling activity over the wild-type ligase[68]. This advance gives better labeling of low-expressing target proteins. However, this method is still limited to the cell surface as polyglycine-probe conjugates are not cell permeable.

Another enzyme-mediated cell-surface labeling method is based on guinea pig transglutaminase (gpTGase)[30]. The natural reaction of transglutaminases is to catalyze amide bond formation between glutamine and lysine side-chains in a calcium-dependent reaction. The guinea pig transglutaminase was selected because it is sequence specific for the glutamine-containing peptide (Q-tag) but is permissive of a range of amine substrates[69]. This is a key feature that is a goal of all enzyme-mediated labeling methods. Q-tags have been used to label recombinant proteins *in vitro* and were extended by our lab for fast labeling of cell surface proteins[30,70-72]. Unfortunately, the method is hindered by nonspecificity on the cell surface. Specific labeling of the glutamine residue in the Q-tag competes against non-specific labeling of endogenous glutamine residues found in cell surface proteins. While intracellular labeling would not be advisable due to nonspecific labeling, it is likely impossible because basal intracellular calcium concentrations are too low for gpTGase activity.

Our lab has developed a method that utilizes *E. coli* biotin ligase (BirA) for site-specific protein labeling[31]. The natural reaction of BirA is to catalyze the ATP-dependent ligation of biotin to a specific lysine residue of biotin carboxyl carrier protein (BCCP). Previous engineering developed a 15-amino acid acceptor peptide (AP) that BirA can biotinylate with identical ligation kinetics to the protein substrate. Biotinylation of the AP tag can be detected using streptavidin-fluorophore or streptavidin-coated quantum dots[73]. Using these strategies, our lab has demonstrated ensemble and single-molecule imaging of AP-tagged surface proteins[73,74]. We also developed an orthogonal yeast biotin ligase and yeast acceptor peptide pair that allows us to perform multi-color imaging[32]. This labeling method is highly specific and fast, but using large streptavidin conjugates for labeling, as previously mentioned, restricts labeling to cell-surface proteins and defeats the purpose of a small tag. To reduce the overall size of the label, we developed a ketone isostere of biotin that can be efficiently ligated by the wild-type BirA enzyme[31]. Ketones can be chemoselectively derivatized with hydrazines or hydroxylamines, thus eliminating the need for streptavidin-based detection of the site-specific labeling. However, reaction kinetics of ketones with hydrazines or hydroxylamines is very slow which limits the temporal resolution of this method. Moreover, this method is still restricted to the cell surface because, while absent from the cell surface, ketones are present in the cytosol.

Intracellular protein labeling

The enzyme-mediated labeling methods described so far are limited to the cell surface and crippled by either slow labeling kinetics or poor specificity. After finding *E. coli* BirA to be difficult to engineer and failing to find biotin ligases from other species with increased small molecule promiscuity[75], our lab sought out a new ligase that performed similar chemistry. We settled on *E. coli* lipoic acid ligase (LplA), which, in its natural reaction, catalyzes the ligation of lipoic acid to proteins involved in oxidative metabolism. Like BirA, it is highly specific for the protein substrates, but LplA is more malleable in terms of small molecule substrates, ligating both lipoic acid and octanoic acid. Our lab quickly determined that LplA can accept alkyl-azide substrates which can then be chemoselectively derivatized, allowing site-specific cell surface labeling[33]. The challenge was to adapt this to an intracellular labeling method. This motivated the work described in Chapters 2 and 3 of this thesis and led to the development of the intracellular

labeling method we call PRobe Incorporation Mediated by Enzymes, or PRIME. Using LplA as a platform for a protein labeling method proved very lucrative; our lab produced ligases for intracellular labeling with coumarin[76-78], resorufin, azides[79], alkynes[33], and tetrazines[80]. More recently we were able to develop an interaction-dependent form of the coumarin ligase labeling technology[81]. However, there were several limitations to this intracellular labeling method (summarized in Chapter 2) including a lack of enzyme activity in the secretory pathway. After attempts to solve this problem through rational design of the ligase failed, we resorted to *in vitro* evolution techniques, the results of which are summarized in Chapters 5-7.

We believe that the LplA labeling technology holds a wealth of potential for site-specific protein labeling. The varieties of probes that can be incorporated using this technology make it one of the most versatile discussed herein. Moreover, the fast ligation kinetics and small tag size make it applicable for minimally perturbative labeling of biologically relevant proteins with good temporal control. The progress described in this thesis of adapting the method to intracellular protein labeling and evolving a ligase with activity on the cell surface and endoplasmic reticulum makes this method one of the most generalizable protein labeling methods to-date. However, there is a significant hurdle for labs that want to adapt the method for use in their own systems; individual labs must synthesize the PRIME labeling substrates if they wish to use the method. The importance of probe availability on method adoption is exemplified by the drastic increase in experiments utilizing HaloTag and SNAP/CLIP after the commercial availability of various alkylhalide-, benzylguanine- and benzylcytosine-probe conjugates[82-90].

Table 1-1: Summary of methods for cell-surface and intracellular protein labeling.

Cell-surface protein labeling methods						
Method	*Genetic Tag*	*Tag size*	*Enzyme*	*Covalent*	*Other comments*	*Refs*
Protein Tags						
Cutinase	Cutinase	213 aa	—	Yes	Large tag	23
Peptide Tags						
PolyHis	PolyHis	6-12 aa	—	No/Yes	Low probe affinity for non-covalent version	20, 21, 29, 39
PolyAsp	PolyAsp	8-22 aa	—	No/Yes	Low probe affinity for non-covalent version	40, 41
Streptavidin and bungarotoxin binding peptides	SA or BT binding peptide	13 aa	—	No	Detection method produces large tag	42
AP	AP	17 aa	Biotin ligase	Yes	Slow kinetics for ketone-biotin	31, 32, 73-75
PPTase	A1/S6	11 aa	AcpS/Sfp	Yes	Fast and specific labeling with small tag	65
Transglutaminase labeling	Q-tag	7 aa	Transglutaminase	Yes	High background from endogenous glutamine residues	30
Sortase labeling	SorTag	5 aa	Sortase A	Yes	Fast and specific labeling with small tag	67, 68

Table 1-1(cont): Summary of methods for cell-surface and intracellular protein labeling.

Intracellular protein labeling methods						
Method	*Genetic Tag*	*Tag size*	*Enzyme*	*Covalent*	*Other comments*	*Refs*
Protein Tags						
FKBP	FKBP12	98 aa	—	No	Non-covalent labeling, eventual loss of labeling signal	15
DHFR	DHFR/DHFR(L28C)	157 aa	—	No/Yes	Fast off-rates for non-covalent version	16, 17, 28, 37
HaloTag	Dehalogenase	296 aa	—	Yes	Fast and specific labeling with large tag	27, 38, 84-88
SNAP/CLIP	O^6-alkylguanine-DNA-alkyltransferase	182 aa	—	Yes	Fast and specific labeling with large tag	25, 26, 34, 80-83
Peptide Tags						
TexasRed Binding Peptide	TexasRed Binding Peptide	23-38 aa	—	No	High mitochondrial background	22
Tetracysteine	CCPGCC or FLNCCPGCCMEP	6-12 aa	—	No	Low specificity	18, 19, 43-52
PRIME/ID-PRIME	LAP2/LAP1	13/17 aa	LplA	Yes	Fast and specific labeling with small tag	32, 76-79
Tagless						
UAA mutagenesis	UAG codon	0 aa	—	Yes	Truncated protein products, native stop-codon read-through	11-13, 56-60
Ligand-directed tosyl chemistry	—	0 aa	—	Yes	Covalent modification near native ligand binding site	14, 61

References

1. Fernandez-Suarez, M.; Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat Rev Mol Cell Biol* **2008**, *9*, 929-43.

2. Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **2006**, *312*, 217-24.

3. Lippincott-Schwartz, J.; Patterson, G. H. Development and use of fluorescent protein markers in living cells. *Science* **2003**, *300*, 87-91.

4. Lisenbee, C. S.; Karnik, S. K.; Trelease, R. N. Overexpression and mislocalization of a tail-anchored GFP redefines the identity of peroxisomal ER. *Traffic* **2003**, *4*, 491-501.

5. 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*, 231-40.

6. Moritz, O. L.; Tam, B. M.; Papermaster, D. S.; Nakayama, T. A functional rhodopsin-green fluorescent protein fusion protein localizes correctly in transgenic Xenopus laevis retinal rods and is expressed in a time-dependent pattern. *J Biol Chem* **2001**, *276*, 28242-51.

7. Brock, R.; Hamelers, I. H.; Jovin, T. M. Comparison of fixation protocols for adherent cultured cells applied to a GFP fusion protein of the epidermal growth factor receptor. *Cytometry* **1999**, *35*, 353-62.

8. Marsh, D. R.; Holmes, K. D.; Dekaban, G. A.; Weaver, L. C. Distribution of an NMDA receptor:GFP fusion protein in sensory neurons is altered by a C-terminal construct. *J Neurochem* **2001**, *77*, 23-33.

9. Elliott, G.; O'Hare, P. Intercellular trafficking of VP22-GFP fusion proteins. *Gene Ther* **1999**, *6*, 149-51.

10. McLean, A. J.; Milligan, G. Ligand regulation of green fluorescent protein-tagged forms of the human beta(1)- and beta(2)-adrenoceptors; comparisons with the unmodified receptors. *Br J Pharmacol* **2000**, *130*, 1825-32.

11. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* **2007**, *4*, 239-44.

12. Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. 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*, 10092-7.

13. Umehara, T., *et al.* N-Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett* **2012**, *586*, 729-33.

14. Tsukiji, S.; Miyagawa, M.; Takaoka, Y.; Tamura, T.; Hamachi, I. Ligand-directed tosyl chemistry for protein labeling in vivo. *Nat Chem Biol* **2009**, *5*, 341-3.

15. Marks, K. M.; Braun, P. D.; Nolan, G. P. A general approach for chemical labeling and rapid, spatially controlled protein inactivation. *Proc Natl Acad Sci U S A* **2004**, *101*, 9982-7.

16. Miller, L. W.; Cai, Y.; Sheetz, M. P.; Cornish, V. W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat Methods* **2005**, *2*, 255-7.

17. Miller, L. W.; Sable, J.; Goelet, P.; Sheetz, M. P.; Cornish, V. W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew Chem Int Ed Engl* **2004**, *43*, 1672-5.

18. Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281*, 269-72.

19. 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*, 6063-76.

20. Guignet, E. G.; Segura, J. M.; Hovius, R.; Vogel, H. Repetitive reversible labeling of proteins at polyhistidine sequences for single-molecule imaging in live cells. *Chemphyschem* **2007**, *8*, 1221-7.

21. Guignet, E. G.; Hovius, R.; Vogel, H. Reversible site-selective labeling of membrane proteins in live cells. *Nat Biotechnol* **2004**, *22*, 440-4.

22. Marks, K. M.; Rosinov, M.; Nolan, G. P. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* **2004**, *11*, 347-56.

23. 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*, 14753-8.

24. Gautier, A., *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* **2008**, *15*, 128-36.

25. Keppler, A.; Pick, H.; Arrivoli, C.; Vogel, H.; Johnsson, K. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc Natl Acad Sci U S A* **2004**, *101*, 9955-9.

26. Gronemeyer, T.; Chidley, C.; Juillerat, A.; Heinis, C.; Johnsson, K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel* **2006**, *19*, 309-16.

27. Los, G. V., *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008**, *3*, 373-82.

28. Gallagher, S. S.; Sable, J. E.; Sheetz, M. P.; Cornish, V. W. An in vivo covalent TMP-tag based on proximity-induced reactivity. *ACS Chem Biol* **2009**, *4*, 547-56.

29. Uchinomiya, S. H., *et al.* Site-specific covalent labeling of His-tag fused proteins with a reactive Ni(II)-NTA probe. *Chem Commun (Camb)* **2009**, 5880-2.

30. Lin, C. W.; Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* **2006**, *128*, 4542-3.

31. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2005**, *2*, 99-104.

32. Chen, I.; Choi, Y. A.; Ting, A. Y. 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*, 6619-25.

33. Fernandez-Suarez, M., *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **2007**, *25*, 1483-7.

34. Vivero-Pol, L.; George, N.; Krumm, H.; Johnsson, K.; Johnsson, N. Multicolor imaging of cell surface proteins. *J Am Chem Soc* **2005**, *127*, 12770-1.

35. Longhi, S.; Cambillau, C. Structure-activity of cutinase, a small lipolytic enzyme. *Biochim Biophys Acta* **1999**, *1441*, 185-96.

36. Sasso, S. P.; Gilli, R. M.; Sari, J. C.; Rimet, O. S.; Briand, C. M. Thermodynamic study of dihydrofolate reductase inhibitor selectivity. *Biochim Biophys Acta* **1994**, *1207*, 74-9.

37. Calloway, N. T., *et al.* Optimized fluorescent trimethoprim derivatives for in vivo protein labeling. *Chembiochem* **2007**, *8*, 767-74.

38. Zhang, Y., *et al.* HaloTag protein-mediated site-specific conjugation of bioluminescent proteins to quantum dots. *Angew Chem Int Ed Engl* **2006**, *45*, 4936-40.

39. Hauser, C. T.; Tsien, R. Y. A hexahistidine-Zn2+-dye label reveals STIM1 surface exposure. *Proc Natl Acad Sci U S A* **2007**, *104*, 3693-7.

40. 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*, 10452-9.

41. Nonaka, H.; Tsukiji, S.; Ojida, A.; Hamachi, I. Non-enzymatic covalent protein labeling using a reactive tag. *J Am Chem Soc* **2007**, *129*, 15777-9.

42. McCann, C. M.; Bareyre, F. M.; Lichtman, J. W.; Sanes, J. R. Peptide tags for labeling membrane proteins in live cells with multiple fluorophores. *Biotechniques* **2005**, *38*, 945-52.

43. Martin, B. R.; Giepmans, B. N.; Adams, S. R.; Tsien, R. Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* **2005**, *23*, 1308-14.

44. Adams, S. R.; Tsien, R. Y. Preparation of the membrane-permeant biarsenicals FlAsH-EDT2 and ReAsH-EDT2 for fluorescent labeling of tetracysteine-tagged proteins. *Nat Protoc* **2008**, *3*, 1527-34.

45. Zurn, A., *et al.* Site-specific, orthogonal labeling of proteins in intact cells with two small biarsenical fluorophores. *Bioconjug Chem*, *21*, 853-9.

46. Panchal, R. G., *et al.* In vivo oligomerization and raft localization of Ebola virus protein VP40 during vesicular budding. *Proc Natl Acad Sci U S A* **2003**, *100*, 15936-41.

47. Ju, W., *et al.* Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nat Neurosci* **2004**, *7*, 244-53.

48. Ignatova, Z.; Gierasch, L. M. Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc Natl Acad Sci U S A* **2004**, *101*, 523-8.

49. Rudner, L., *et al.* Dynamic fluorescent imaging of human immunodeficiency virus type 1 gag in live cells by biarsenical labeling. *J Virol* **2005**, *79*, 4055-65.

50. Szecsi, M.; Spindler-Barth, M. Flash labeling of a nuclear receptor domain (D domain of ultraspiracle) fused to tetracysteine tag. *Acta Biol Hung* **2006**, *57*, 181-90.

51. Wilkins, B. J.; Yang, X.; Cropp, T. A. Photochemical control of FlAsH labeling of proteins. *Bioorg Med Chem Lett* **2009**, *19*, 4296-8.

52. Pattnaik, A. K.; Panda, D. Biarsenical labeling of tetracysteine-tagged proteins for tracking existing and newly synthesized pools of proteins. *Cold Spring Harb Protoc* **2009**, *2009*, pdb prot5343.

53. Stroffekova, K.; Proenza, C.; Beam, K. G. 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*, 859-66.

54. Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **2006**, *35*, 225-49.

55. Zhang, Z., *et al.* A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **2003**, *42*, 6735-46.

56. Ohno, S., *et al.* Site-selective post-translational modification of proteins using an unnatural amino acid, 3-azidotyrosine. *J Biochem* **2007**, *141*, 335-43.

57. Lemke, E. A. Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. *Methods Mol Biol*, *751*, 3-15.

58. Farrell, I. S.; Toroney, R.; Hazen, J. L.; Mehl, R. A.; Chin, J. W. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. *Nat Methods* **2005**, *2*, 377-84.

59. Wang, J.; Xie, J.; Schultz, P. G. A genetically encoded fluorescent amino acid. *J Am Chem Soc* **2006**, *128*, 8738-9.

60. Chin, J. W., *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc* **2002**, *124*, 9026-7.

61. Wang, H., *et al.* Chemical cell-surface receptor engineering using affinity-guided, multivalent organocatalysts. *J Am Chem Soc*, *133*, 12220-8.

62. George, N.; Pick, H.; Vogel, H.; Johnsson, N.; Johnsson, K. Specific labeling of cell surface proteins with chemically diverse compounds. *J Am Chem Soc* **2004**, *126*, 8896-7.

63. Yin, J.; Liu, F.; Li, X.; Walsh, C. T. Labeling proteins with small molecules by site-specific posttranslational modification. *J Am Chem Soc* **2004**, *126*, 7754-5.

64. Meier, J. L.; Mercer, A. C.; Rivera, H., Jr.; Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J Am Chem Soc* **2006**, *128*, 12174-84.

65. Zhou, Z., *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol* **2007**, *2*, 337-46.

66. Yin, J., *et al.* Single-cell FRET imaging of transferrin receptor trafficking dynamics by Sfp-catalyzed, site-specific protein labeling. *Chem Biol* **2005**, *12*, 999-1006.

67. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: a versatile method for protein labeling. *Nat Chem Biol* **2007**, *3*, 707-8.

68. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011**, *108*, 11399-404.

69. Lorand, L., *et al.* Specificity of guinea pig liver transglutaminase for amine substrates. *Biochemistry* **1979**, *18*, 1756-65.

70. Sato, H.; Hayashi, E.; Yamada, N.; Yatagai, M.; Takahara, Y. Further studies on the site-specific protein modification by microbial transglutaminase. *Bioconjug Chem* **2001**, *12*, 701-10.

71. Sato, H.; Ikeda, M.; Suzuki, K.; Hirayama, K. Site-specific modification of interleukin-2 by the combined use of genetic engineering techniques and transglutaminase. *Biochemistry* **1996**, *35*, 13072-80.

72. Taki, M.; Shiota, M.; Taira, K. Transglutaminase-mediated N- and C-terminal fluorescein labeling of a protein can support the native activity of the modified protein. *Protein Eng Des Sel* **2004**, *17*, 119-26.

73. Howarth, M.; Takao, K.; Hayashi, Y.; Ting, A. Y. Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc Natl Acad Sci U S A* **2005**, *102*, 7583-8.

74. Zou, P.; Ting, A. Y. Imaging LDL receptor oligomerization during endocytosis using a co-internalization assay. *ACS Chem Biol*, *6*, 308-13.

75. Slavoff, S. A.; Chen, I.; Choi, Y. A.; Ting, A. Y. Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *J Am Chem Soc* **2008**, *130*, 1160-2.

76. Uttamapinant, C., *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A*, *107*, 10914-9.

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

78. Cohen, J. D.; Thompson, S.; Ting, A. Y. Structure-guided engineering of a Pacific Blue fluorophore ligase for specific protein imaging in living cells. *Biochemistry*, *50*, 8221-5.

79. Yao, J. Z., *et al.* Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J Am Chem Soc*, *134*, 3720-8.

80. Liu, D. S., *et al.* Diels-Alder cycloaddition for fluorophore targeting to specific proteins inside living cells. *J Am Chem Soc*, *134*, 792-5.

81. Slavoff, S. A.; Liu, D. S.; Cohen, J. D.; Ting, A. Y. Imaging protein-protein interactions inside living cells via interaction-dependent fluorophore ligation. *J Am Chem Soc*, *133*, 19769-76.

82. Keppler, A.; Ellenberg, J. Chromophore-assisted laser inactivation of alpha- and gamma-tubulin SNAP-tag fusion proteins inside living cells. *ACS Chem Biol* **2009**, *4*, 127-38.

83. Kampmeier, F., *et al.* Site-specific, covalent labeling of recombinant antibody fragments via fusion to an engineered version of 6-O-alkylguanine DNA alkyltransferase. *Bioconjug Chem* **2009**, *20*, 1010-5.

84. Regoes, A.; Hehl, A. B. SNAP-tag mediated live cell labeling as an alternative to GFP in anaerobic organisms. *Biotechniques* **2005**, *39*, 809-10, 812.

85. Mao, S., *et al.* Optical lock-in detection of FRET using synthetic and genetically encoded optical switches. *Biophys J* **2008**, *94*, 4515-24.

86. Padilla-Parra, S., *et al.* Quantitative comparison of different fluorescent protein couples for fast FRET-FLIM acquisition. *Biophys J* **2009**, *97*, 2368-76.

87. Huybrechts, S. J., *et al.* Peroxisome dynamics in cultured mammalian cells. *Traffic* **2009**, *10*, 1722-33.

88. Yamaguchi, K.; Inoue, S.; Ohara, O.; Nagase, T. Pulse-chase experiment for the analysis of protein stability in cultured mammalian cells by covalent fluorescent labeling of fusion proteins. *Methods Mol Biol* **2009**, *577*, 121-31.

89. Gallo, S.; Beugnet, A.; Biffo, S. Tagging of functional ribosomes in living cells by HaloTag(R) technology. *In Vitro Cell Dev Biol Anim*, *47*, 132-8.

90. BenYounes, A., *et al.* A fluorescence-microscopic and cytofluorometric system for monitoring the turnover of the autophagic substrate p62/SQSTM1. *Autophagy*, *7*, 883-91.

Chapter 2. Engineering *E. coli* lipoic acid ligase for site-specific labeling of intracellular proteins

The work presented in this chapter has been published in part in C. Uttamapinant, K.A. White, H. Baruah, S. Thompson, M. Fernández-Suárez, S. Puthenveetil, A.Y. Ting A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A*, **2005** *107*, 10914-9. The experimental section of that publication is reproduced here in part and with permission. Hemanta Baruah performed the initial alanine mutant screening and Samuel Thompson performed much of the W37 mutant screening on E2p. Chayasith Uttamapinant performed synthesis of the protected coumarin probes and determined kinetics for coumarin ligation using ^{W37V}LplA and ^{W37I}LplA.

Introduction

As outlined in Chapter 1, the Ting lab has developed several different enzyme-mediated methods for site-specific labeling of cell surface proteins. The most promising method utilizes *E. coli* lipoic acid ligase (LplA) to ligate an azidoalkanoic acid, which can then be chemoselectively derivatized using cyclooctyne-fluorophore conjugates[1]. This method is fast, specific, and can be used to incorporate many different fluorophores and probes. While the other enzyme-mediated labeling methods developed by our lab are limited to the cell surface by either incompatible chemistry[2] (in the case of BirA) or poor specificity[3] (in the case of transglutaminase), the LplA labeling technology has no intrinsic feature that would prohibit its use for intracellular protein labeling. In this chapter we will introduce the biochemical properties of *E. coli* LplA, discuss the considerations that must be made for intracellular protein labeling, and then describe how we have extended this robust cell-surface labeling method to intracellular protein labeling.

Properties of E. coli lipoic acid ligase

Lipoic acid is an essential cofactor found in multi-enzyme complexes involved in oxidative metabolism. Through the activity of lipoate protein ligases (Lpls), free lipoic acid is covalently attached to a specific lysine residue of lipoyl domains in these complexes[4]. LplA is the lipoate protein ligase found in *E. coli* and it catalyzes the ATP-dependent ligation of lipoic acid in a two-step reaction. In the first step, LplA uses ATP to activate lipoic acid, forming the lipoyl-AMP ester and releasing pyrophosphate. In the second step of the reaction, the enzyme binds the protein substrate and the target lysine residue attacks the adenylate ester, resulting in amide bond formation and covalent attachment of lipoic acid to the lysine (Figure 2-1). The function of the resulting lipoyl-lysine modification is to serve as a "swinging arm" that shuttles reaction intermediates and reducing equivalents between the active sites of the enzyme subunits[5]. While *E. coli* LplA can catalyze both the adenylation and the transfer of lipoic acid, most mammalian Lpls[6-9] (and some non-mammalian Lpls)[10-16] can catalyze only the transfer step. In these

systems, a separate enzyme, lipoate activating enzyme (LAE), is required to produce the adenylate ester prior to the action of the Lpl enzyme[7].

In *E. coli* there are three protein substrates that must be lipoylated for function: the E2 dihydrolipoamide acetyltransferase subunit (E2p) of pyruvate dehydrogenase (PDH), the E2 dihydrolipoamide succinyltransferase subunit (E2o) of 2-oxogluterate dehydrogenase, and the H-protein of the glycine cleavage system[5]. Mammals have an additional lipoylated protein: the E2 acyltransferase subunit of the branched chain α-ketoacid dehydrogenase. These protein substrates and analysis of Lpl sequence and structure conservation between species will be discussed in depth in Chapter 4 of this thesis.

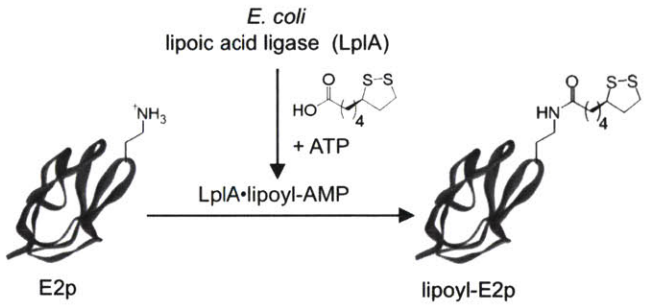


Figure 2-1. The site-specific ligation of lipoic acid catalyzed by *E. coli* lipoic acid ligase. The lipoylation of endogenous *E. coli* proteins is catalyzed by lipoic acid ligase (LplA) in two distinct steps. First, the enzyme uses ATP to activate the carboxylic acid of lipoic acid, producing the adenylate ester (lipoyl-AMP). The enzyme then binds the protein substrate (here E2p, indicated in grey ribbon) and the reactive lysine residue attacks the lipoyl-AMP resulting in amide bond formation and release of AMP.

Even though the endogenous abundance of LplA is extremely low (reported to be as few as 10 copies per cell[4]), endogenous lipoyl domains in *E. coli* are nearly 100% lipoylated[4], which could be explained by the fast kinetics of the lipoic acid ligation (k_{cat} is 0.25 sec^{-1})[1]. Since LplA plays a key role in the production of active enzyme complexes involved in oxidative metabolism, it first appeared that LplA might be an essential enzyme; however, LplA null strains are growth-competent and have the full complement of lipoylated substrates[4]. While LplA uses exogenously supplied or scavenged lipoic acid to lipoylate protein substrates, there is a complementary pathway in *E. coli* that produces lipoylated substrates using endogenously synthesized lipoic acid. In this pathway, endogenously synthesized octanoic acid is first transferred from octanoyl-acyl carrier protein onto the lipoyl acceptor protein by the enzyme LipB. Then LipA, an S-adenosyl

methionine-dependent enzyme, acts on the octanoyl-protein substrate and inserts sulfur atoms at the C6 and C8 positions of the octanoyl-lysine modification, producing the lipoyllysine modification.

Considerations for intracellular labeling

As we have already mentioned, one feature that makes LplA such an attractive candidate for an intracellular labeling method is its exquisite sequence specificity. Another intriguing feature is the ability of ^{WT}LplA to catalyze the ligation of other substrates such as selenolipoic acid, octanoic acid, and the unnatural azidoalkanoic acid substrates that we use for cell surface protein labeling[1,4,17](Figure 2-2). While BirA is also highly sequence specific, it is not permissive in terms of small-molecule substrates, only ligating biotin or a ketone isostere of biotin, neither of which is applicable to live-cell intracellular labeling.

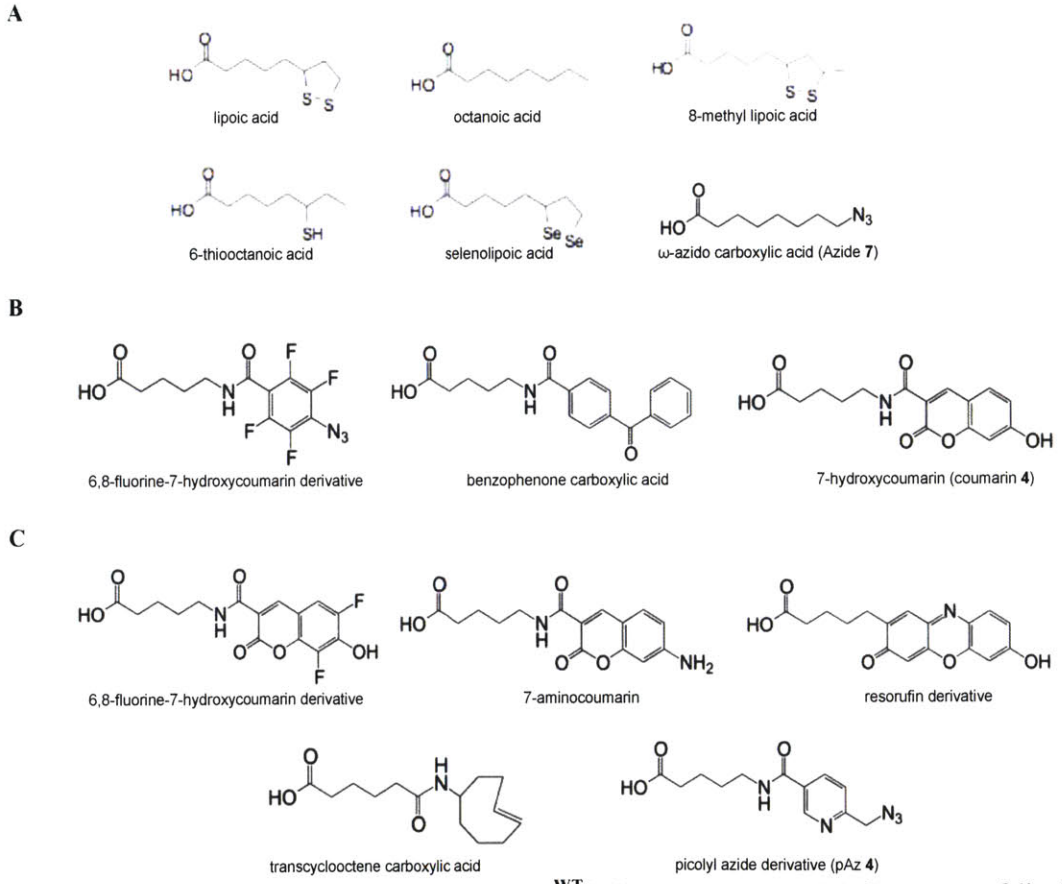


Figure 2-2. Range of substrates incorporated by ^{WT}LplA or a mutant. (A) Structures of lipoic acid analogues and probes ligated by ^{WT}LplA *in vitro* or on the cell surface. **(B)** Structures of useful probes that had been ligated by LplA mutants *in vitro* or on the cell surface before starting the intracellular labeling project. **(C)** Structures of probes incorporated by LplA mutants since the publication of the intracellular labeling paper.

The versatile two-step cell-surface protein labeling method previously developed by our lab allows for the incorporation of virtually any fluorophore or reporter (Figure 2-3, middle). However, we recognized that extending this two-step labeling scheme to intracellular protein labeling presented several challenges. First, the sensitivity of the labeling would be limited by the kinetics of the [3 + 2] strain-promoted cycloaddition, which has a rate constant of 4.3×10^{-3} $M^{-1}sec^{-1}$ [18]. Second, performing a two-step labeling protocol inside cells would require two washout steps, first to remove any unligated azide and then to wash out excess cyclooctyne-fluorophore conjugate. Third, and perhaps most importantly, preliminary attempts at intracellular azide labeling indicated that azide **7** labeling with wild-type LplA was nonspecific inside living cells (data not shown). While the two-step labeling protocol was eventually extended to site-specific intracellular live-cell labeling in beautiful work done by Jen Yao[19], we decided to explore a simpler one-step fluorophore ligation scheme for the first inroads into intracellular protein labeling.

The first concern with this plan was that fluorophores applicable to live-cell imaging[20] are much larger than any substrate ^{WT}LplA had previously been shown to ligate (Figure 2-2A). Fortunately, as we began the work described in this chapter, Dr. Hemanta Baruah demonstrated that a mutant of LplA could accept and ligate a larger arylazide photocrosslinker with no change in sequence specificity[21]. This was an encouraging result because it indicated that larger small molecule substrates might be incorporated with some engineering of the enzyme's substrate binding pocket. With this information, we settled on a 7-hydroxycoumarin as our target fluorophore for this project. This blue fluorophore was selected because it is one of the smallest fluorophores suitable for live-cell imaging. We also hypothesized that its small size and hydrophobicity would make it a good substrate for LplA. While the photophysical properties of 7-hydroxycoumarin are not ideal (ε = 36,700 M^{-1} cm^{-1} and QY 0.7)[22] it is brighter than EBFP (ε = 31,500 M^{-1} cm^{-1} and QY 0.2)[23] and hundreds of times smaller. The goal is to engineer LplA to ligate the fluorophore directly (Figure 2-3, bottom).

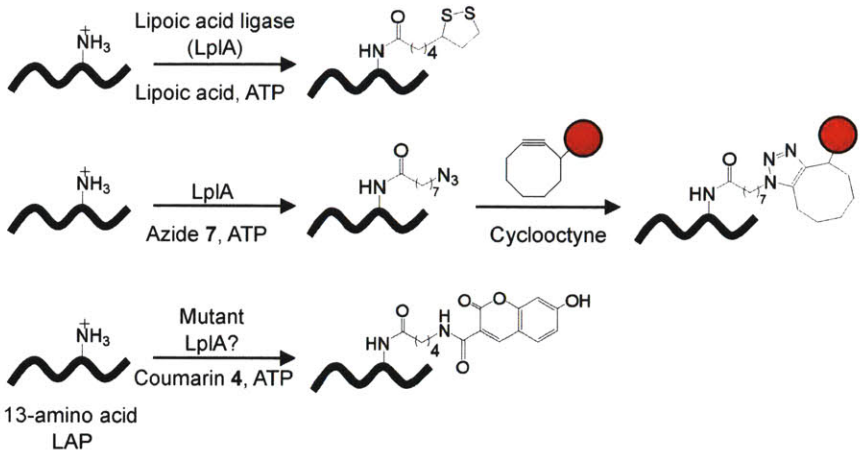


Figure 2-3. Natural and engineered ligation reactions catalyzed by LplA. Top row shows the natural reaction catalyzed by LplA. Middle row shows the cell-surface two-step probe targeting via alkyl azide **7** ligation followed by [3 + 2] cycloaddition. Bottom row shows proposed intracellular one-step 7-hydroxy coumarin fluorophore ligation by a mutant LplA. LAP = LplA Acceptor Peptide.

We were also concerned that the kinetics of ligation onto the rationally designed LplA Acceptor Peptide (LAP) would need to be improved in order to achieve fast and specific intracellular labeling. While LAP1 (DEVLVEIETDKAVLEVP)[1] is labeled by ^{WT}LplA with similar *in vitro* ligation kinetics to E2p (Az **7** k_{cat} = 0.05 sec^{-1} for LAP1 vs. 0.11 sec^{-1} for E2p), the enzyme affinity for LAP1 is very low (K_m > 300 μM). Unfortunately, with an affinity that low, it is likely that labeling of LAP1-tagged intracellular proteins will not be highly sensitive. To address this, a higher affinity LAP peptide (LAP2 = GFEIDKVWYDLDA) was evolved using yeast-display[24]. This peptide has similar lipoylation kinetics to the protein substrate (lipoic acid k_{cat} = 0.2 sec^{-1} for LAP2 vs. 0.25 sec^{-1} for E2p) but a much higher affinity (K_m = 13 μM) compared to LAP1.

In the results section of this chapter we describe the engineering of a blue fluorophore ligase and its application to intracellular protein labeling of peptide-tagged recombinant proteins. First, we present the structure-guided rational mutagenesis of LplA to produce an enzyme that is capable of ligating a 7-hydroxycoumarin probe. Second, we describe the *in vitro* characterization of this enzyme and confirmation of specificity on mammalian cell lysate. Third, we describe the application of this enzyme to intracellular protein labeling. Finally, we describe several shortcomings of this labeling method that motivated the enzyme engineering work that is subsequently described in Chapters 5-7 of this thesis.

Results

Structure-guided mutagenesis and screening for a coumarin fluorophore ligase

Identification of Trp37 as a 'gatekeeper' residue

For the first stage of LplA engineering, we synthesized four 7-hydroxycoumarin substrates with varying linker lengths (Figure 2-4). These probes were synthesized by reacting succinimidyl 7-hydroxycoumarin-3-carboxylate with one of four amino alkanoic acids (n =3, 4, 5, or 6). We performed an *in vitro* activity assay to determine if wild-type LplA could accommodate any of these coumarin substrates. High performance liquid chromatography (HPLC) was used to monitor the extent of ligation onto a 9kD hybrid lipoyl domain from pyruvate dehydrogenase (E2p)[4]. Using this assay, we determined that the wild-type enzyme was unable to ligate any of the coumarin substrates.

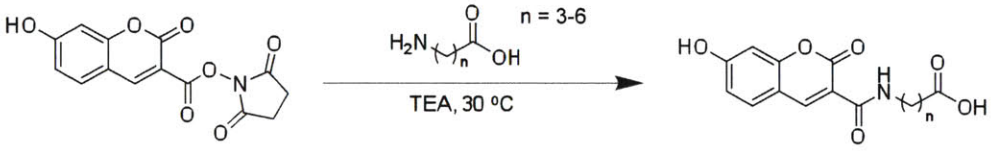


Figure 2-4. Synthesis of 7-hydroxycoumarin substrates with varying linker lengths.

This result is not surprising when we consider the crystal structure of *E. coli* LplA bound to the lipoyl-AMP intermediate (PDB: 3A7R)[25] (Figure 2-5). Note that the dithiolane ring of the bound lipoyl-AMP is located at the very end of the hydrophobic binding tunnel and there is not much room deeper into the existing binding pocket. We proposed mutating LplA residues that were in close proximity to the dithiolane ring of lipoyl-AMP in order to expand the binding pocket and better accommodate the larger, two-ring coumarin structure (Figure 2-5, left). To that end, we selected thirteen residues within 7.5Å of the dithiolane moiety of lipoic acid: N16, L17, V19, E20, F35, W37, S71, S72, H79, T87, R140, H149 (shown in green stick in Figure 2-5). We made LplA point mutants wherein each of these residues was mutated to alanine and then screened these alanine point mutants for ligation activity using the two smaller 7-hydroxycoumarin substrates (n = 3, 4). Of the thirteen alanine point mutants tested, only W37A produced detectable coumarin-E2p ligation product. This result makes sense given the orientation of W37 (shown in red stick in Figure 2-5) at the end of the lipoate binding pocket and

given its close proximity to the dithiolane ring (3.6Å). Interestingly, the W37 residue of LplA was also identified in the engineering of the aryl azide ligase as the position most important in expanding substrate specificity of LplA[21]. Since mutations at W37 seem to open the door to new and useful probe ligation using LplA, we call W37 the "gatekeeper" residue.

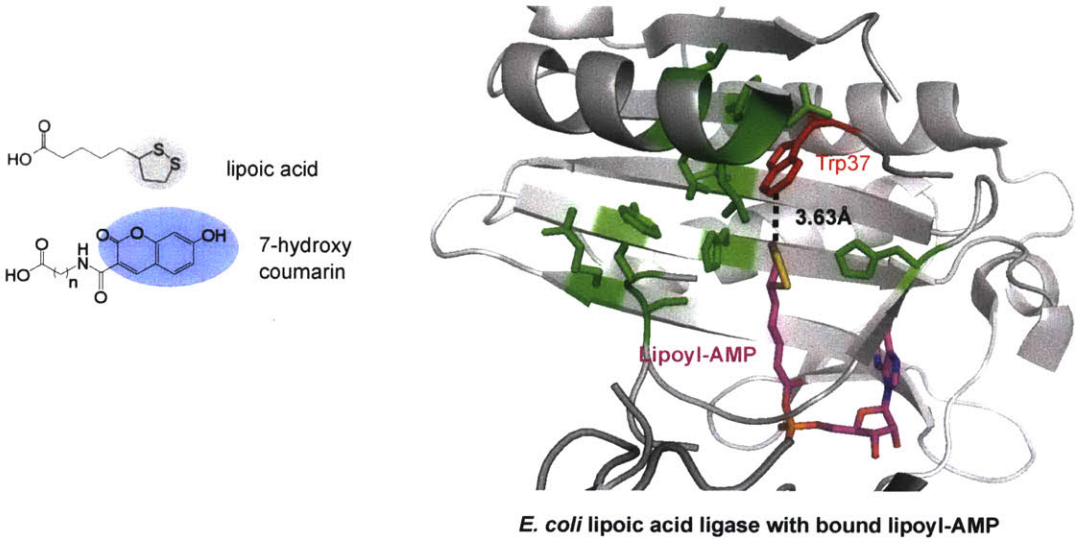
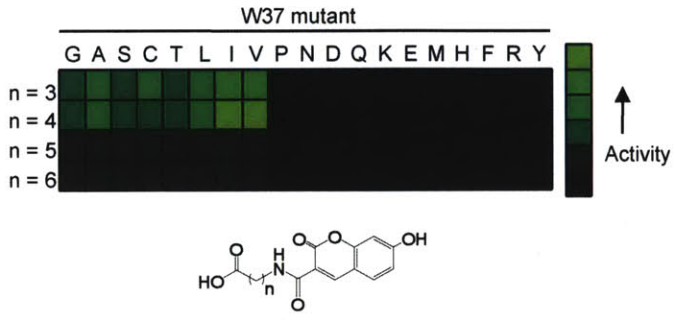


Figure 2-5. LplA active site residues selected for mutagenesis. Residues within 7.5Å of the dithiolane ring of lipoyl-AMP are rendered in green or red stick. These residues were mutated to alanine in the initial screen in order to expand the binding pocket to accommodate the larger coumarin structure (left). Trp37, indicated in red, was identified as the gatekeeper residue. This image was generated from the original PDB file (3A7R)[25] using PyMol software.

Secondary screen of Trp37 mutants identifies two coumarin ligases

While coumarin ligation with ^{W37A}LplA was detectable in short *in vitro* reactions, we wondered whether different mutations at the gatekeeper residue might further improve coumarin ligation activity. To that end, we prepared a panel of LplA mutants wherein W37 was mutated to all possible amino acids. We screened this panel of W37 point mutants against the four 7-hydroxycoumarin probes (Figure 2-6A). When we had completed analysis of the enzyme/probe matrix, we noticed some clear trends in the coumarin ligation activities of these mutants. There were 11 inactive mutants that exhibited no labeling of E2p in the in vitro screen. These inactive enzymes were ones where W37 had been mutated to charged residues (K, R, D, E), bulky residues (F, Y, N, Q, M, H), or proline (P). It is likely that large groups at the W37 position are not able to

accommodate even the shortest coumarin probes, while charged residues may interfere with binding of the hydrophobic coumarin probe.



LplA mutant	% Conversion with Coumarin 4
W37V	100 ± 8.3
W37I	27.9 ± 5.6
W37A	15.5 ± 2.8
W37L	5.9 ± 1.1
W37G	< 2
W37S	< 2
W37C	< 2
W37T	< 2

Figure 2-6. ^{W37V}LplA and ^{W37I}LplA are identified as the best coumarin ligases. Assays were performed for 12 hours using the indicated probes and mutants. Production of the coumarin-E2p conjugate was monitored by HPLC. Array shows relative activities of W37 LplA mutants with 7-hydroxycoumarin substrates, black quadrants indicate that no product was detected. Active ligases were subsequently screened for coumarin **4** LAP4.3D peptide in a 40-minute labeling protocol. Table shows quantification of coumarin **4** ligation onto LAP4.3D. Assay performed in triplicate, errors ± 1 s.d.

Eight mutants demonstrated activity for coumarin ligation onto E2p in this screen. Without exception, these mutants preferred the smaller coumarin substrates (n = 3, 4) and very little product was seen with the longer substrates. This indicates that there is a limit to how much we can enlarge the binding pocket using gatekeeper mutations and also suggests that engineering LplA to incorporate fluorophores larger than coumarin will require additional mutations to carve out even more space in the enzyme. Interestingly, simply enlarging the binding pocket of LplA by mutating the gatekeeper residue to the smallest residues (alanine and glycine) did not actually produce the highest coumarin ligation activity; we found that mutations with intermediate-sized hydrophobic side chains such as isoleucine and valine gave the biggest activity increase. We hypothesize that these side chains are best able to complement the size and hydrophobicity of the coumarin probe. This trend of increased ligation activity with hydrophobic residues suggests that the ligase actually binds the neutral form of 7-hydroxycoumarin, rather than the deprotonated form (pKa =7.5)[22].

To better distinguish between the activities of the eight active mutants, we performed a secondary screen for coumarin **4** (n=4) ligation activity where we replaced the E2p protein substrate with one of the evolved LAPs fused to the carrier protein HP1 (LAP4.3D = GFEHEKVWYDLDA)[24]. Note that we chose LAP4.3D as opposed to the best yeast-evolved peptide (LAP2 = GFEIDKVWYDLDA) simply because it gives superior chromatographic traces on HPLC when fused to the carrier protein HP1, making product conversion easier to quantify. The results of this more stringent labeling protocol can be found in the table in Figure 2-6. Although ^{W37V}LplA is by far the best ligase *in vitro*, we had some evidence that intracellular expression of ^{W37V}LplA might be problematic; Marta Fernandez-Suarez had demonstrated that ^{W37V}LplA, when expressed as a fusion to the fluorescent protein mCherry, was prone to aggregation inside mammalian cells (data not shown). Primarily because of this issue, we decided to carry both ^{W37V}LplA and ^{W37I}LplA through for further *in vitro* characterization and downstream testing in mammalian cells.

In vitro characterization of coumarin ligases

^{W37V}LplA and ^{W37I}LplA catalyze fast and specific covalent ligation of coumarin onto LAP2

We first characterized the coumarin ligation *in vitro*. Using our most kinetically efficient LAP2 peptide, we performed *in vitro* coumarin **4** ligation reactions with ^{W37V}LplA and ^{W37I}LplA. Figure 2-7A shows that both ^{W37V}LplA and ^{W37I}LplA covert LAP2 to LAP2-coumarin in a 20-minute labeling protocol. The LAP2-coumarin **4** covalent conjugate was verified via mass spec (Figure 2-7B). Controls where ATP is omitted or wild-type LplA is used demonstrate that the coumarin ligation is ATP-dependent and requires the specific mutation we have introduced into the enzyme.

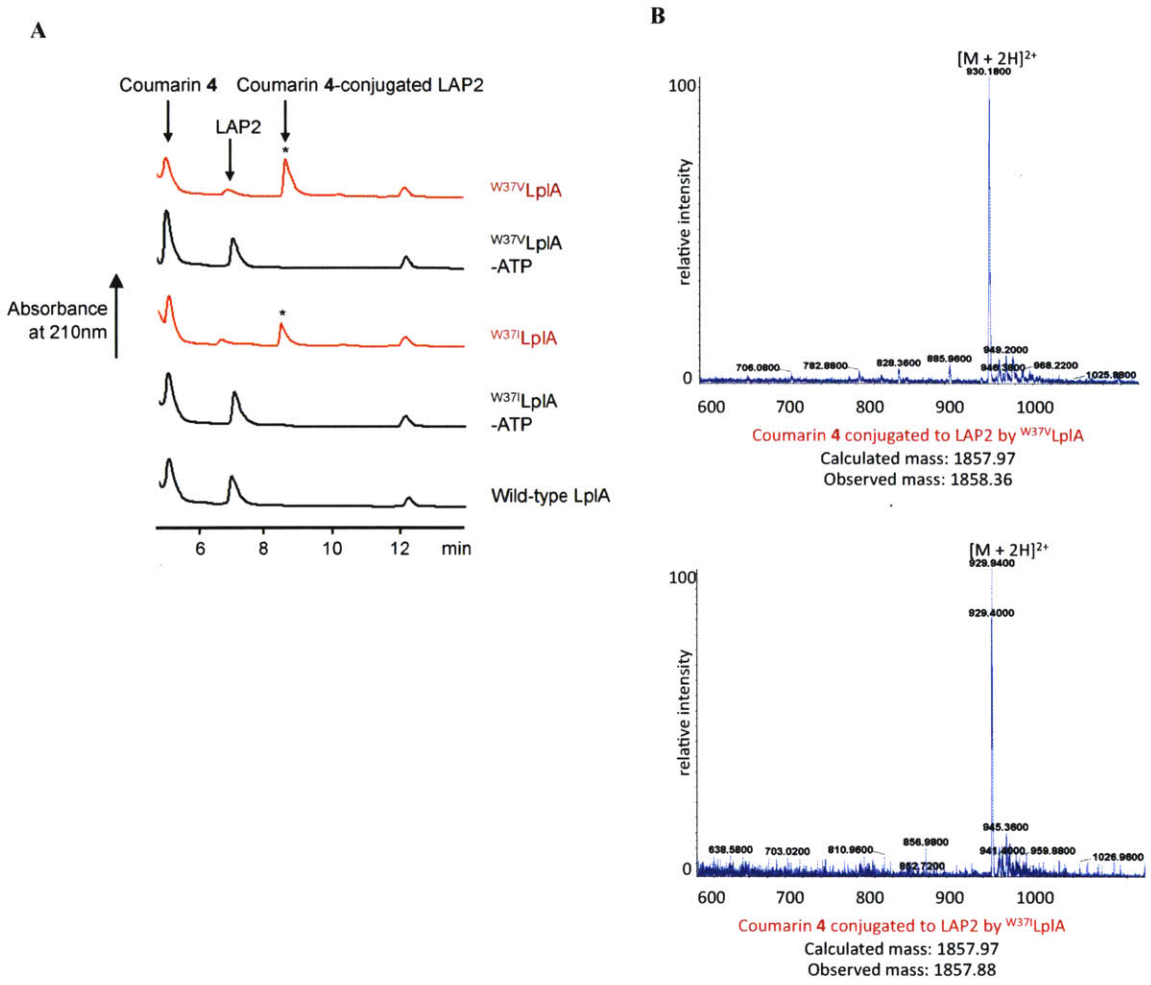


Figure 2-7. *In vitro* characterization of coumarin ligases. (A) HPLC traces showing conversion of LAP2 peptide to the LAP2-coumarin conjugate with ^{W37V}LplA and ^{W37I}LplA. Negative controls are shown where ^{WT}LplA is used or ATP is omitted. **(B)** Starred peaks in (B) were collected and analyzed via mass spectrometry, confirming the identity of the LAP2-coumarin conjugates.

We next characterized the kinetics of coumarin ligation (Figure 2-8). For ^{W37V}LplA we obtained a k_{cat} of 0.019 ± 0.004 sec^{-1} and a K_m of 56 ± 20 μM. While ^{W37I}LplA has a similar k_{cat} (0.016 ± 0.002 sec^{-1}), we found it has a much higher K_m of 261 ± 69 μM. While the calculated coumarin k_{cat} values are poor compared to the wild-type lipoylation reaction[1,4,24,26] and even some of the unnatural substrates[1,21], the coumarin K_m for ^{W37V}LplA compares favorably to previous K_m values for unnatural probes [1,21](Table 2-1). For comparison, wild-type LplA has a catalytic efficiency (k_{cat}/K_m) of 1.2 x 10^5 while ^{W37V}LplA has a catalytic efficiency of 3 x 10^2.

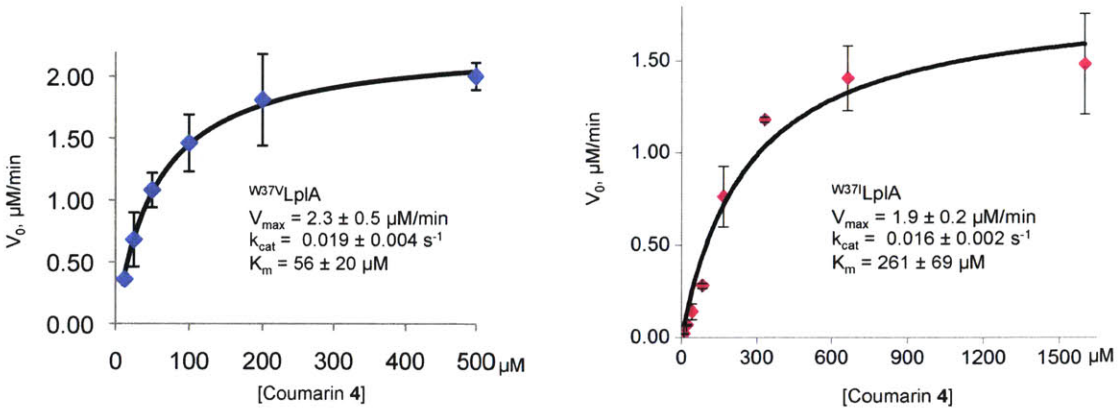


Figure 2-8. Kinetics of coumarin ligation with ^{W37V}LplA and ^{W37I}LplA. Michaelis-Menten curves give k_{cat} and K_m values for coumarin **4** ligation with ^{W37V}LplA (left) and ^{W37I}LplA (right). Curves were generated from HPLC measurements of initial rates (V_0) using 2 µM enzyme, 150 µM LAP2 and coumarin **4** concentrations ranging from 12.5 and 1,600 µM. Plots were made using Origin 6.1 software. Error bars ± 1 s.d.

Probe	k_{cat} (sec^{-1})	K_m (µM)	LplA mutant	Acceptor sequence	Reference
lipoic acid		4.5	wild-type	H-protein	[11]
lipoic acid		1.7	wild-type	PDH*	[4]
lipoic acid	0.25		wild-type	E2p	[1]
lipoic acid	0.22		wild-type	LAP2	[24]
alkyl azide **7**	0.111	127 ± 11	wild-type	E2p	[1]
alkyl azide **7**	0.048		wild-type	LAP1	[1]
aryl azide	0.31	<80	W37V	LAP1	[21]
coumarin **4**	0.019	56 ± 20	W37V	LAP2	(this work)
coumarin **4**	0.016	261 ± 69	W37I	LAP2	(this work)

*Pyruvate dehydrogenase: protein substrate for LplA with 3 E2p domains

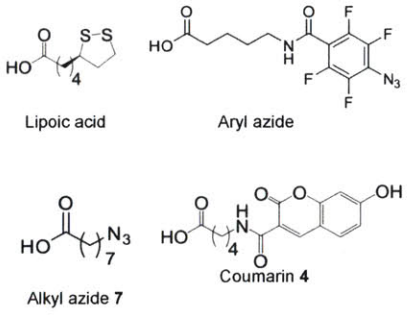


Table 2-1. Comparison of kinetic constants for ligation of lipoic acid and unnatural probes using LplA and mutants

Coumarin ligases are site-specific and do not recognize endogenous mammalian proteins

It was unclear whether or not remodeling the binding pocket of LplA would affect its protein and peptide substrate specificity. To probe this issue, we tested the specificity of both ^{W37V}LplA and ^{W37I}LplA on mammalian cell lysates. First we created a LAP2 fusion to yellow fluorescent protein (YFP) and expressed it in human embryonic kidney (HEK) cells. HEK lysates were prepared and then labeled with coumarin **4**, ATP, and

either ^{W37V}LplA or ^{W37I}LplA. We then analyzed these samples by in-gel coumarin fluorescence. Figure 2-9 shows that even in the presence of thousands of other mammalian proteins in the cell lysate, only LAP2-YFP is labeled by coumarin (Lane 1 and 4). To confirm site-specificity we performed labeling on HEK lysate from cells expressing LAP2(ala)-YFP wherein the lysine modification site of LAP2 is mutated to an alanine and this abolishes labeling (Lane 2 and 5). This demonstrated that coumarin labeling is site-specific and no endogenous mammalian proteins are labeled with coumarin. We were somewhat surprised at the lack of background, given that there are four endogenous substrates for mammalian Lpl and LplA has been demonstrated to cross-react with substrates from other species[4,27,28]. Based on our data, we hypothesized that mammalian Lpl substrates present in the lysate are not labeled with coumarin because labeling sites are already saturated with lipoic acid. To test this hypothesis, we performed anti-lipoic acid western blotting of the samples (Figure 2-9, bottom). We found that while the endogenous lipoylated protein substrates of mammalian Lpl are present, *in vitro* lipoylation of the lysate does not increase lipoylation level of these proteins.

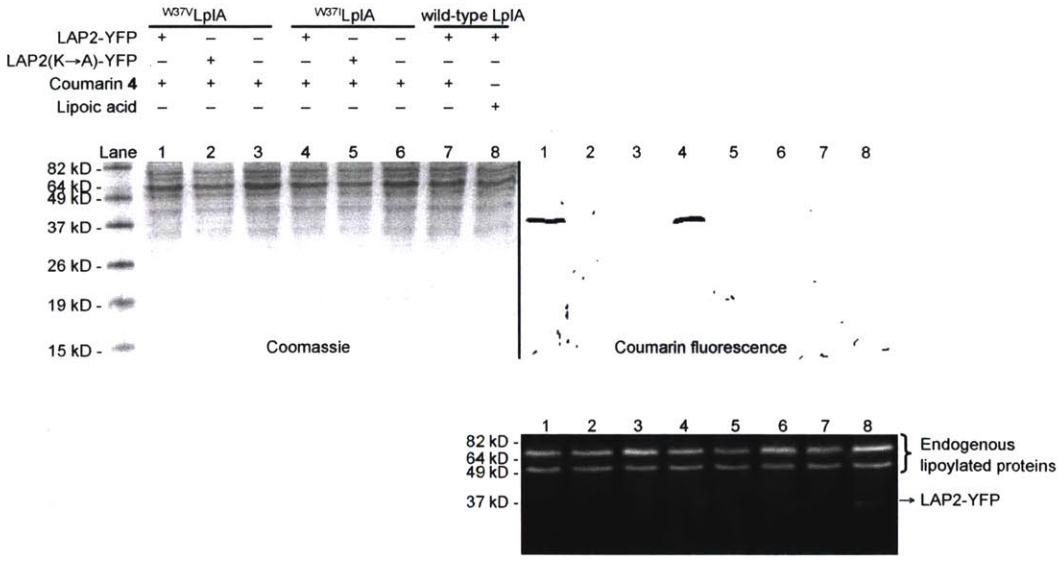


Figure 2-9. Coumarin labeling is site-specific on mammalian cell lysate. Lanes 1 and 4 show specific labeling of LAP2-YFP in HEK cell lysate monitored by in-gel coumarin fluorescence. Negative controls are shown for coumarin labeling on cells expressing LAP2(ala)-YFP (lanes 2 and 5), untransfected lysate (lanes 3 and 6) and using ^{WT}LplA, The samples were also analyzed by western blot using anti-lipoic acid antibody (bottom). Coomassie staining of the gel shows all proteins present in the lysate.

Optimizing conditions for intracellular protein labeling

In order to extend this robust in vitro coumarin labeling method to intracellular labeling, several considerations must be made (Figure 2-10). First we must co-express the enzyme and LAP2-fused substrate inside mammalian cells. Second, we must deliver the coumarin probe across the cell membrane and then washout excess unligated probe. Note that we can rely on endogenous ATP for the ligation, since ATP is present in millimolar concentrations inside the cell. Third, we needed to test the evolved LAP peptides in an intracellular context and select the best sequence for coumarin labeling. While LAP2 is the most kinetically efficient peptide *in vitro*, it was not inevitable that it would be best for intracellular applications; for example, we have determined that LAP4.2 gives the best surface expression and that LAP4.3D is the best-behaved for *in vitro* analysis as a fusion.

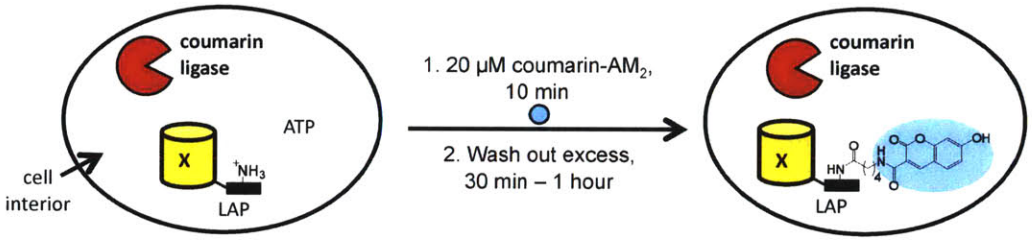


Figure 2-10. Scheme for intracellular coumarin labeling. First, the coumarin ligase and LAP-tagged protein of interest must be co-expressed inside living cells. Then we must deliver the coumarin probe across the cell membrane. After labeling, we must be able to remove excess unligated coumarin in order to observe specific labeling.

Coumarin-AM$_2$ is cell permeable and has clean washout from cells

We first tackled the issue of probe delivery. Since the coumarin **4** probe has 1-2 charges at physiological pH, it cannot cross the cell membrane. To achieve intracellular delivery we protected the carboxylate and 7-hydroxy group with acetoxymethyl groups, which have been shown to be cleaved by endogenous esterases in mammalian cells[29]. Tao Uttamapinant synthesized the protected coumarin **4** probe (coumarin-AM$_2$; structure shown in Figure 2-11) by reacting the 7-hydroxycoumarin probe with bromomethylacetate to get nucleophilic substitution at both the carboxylate (coumarin-AM) and, after an overnight reaction, substitution at the hydroxyl, giving the doubly-protected coumarin-AM$_2$. We discovered that, while coumarin-AM was not membrane permeable, coumarin-AM$_2$ entered cells readily (data not shown). We determined that

after incubating HEK cells with 20 μM coumarin-AM_2 for just 10 minutes, intracellular probe concentrations were >400 μM. This was reassuring because this concentration is well above the calculated K_m for either ^{W37V}LplA or ^{W37I}LplA. We were also able to achieve complete washout of the probe by replacing the media several times over 30 minutes.

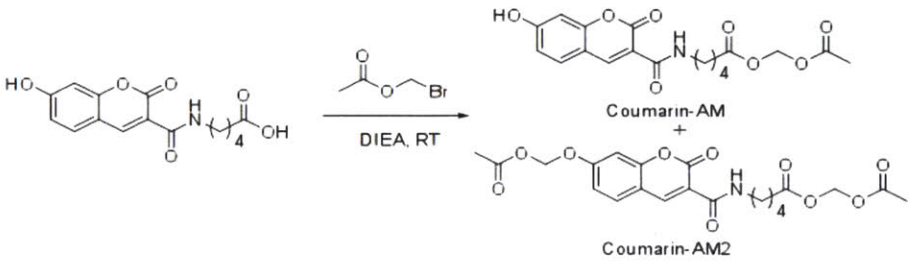


Figure 2-11. Synthesis of coumarin-AM and cell-permeable coumarin-AM_2 probes.

^{WT}LplA expression level must be controlled for specific intracellular lipoylation

The next challenge was to achieve good expression of LplA inside mammalian cells. As we mentioned above, we found that *E. coli* LplA did not express well, aggregating inside cells when expressed as a fusion to mCherry and giving very low expression otherwise (data not shown). However, we found that by using human-preferred codons for LplA expression, we were able to see good intracellular expression of mCherry-LplA. Encouraged by this, we decided to first test whether ^{WT}LplA could catalyze intracellular lipoylation of E2p and LAP peptide substrates. We added lipoic acid to cells expressing ^{WT}LplA and either nuclear E2p (E2p-BFP-NLS) or nuclear LAP1 (LAP1-BFP-NLS) and incubated for several hours; we then fixed the cells and detected substrate lipoylation using immunofluorescence staining (Figure 2-12A). We routinely use nuclear-targeted substrates to have an internal control for specificity; if the enzyme is specific, we should see only nuclear labeling. Unfortunately, we found that while lipoylation of E2p was specific, there was non-specific labeling when using nuclear-targeted LAP1 and almost as much signal when using LAP1(ala) where the target lysine has been mutated to alanine. Through various optimizations we discovered that the nonspecificity towards the LAP1 peptide correlated with ^{WT}LplA expression level (data not shown). In order to achieve specific lipoylation inside mammalian cells, we needed to control the expression level of mCherry-WTLplA. For intracellular lipoylation, the non-specific labeling is such a

problem that we needed to use tetracycline-controlled expression system to consistently achieve the low expression required for specific labeling of LAP peptides (Figure 2-12B).

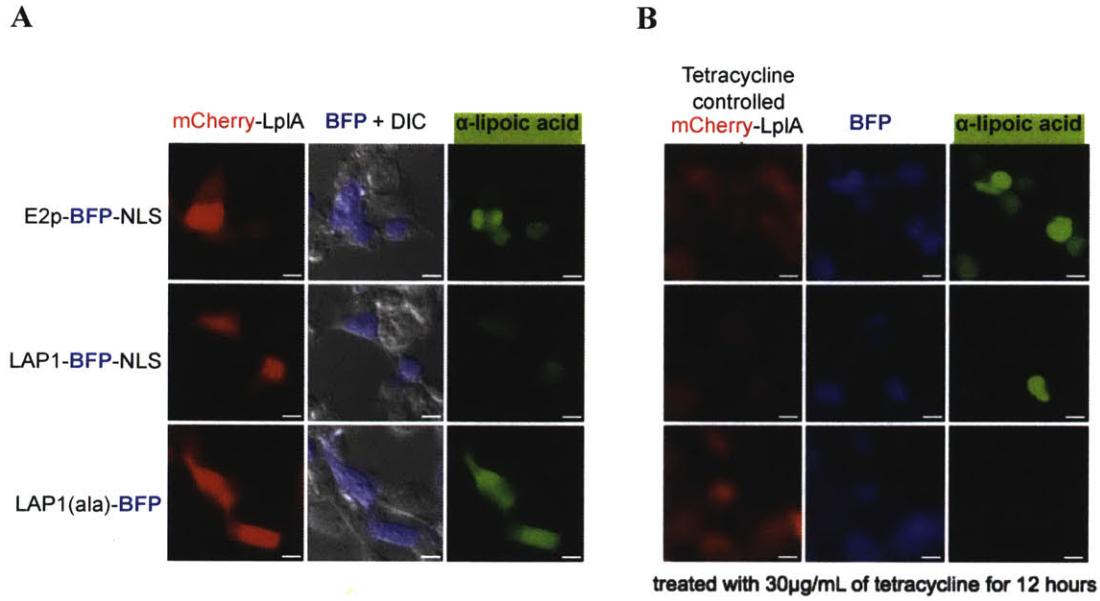


Figure 2-12. Intracellular expression of ^{WT}LplA must be controlled in order to achieve specific lipoylation. (A) Intracellular non-specific lipoylation of E2p-BFP-NLS, LAP1-BFP-NLS and LAP1(ala) when ^{WT}LplA is overexpressed. Lipoylation was performed in live cells and detected using immunofluorescence staining post-fixation. **(B)** Specific lipoylation of E2p-BPF-NLS and LAP1-BFP-NLS is achieved when using tetracycline-inducible expression plasmid (tet-controlled-^{WT}LplA); under these conditions, LAP1(ala) is not lipoylated.

^{W37V}LplA gives higher labeling and sensitivity but ^{W37I}LplA is the more specific ligase

The previous result was disheartening because we realize that when a labeling method is restricted a particular cell line (in this case a stable cell line expressing the tetracycline repressor) there is a huge barrier to adoption by outside labs; an example of this is the limited use of DHFR labeling until the requirement for DHFR- cell lines was relieved by trimethoprim derivatives[30,31]. Fortunately, when we tested the intracellular expression of ^{W37V}LplA and ^{W37I}LplA we found that these enzymes have lower intrinsic intracellular expression than the wild-type ligase. More importantly, we were able to control expression level of these enzymes by simply titrating the amount of plasmid we transfected into the cells (Figure 2-13A). We therefore decided to test coumarin labeling using both ^{W37V}LplA and ^{W37I}LplA under high expression conditions (200 ng transfected DNA) and low expression conditions (20 ng transfected DNA). For each of the four experimental conditions we used LAP2-YFP substrate as a readout of specific labeling

(Figure 2-13B, upper panels). We additionally tested each condition using LAP2(ala)-YFP (Figure 2-13B, lower panels), which allowed us to compare specific vs. non-specific labeling. With both ligases, we observed specific coumarin labeling of intracellular LAP2-YFP at low expression levels. However, when overexpressed, we observed some nonspecific labeling with ^{W37V}LplA (Figure 2-13). We quantified the imaging data from Figure 2-13B and plotted the ratio of coumarin labeling to LAP2-YFP or LAP2(ala)-YFP expression in Figure 2-13C. From these plots, we concluded that ^{W37V}LplA gives the best coumarin signal to background when its expression level could be controlled, but ^{W37I}LplA would be the preferred ligase when ligase overexpression cannot be avoided.

Our original hypothesis for the nonspecific coumarin signal observed with ^{W37V}LplA was that the enzyme might be retaining the coumarin-AMP ester in the active site, as biotin ligase is known to do with biotin-AMP[32], which would result in increasing background signal when the enzyme is over-expressed. This theory also explains why we do not see this behavior with ^{W37I}LplA, which has a much higher K_m for coumarin. To support this claim of non-covalent probe-AMP retention, we demonstrated that nonspecific labeling follows the localization of over-expressed ^{W37V}LplA (Figure 12-14).

However, these results do not reveal whether the coumarin signal is due to covalent enzyme self-modification or non-covalent probe retention. When we performed in-gel fluorescence analysis on *in vitro* reactions with highly concentrated enzyme, we found that there appeared to be covalent modification of the enzyme that was present even after boiling the samples (Figure 12-15). We performed in-gel trypsin digest of the samples and submitted the sample to mass spectrometry, hoping to determine whether the enzyme was being self labeled at a specific residue or non-discriminately. However, the data obtained from this experiment was inconclusive (data not shown). This curiosity was not investigated further, since specific labeling was achieved with optimization of ligase expression conditions.

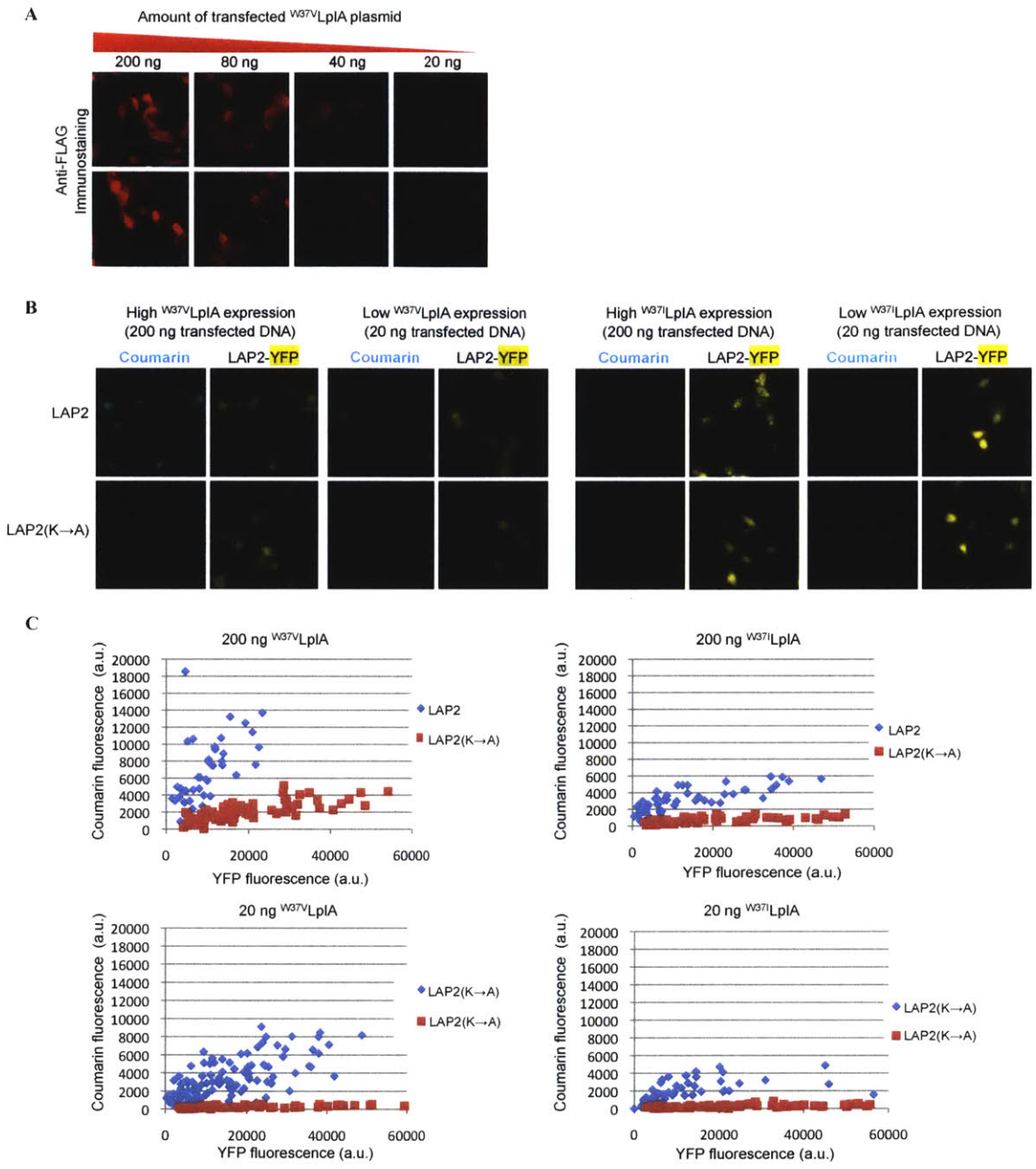


Figure 2-13. Comparison of intracellular expression and coumarin ligation by ^{W37V}LplA and ^{W37I}LplA. (A) Expression level of ^{W37V}LplA can be controlled through plasmid titration (similar results for ^{W37I}LplA). **(B)** Images show coumarin labeling specificities at different enzyme expression levels. Coumarin labeling signal is shown with LAP2-YFP (specific labeling) and LAP2(ala)-YFP (non-specific labeling). **(C)** Quantification of imaging data in **(B)**. Coumarin labeling intensity was plotted against YFP intensity (reflecting LAP2 or LAP2(ala) expression level). Data shown for >60 cells for each condition.

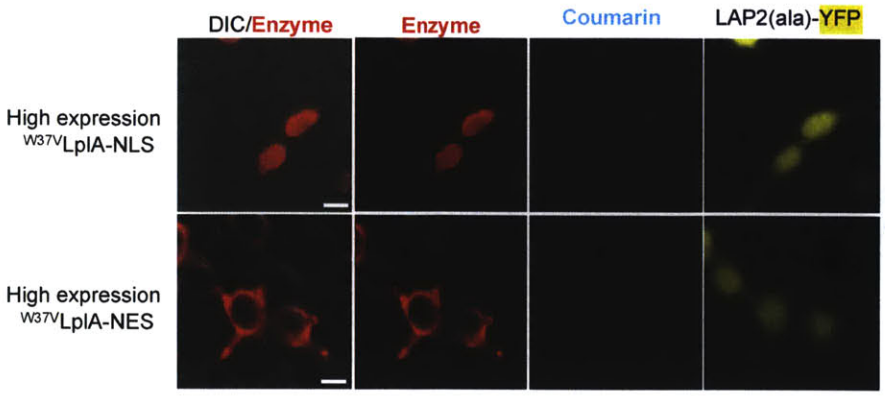


Figure 2-14. Non-specific coumarin signal follows ^{W37V}LplA expression pattern when the enzyme is overexpressed. HEK cells expressing whole-cell LAP2(ala)-YFP were labeled using overexpressed ^{W37V}LplA targeted to the nucleus or the cytosol. Non-specific coumarin labeling pattern follows enzyme expression level as detected by anti-FLAG staining after fixation.

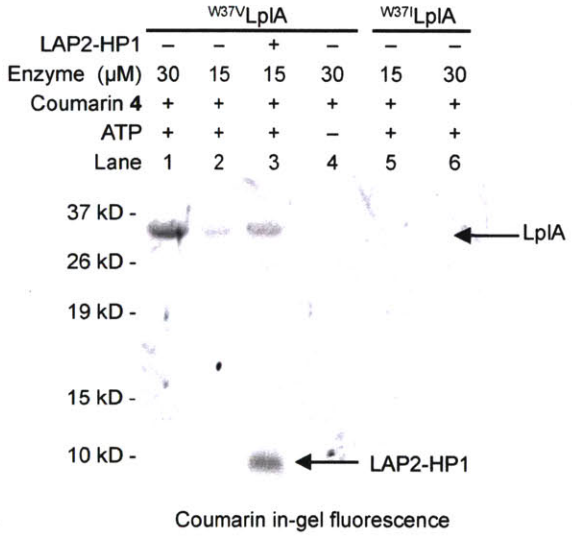


Figure 2-15. In-gel coumarin fluorescence indicates covalent coumarin self-labeling with high enzyme concentrations. In vitro reactions containing 15 or 30 μM of the indicated coumarin ligase, 150 μM LAP2-HP1 (where indicated), 500 μM coumarin **4**, and ATP were incubated for 3 hours. Reactions were quenched, boiled, and run on an SDS-PAGE gel. Coumarin signal on ^{W37V}LplA persists after boiling, indicating a covalent modification with coumarin probe.

We next set out to determine which LAP peptide works best for intracellular protein labeling applications. Although we had already demonstrated that LAP2 could be labeled as a fusion to YFP, we compared the coumarin labeling signal obtained with the protein substrate E2p with the four best LAP peptides from yeast-display evolution: LAP4.1, LAP4.2, LAP4.3D, and LAP2[24]. We expressed each substrate as a fusion to YFP

along with coumarin ligase and then performed a 10-minute coumarin labeling protocol. We analyzed the imaging data and plotted coumarin labeling vs. substrate expression level (Figure 2-16). LAP2 was the only peptide that gave improved coumarin labeling signal over E2p. Therefore, for all future applications of coumarin labeling, we used LAP2 as the peptide tag.

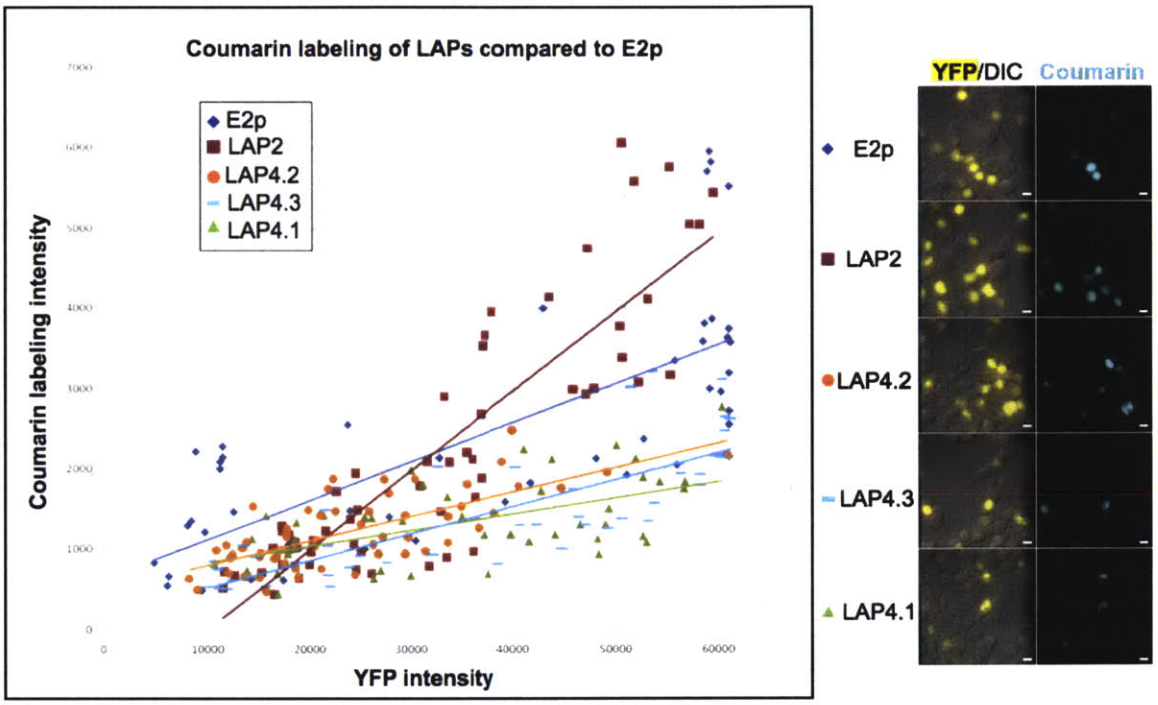


Figure 2-16. LAP2 gives best coumarin labeling signal in intracellular coumarin labeling assay. Ten-minute coumarin labeling was performed in HEK cells expressing E2p or one of the evolved LAP peptides fused to YFP. For at least 50 single cells for each condition, coumarin labeling intensity was plotted against YFP expression level. LAP2 was the only evolved peptide that gave improved coumarin labeling over E2p.

Characterization of coumarin labeling inside mammalian cells

We characterized intracellular coumarin labeling with ^{W37V}LplA in HEK cells expressing LAP2-YFP. We performed the labeling using just 10 minutes of incubation with coumarin-AM_2 followed by probe washout over an hour. Negative controls where we mutate the target lysine residue of LAP2 or replace the coumarin ligase with ^{WT}LplA show no coumarin labeling (Figure 2-17A). These controls demonstrate that the intracellular labeling is site-specific and dependent on the activity of our engineered coumarin ligase. We also demonstrated specific coumarin labeling of LAP2 fusions targeted to the nucleus, cytosol, and plasma membrane as well as LAP2 fused to MAP2

(microtubule-associated protein 2) and the cytoskeletal proteins vimentin and β-actin (Figure 2-17B). The LAP2 peptide is also transposable, being efficiently labeled at N-terminal, C-terminal, and internal fusion sites. We also estimated the intracellular coumarin ligation yield and determined that, using the standard 10 minute labeling protocol, 25 ± 13% of LAP2-YFP was labeled; extending the labeling time to 1 hour increased labeling yield to 53 ± 25%. We call this robust and generalizable intracellular protein labeling method PRIME (PRobe Incorporation Mediated by Enzymes).

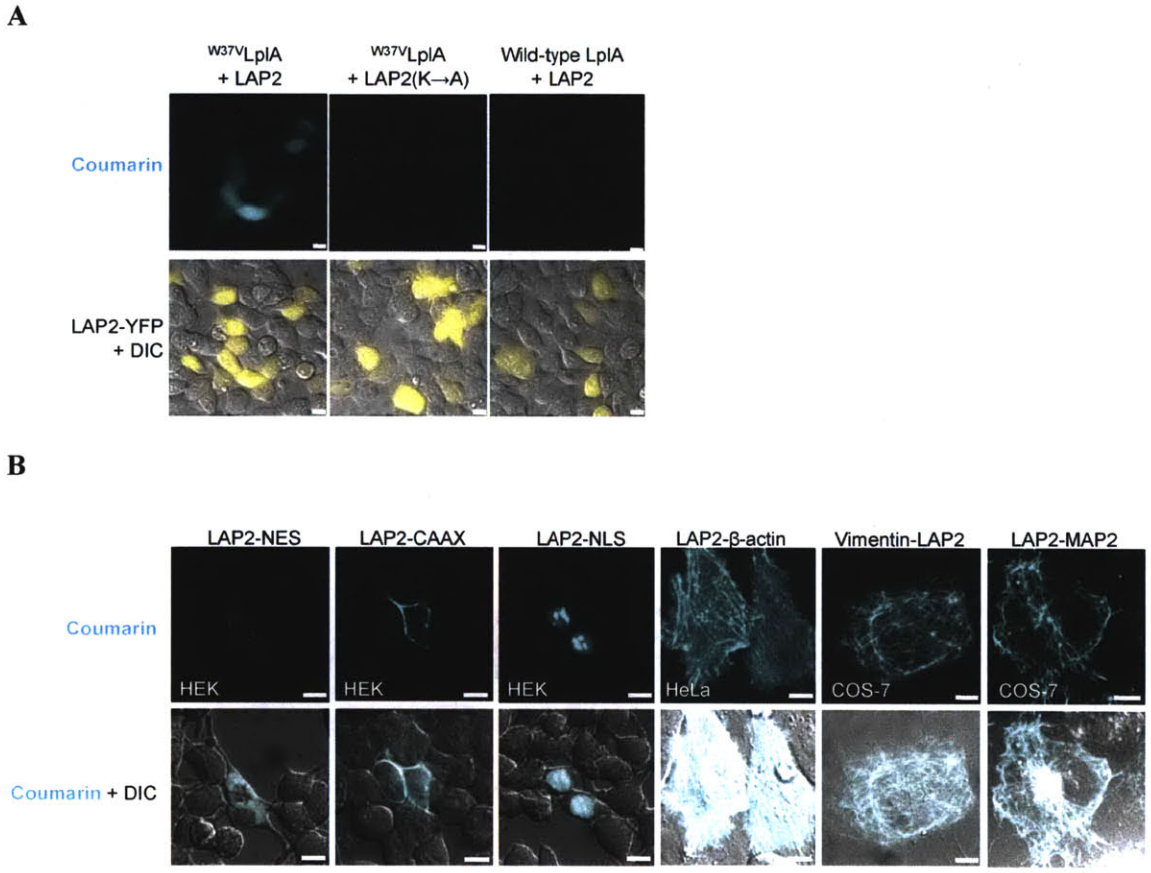


Figure 2-17. Characterization of intracellular coumarin labeling of LAP2-tagged substrates using ^{W37V}LplA. (A) PRIME labeling of LAP2-YFP in live HEK cells. Negative controls using an alanine mutant of LAP2 and wild-type LplA are also shown. **(B)** PRIME labeling of different LAP2-tagged proteins in living cells. NES = Nuclear Export Sequence. CAAX = Farnesylation motif. NLS = nuclear localization sequence. MAP2 = Microtubule-associated Protein 2.

Comparison of coumarin PRIME and FlAsH labeling methods

After we fully characterized our new PRIME labeling method inside living mammalian cells, we wanted to compare this new method to existing intracellular protein labeling methods. We argue that best comparison would be to FlAsH, as it is the only existing labeling method that uses a small peptide tag and works inside cells[33,34]. We wanted to compare the methods in terms of specificity, sensitivity, toxicity, and overall labeling time. In order to perform a direct comparison of the labeling methods, we prepared a single nuclear-targeted construct containing LAP2, the optimized 12-amino acid tetracysteine peptide (FLNCCPGCCMEP)[34] and mCherry for visualization (Figure 2-18). Since both recognition peptides are localized to the nucleus, we can easily distinguish and quantify specific (nuclear) labeling and nonspecific (cytosolic) labeling.

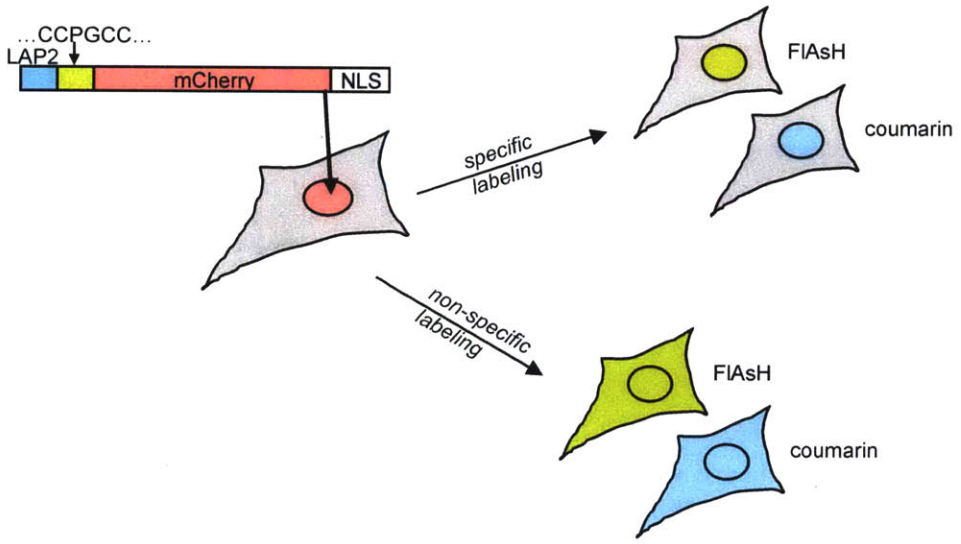


Figure 2-18. Scheme for direct comparison of PRIME and FlAsH labeling. Both peptide tags are fused to a nuclear localized mCherry protein for visualization. Because both substrates are nuclear, specific labeling with either method will result in nuclear signal in either coumarin (PRIME) or fluorescein (FlAsH) channels. Any cytosolic signal is the result of non-specific labeling or dye retention.

For the specificity comparison, we transfected HEK cells with this construct as well as ^{W37V}LplA and performed the standard 10 minute coumarin labeling protocol, followed by 30 minutes of FlAsH labeling (standard labeling protocol[34]). We found that coumarin labeling was very specific, giving primarily nuclear labeling signal that co-localized with substrate expression (Figure 2-19). With FlAsH labeling, on the other hand, we observed two types of nonspecific labeling: untransfected cells labeled by the FlAsH biarsenical dye (Figure 2-19, red arrows) and cytosolic background in cells

expressing the nuclear-targeted substrate (white arrow). Note that we also tried the reverse labeling protocol where we first perform FlAsH followed by coumarin labeling and noted no difference in specific labeling intensities. Unfortunately, we found that the FlAsH background was much higher when we followed with coumarin labeling (data not shown), which we suspect is due to endocytosis of excess FlAsH dye during the coumarin labeling 37°C incubation steps. For quantitative comparison of these imaging data, we plotted the coumarin and fluorescein nuclear vs. cytosolic fluorescence intensities for n = 225 individual cells (Figure 2-19B). A high nuclear to cytosolic ratio indicates a highly specific labeling method, and from the plot we concluded that coumarin labeling is more specific than FlAsH labeling.

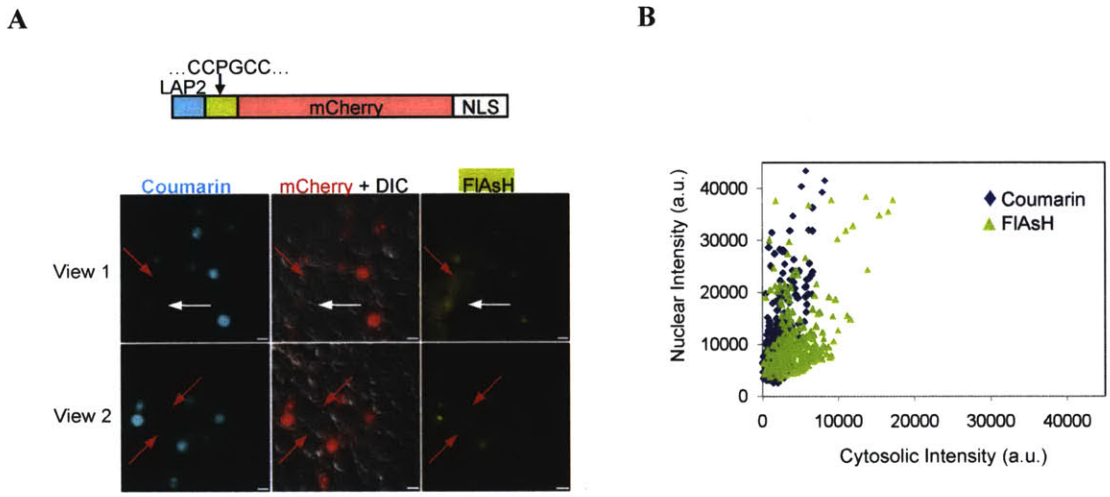
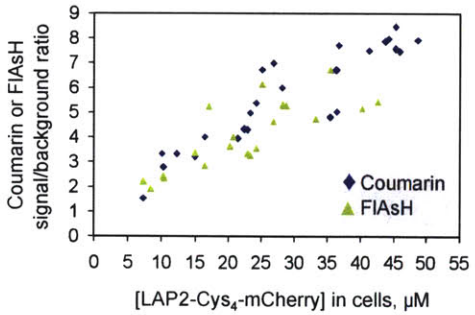


Figure 2-19. PRIME labeling is more specific than FlAsH labeling. (A) HEK cells expressing the LAP2-Cys4-mCherry NLS were labeled with coumarin followed by FlAsH labeling. Red arrows indicate untransfected cells with FlAsH dye retention. White arrows indicate transfected cells with obvious cytosolic background. **(B)** Imaging data from **(A)** was processed and the nuclear vs. cytosolic labeling intensities plotted for both coumarin and FlAsH for ~275 single cells. A steeper slope indicates a more specific labeling method.

We used the same imaging data to determine the sensitivity of the two labeling methods. First, we used the wedge method[33] to calculate the substrate concentration in individual cells. We then were able to plot the signal/background ratio (from data in 2-19C) vs. the substrate concentration (Figure 2-20A). For both coumarin and FlAsH labeling, we can detect as little as 10 μM target protein with signal to noise >2:1. Finally, we compared the toxicity of the two labeling protocols using an assay for mitochondrial

respiration and determined that FlAsH is more toxic than PRIME labeling (Figure 2-20B).

A



B

	% Viability	
Condition	Untransfected	Transfected
Untreated	100 ± 6	94 ± 4
10 min of 20 μM coumarin-AM_2	98 ± 5	94 ± 7
6 hr of 20 μM coumarin-AM_2	93 ± 6	88 ± 1
30 min of 2.5 μM FlAsH-EDT_2	92 ± 5	80 ± 8
6 hr of 2.5 μM FlAsH-EDT_2	82 ± 5	71 ± 8
2 hr of 100 mM NaN_3	73 ± 3	69 ± 3

Figure 2-20. FlAsH and PRIME methodologies are similarly sensitive, but FlAsH is more toxic to cells. (A) Intracellular substrate concentrations were estimated using the wedge method. For ~50 single cells the coumarin and FlAsH nuclear/cytosolic intensity ratios were plotted against the absolute concentration of the mCherry-fused substrates. When substrate levels are >10 μM, both methods produce labeling signal >2:1. **(B)** An assay for mitochondrial respiration showed that FlAsH is more toxic than the PRIME methodology.

LplA has decreased activity in oxidizing compartments of the cell

PRIME labeling is the first intracellular protein labeling method that combines a peptide tag with high specificity and fast labeling. While we demonstrated effective coumarin labeling in the cytosol, nucleus, and at the plasma membrane of living cells, we also wanted to test labeling of proteins expressed in the secretory pathway. However, when we targeted the ^{W37V}LplA to the endoplasmic reticulum (ER) we did not see any labeling of LAP2-YFP-ER or of LAP2 fused to a number of transmembrane proteins (data not shown). We initially suspected that the coumarin-AM_2 probe concentration is not high enough inside the ER to achieve good labeling; but we did not observe coumarin labeling in the ER with increased the probe concentration or when we expressed esterases targeted to the endoplasmic reticulum (data not shown). To determine whether the lack of activity in the ER was probe-specific, we attempted lipoylation of a LAP2-tagged surface protein using ^{WT}LplA targeted to the endoplasmic reticulum. The LAP2-fusion gets lipoylated by ER targeted enzyme as it is being trafficked through the ER, and we can detect the lipoylation using immunofluorescence at the cell surface. The extent of

lipoylation using an ER targeted enzyme (Figure 2-21, bottom) is extremely low compared to lipoylation with enzyme supplied exogenously (Figure 2-21, top).

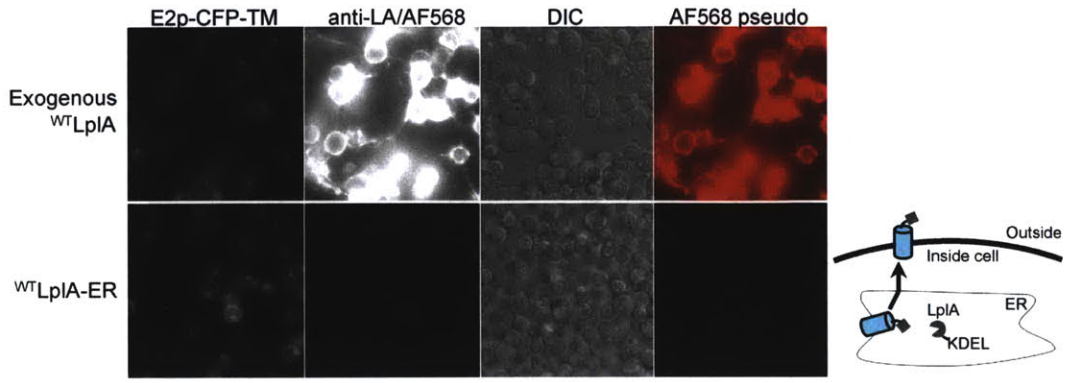


Figure 2-21. Lipoylation activity of ^{WT}LplA is decreased in the endoplasmic reticulum. Top, LAP2-CFP is expressed on the mammalian cell surface. When the LAP2 is lipoylated using purified ^{WT}LplA, lipoylation signal is extremely strong (top). When the LAP2 is lipoylated using ER-targeted ^{WT}LplA as it is trafficked to the cell surface, lipoylation is barely detectable (bottom).

This tells us that the inherent activity of LplA is decreased in the endoplasmic reticulum. Since ^{WT}LplA has 6 cysteine residues, we hypothesized that aberrant disulfide formation or disulfide-induced misfolding could reduce activity in oxidizing compartments like the ER. However, when we made mutants of ^{WT}LplA wherein the cysteine residues were mutated to serines, we did not see any improvements in ER activity (data not shown). Another potential problem could be the high calcium concentration in the ER, which we found inhibits LplA activity *in vitro* (data not shown). Without fully understanding why LplA has decreased activity in the ER, and without clear avenues for rational design, we decided to move forward with *in vitro* evolution to select for a probe ligase with increased activity in the endoplasmic reticulum and other oxidizing cellular compartments (described in Chapters 6-8).

Conclusions

In this chapter we have described the rational engineering of a coumarin fluorophore ligase and the development of an intracellular protein labeling method (PRIME). The method is based on E. coli lipoic acid ligase, which had previously been used by the lab as a method for cell-surface protein labeling with an alkylazide probe. For intracellular labeling, we envisioned using direct ligation of a small coumarin fluorophore. We used rational mutagenesis to engineer a coumarin binding pocket in the enzyme and then confirmed that the ligation was still highly sequence specific both in vitro and on mammalian cell lysate. After optimization of coumarin probe delivery, intracellular enzyme expression, and selecting the best peptide tag, we were able to achieve site-specific coumarin labeling of intracellular proteins.

This method was robust and sufficiently generalizable that we were able to label proteins in multiple subcellular compartments. When compared to existing labeling methods, the PRIME method uniquely provides the combination of a peptide tag, high labeling specificity, and compatibility with intracellular labeling. However, there are two fundamental problems with this intracellular labeling method. First, the coumarin ligation kinetics are not optimal when compared to the endogenous reaction or even the ligation of other unnatural probes. Second, we do not see unnatural probe ligation activity in oxidizing compartments such as the endoplasmic reticulum. Attempts at correcting these problems through rational design and mutagenesis were not successful. Therefore, we used *in vitro* evolution techniques to select for ligases with not only better ligation kinetics, but also improved activity in the secretory pathway. This work is presented in Chapters 6-8 of this thesis.

Experimental

Protein expression and purification.

LplA, E2p, and LAP-Heterochromatin Protein 1 (HP1) fusion proteins were expressed in BL21-DE3 bacteria with IPTG induction as previously described[1,21]. After Ni-NTA purification, LplA enzymes were stored in 20 mM Tris pH 7.5, 1 mM DTT and 10% glycerol in aliquots at -80 °C. E2p and LAP-HP1 proteins were stored in PBS pH 7.4. Protein concentrations were determined using the bicinchoninic acid assay (Pierce).

In vitro coumarin ligation reactions (Figure 2-6, 2-7)

General reaction conditions using purified proteins were as follows: 100–150 μM LAP or E2p substrate was incubated with 1–2 μM LplA, 500 μM coumarin probe, 5 mM ATP, and 5 mM $Mg(OAc)_2$ in 25 mM Na_2HPO_4 pH 7.2 at 30 °C for various lengths of time. Reactions were quenched with EDTA (C_f 100 mM) and analyzed on a Varian Prostar HPLC using a reverse-phase C18 Microsorb-MV 100 column (250 × 4.6 mm). Chromatograms were recorded at 210 nm. For analysis of E2p and LAP4.3D-HP1 (LAP4.3D sequence: GFEHEKVWYDLDA), we used a 20-minute gradient of 40–57% acetonitrile in water with 0.1% trifluoroacetic acid at 1 ml/minute flow rate. Both E2p and LAP4.3D-HP1 had retention times of 12 minutes, which increased to 15-16 minutes after ligation to coumarin probes. For analysis of reactions using the LAP2 synthetic peptide (GFEIDKVWYDLDA), we used a 10-minute gradient of 30–60% acetonitrile in water with 0.1% trifluoroacetic acid at 1 ml/minute flow rate. Retention times are shown in Figure 2-7. Percent conversions were calculated by dividing the product peak area by the sum of (product + starting material) peak areas.

Initial alanine scanning was performed using lysate from bacterial cells transformed with LplA expression plasmids. 5 ml cultures at OD_{600} of 0.5 were induced with 424 μM IPTG for 4 hours, before harvesting and lysis in 100 μl BPER reagent (Thermo Scientific). 10 μl bacterial lysate was incubated overnight at 30°C with 100 μM E2p, 500 μM probe, 5 mM ATP, 5 mM $Mg(OAc)_2$ in 25 mM Na_2HPO_4, and LplA mutant/probe pairs giving any product were identified from HPLC analysis.

Secondary screening of W37 LplA mutants was performed as follows. 2 μM purified enzyme was combined with 100 μM E2p and incubated overnight under the general reaction conditions. Mutant/probe pairs showing activity under these conditions were assayed in a tertiary screen using 1 μM purified enzyme, 100 μM LAP4.3D-HP1, and a shorter reaction time of 40 minutes to distinguish the activity of different mutants. The data from the secondary and tertiary screens were used to generate the activity matrix shown in Figure 2-6; the percent product conversions from the tertiary screen are presented in the table in Figure 2-6.

For Figure 2-7 5 μM LplA (W37V, W37I, or wild-type) was combined with 500 μM coumarin **4** and 150 μM LAP2 peptide, which was synthesized by the Tufts Peptide Synthesis Core Facility, under the general reaction condition. Reaction time was 2 hours at 30 °C.

Mass spectrometric analysis (Figure 2-7)

Starred peaks from HPLC traces shown were manually collected and injected onto an Applied Biosystems 200 QTRAP mass spectrometer. The flow rate was 10 μl/min and mass spectra were recorded under the positive-enhanced multi-charge mode.

Kinetic measurements (Figure 2-8)

K_m and k_{cat} values were determined using coumarin **4** probe and LAP2 peptide. 2 μM ^{W37V}LplA or ^{W37I}LplA was combined with 150 μM LAP2, 1 mM ATP, and 5 mM $Mg(OAc)_2$ in 25 mM Na_2HPO_4 pH 7.2. Coumarin **4** concentration was varied between 12.5 and 500 μM for the ^{W37V}LplA measurement, and between 16 and 1600 μM for ^{W37I}LplA. Small aliquots were removed at different time intervals from the reaction mixture at 30°C and quenched with 100 mM EDTA (final concentration). The time intervals, ranging from 1–192 minutes, were determined empirically for each coumarin **4** concentration, such that the product conversion at the final initial rate measurement for each coumarin **4** concentration did not exceed 10%. Samples were analyzed by HPLC using a LAP2 peptide gradient as described under "*In vitro* coumarin ligation reactions". To compensate for differences in extinction coefficients of LAP2 and LAP2-coumarin **4**

conjugate, a calibration curve was obtained that correlated the ratio of integrated HPLC peak areas of LAP2:LAP2-coumarin **4** conjugate to the actual peptide concentration ratio. The amount of product obtained at each time point was plotted against time to obtain the initial velocity (V_0) for each concentration of coumarin **4**. V_0 values were then plotted against coumarin **4** concentration and fit to the Michaelis-Menten equation ($V_0 = V_{max}$ [coumarin **4**]/(K_m + [coumarin **4**])), using Origin 6.1 software. From the V_{max} value, k_{cat} was calculated using the equation $V_{max} = k_{cat}$ $[E]_{total}$.

Mammalian cell culture

HEK, HeLa, and COS-7 cells were cultured in Dulbecco's modified Eagle medium (DMEM; Cellgro) supplemented with 10% v/v fetal bovine serum (FBS; PAA Laboratories) at 37°C and under 5% CO_2. CHO K1 cells were cultured in DMEM supplemented with 10% v/v FBS, 70 mg/l L-proline and 220 mg/l sodium pyruvate. NIH-3T3 cells were cultured in DMEM supplemented with 10% v/v newborn calf serum (PAA Laboratories) at 37 °C and under 5% CO_2. For imaging, cells were plated as a monolayer on glass cover slips. Adherence of HEK cells was promoted by precoating the coverslip with 50 μg/ml fibronectin (Millipore).

Coumarin labeling of mammalian cell lysate (Figure 2-9)

HEK cells expressing LAP2-YFP (or LAP2(K→A)-YFP) were lysed under hypotonic conditions in 1 mM HEPES pH 7.5 with 5 mM $MgCl_2$, 1 mM phenylmethylsulphonyl fluoride, and protease inhibitor cocktail (Calbiochem). After three cycles of freeze-thaw, cells were mixed by vortexing for 2 minutes. Lysate was cleared by centrifugation and stored in aliquots at −80 °C. To label with coumarin or lipoic acid, lysates were incubated with 200 nM LplA, 500 μM coumarin **4** or 250 μM lipoic acid, 1.5 mM ATP, and 5 mM $Mg(OAc)_2$ in 25 mM Na_2HPO_4 pH 7.2. After overnight incubation at 30 °C, reactions were boiled in protein loading buffer containing 2-mercaptoethanol for 7 minutes, then separated on a 12% SDS-PAGE gel. Coumarin fluorescence was visualized on an Alpha Innotech ChemiImager 5500 instrument using 365 nm UV light for excitation.

Western blotting of mammalian cell lysate (Figure 2-9)

Lysates were labeled with coumarin **4** or lipoic acid as described in the Main Methods, boiled, run on a 12% SDS-PAGE gel, and transferred onto a nitrocellulose membrane. The membrane was blocked with 3% BSA solution in TBS + 0.5% Tween-20 (TBS-T) for 1 hour. For anti-lipoic acid blotting, the membrane was incubated with rabbit polyclonal anti-lipoic acid antibody (Calbiochem) in 3% BSA in TBS-T at 1:1000 dilution for 40 minutes, followed by three five-minute washes with TBS-T. The membrane was then incubated with anti-rabbit horseradish peroxidase conjugate (Bio-Rad) in 3% BSA in TBS-T at 1:20000 dilution for 40 minutes, followed by washes with TBS-T. The blots were developed using Supersignal West Femto substrate (Pierce) and visualized on an Alpha Innotech ChemiImager 5500 instrument.

Fluorescence imaging

Cells were imaged in Dulbecco's Phosphate Buffered Saline (DPBS) in epifluorescence or confocal modes. For epifluorescence imaging, we used a Zeiss AxioObserver inverted microscope with a 40x or 63x oil-immersion objective. Coumarin (400/20 excitation, 425 dichroic, 435/30 emission), YFP (493/16 excitation, 506 dichroic, 525/30 emission), mCherry (570/20 excitation, 585 dichroic, 605/30 emission) and differential interference contrast (DIC) images were collected and analyzed using Slidebook software (Intelligent Imaging Innovations). For confocal imaging, we used a Zeiss Axiovert 200M inverted microscope with a 40x oil-immersion objective. The microscope was equipped with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), and 561 nm (DPSS) lasers (all 50 mW). Coumarin (405 laser excitation, 445/40 emission), YFP (491 laser excitation, 528/38 emission), mCherry (561 laser excitation, 617/73 emission), and DIC images were collected using Slidebook software. Fluorescence images in each experiment were normalized to the same intensity ranges. Acquisition times ranged from 10–1000 milliseconds.

General protocol for coumarin PRIME labeling in living cells. (Figure 2-13, 2-16, 2-17) HEK cells were transfected at ~70% confluency with expression plasmids for

^{W37V}LplA (20 ng for a 0.95 cm^2 dish) and the E2p, LAP4.x or LAP2 fusion protein of interest (400-600 ng) using Lipofectamine 2000 (Invitrogen). 8–24 hours after transfection, cells were incubated with 20 μM coumarin-AM_2 in serum-free DMEM (Dulbecco's modified Eagle medium, Cellgro) for 10 minutes at 37 °C. Note: 0.1% w/v Pluronic F-127 (Invitrogen) can be optionally added to the labeling solution to give more even coumarin distribution in cells. The media was then replaced 3–4 times over 30–60 min at 37 °C, with DMEM supplemented with 10% FBS (Fetal Bovine Serum), to wash out excess coumarin.

For Figure 2-16, HEK cells were transfected and labeled with coumarin as described above. Using Slidebook software, the average mean fluorescence intensity for each imaging channel was determined for >60 single cells for each condition. Plots show coumarin labeling intensity plotted against YFP intensity (reflecting substrate-YFP expression level) for >60 single cells.

Comparison of ^{W37V}LplA and ^{W37I}LplA labeling specificity in cells (Figure 2-13)

The indicated quantities of FLAG-^{W37V}LplA or FLAG- ^{W37I}LplA plasmid DNA (plus 400 ng LAP2-YFP DNA or LAP2(K→A)-YFP) were transfected into HEK cells. Coumarin labeling was performed as previously described. Epifluorescence images were taken at 40x magnification. Using Slidebook software, the average mean fluorescence intensity for each imaging channel was determined for >60 single cells for each condition. Plots show coumarin labeling intensity plotted against YFP intensity (reflecting LAP2-YFP or LAP2(K→A)-YFP expression level) for >60 single cells.

Determining source of background coumarin labeling (Figure 2-14)

HEK cells were transfected with either 200 ng of ^{W37V}LplA alone or co-transfected with 400 ng of LAP2-YFP or LAP2(K→A)-YFP. Coumarin labeling was performed as previously described. Plots show quantification of coumarin labeling vs. YFP expression from average mean fluorescence intensities, which were collected using Slidebook software as described above.

Images show HEK cells that were transfected with 200 ng of ^{W37V}LplA targeted to the nucleus or the cytosol. Where indicated, enzymes were co-transfected with LAP2(K→A)-YFP. After coumarin labeling was performed, cells were fixed were fixed using 3.7% formaldehyde in DPBS followed by cold precipitation with methanol. Subcellular localization of FLAG-tagged LplA was confirmed using immunofluorescence staining with 4 μg/ml anti-FLAG M2 antibody (Stratagene) in 1% BSA in DPBS for 30 minutes. After 15-minute washes, cells were incubated with 4 μg/ml goat anti-mouse antibody conjugated to Alexa Fluor 568 (Invitrogen) in 1% BSA in DPBS for 30 minutes, then washed and imaged.

Quantification of labeling yield

HEK cells were transfected with ^{W37V}LplA or ^{W37I}LplA and LAP2-YFP plasmids as described. 12 hours after transfection, cells were labeled with coumarin-AM_2 as described above for 10 or 60 minutes. After washing, epifluorescence images were acquired in DPBS at 40x magnification. We used the "wedge method"[33] to estimate absolute concentrations of coumarin and LAP2-YFP inside single cells. A wedge-shaped microchamber was constructed from three glass coverslips. The length along the x direction was 5 mm and the height of the chamber (z-direction) increased linearly from 0 to 150 μm. The chamber was filled with 9 μM purified YFP and 12 μM purified LAP2-coumarin-**4** conjugate in PBS pH 7.4. The fluorescence of the wedge was imaged under conditions identical to those used for cellular imaging. We assumed an average cell thickness of 5 μm and therefore interpolated to the region of the wedge with thickness of 5 μm and used YFP and coumarin fluorescence intensities measured there as reference standards for their concentrations in single cells. The ratio [coumarin]/[LAP2-YFP] represented coumarin ligation yield in cells. Data for each enzyme and coumarin incubation time (10 or 60 minutes) were obtained from ~40 cells.

Intracellular FlAsH and coumarin labeling (Figure 2-19)

HEK cells were co-transfected with LAP2-Cys_4-mCherry-NLS and ^{W37I}LplA. 24 hours after transfection, cells were labeled with coumarin-AM_2 for 10 minutes as described above. After five washes in growth medium, cells were cooled to room temperature, and

labeled with 2.5 μM FlAsH-EDT_2 (Invitrogen) for 30 minutes at 25 °C in serum-free DMEM. The cells were rinsed twice with 250 μM BAL wash buffer (Invitrogen). Labeled cells were imaged in epifluorescence mode, as described above (FlAsH filter set: 493/16 excitation, 488 dichroic, and 525/30 emission).

Comparison of labeling specificities (Figure 2-19)

Images obtained as described above were analyzed as follows. Cells with signal to background ratio >2 (in the fluorescein or coumarin channel) were selected for analysis. For each of these cells, two regions of interest (ROIs) were selected, one in the nucleus and one in the cytosol. The background-corrected mean fluorescence intensity was determined for each ROI using Slidebook. Excel was used to plot the nuclear versus cytosolic fluorescence intensity for each cell, in the FlAsH or coumarin channel. We note that background FlAsH labeling in the cytosol was still observed when we increased the wash time from 30 minutes to 1.5 hours.

Quantification of labeling sensitivities (Figure 2-20)

Images obtained as described above were analyzed as follows. We used the "wedge method"[33] to estimate absolute concentrations of intracellular LAP2-Cys_4-mCherry-NLS. A wedge-shaped microchamber was constructed from three glass coverslips. The length along the x direction was 5 mm and the height of the chamber (z-direction) increased linearly from 0 to 150 μm. The chamber was filled with 20 μM purified mCherry protein in PBS pH 7.4. The fluorescence of the wedge was imaged under conditions identical to those used for cellular imaging. We assumed an average cell thickness of 5 μm and therefore interpolated to the region of the wedge with thickness of 5 μm and used mCherry intensities measured there as reference standards for the concentration in single cells. Intracellular LAP2-Cys_4-mCherry-NLS concentrations ranged from 3 μM to >1 mM. To analyze coumarin labeling, cells showing coumarin signal to background (background from untransfected cells) >3:1 were selected for analysis. To analyze FlAsH labeling, cells were selected with signal to background (from untransfected cells) >3:1 in *both* the fluorescein and mCherry channels, to exclude untransfected cells that were non-specifically labeled by FlAsH. For each cell, the background-corrected nuclear/cytosolic

intensity ratio was calculated using Slidebook. Excel was used to plot these ratios as a function of intracellular mCherry concentration.

Determination of labeling toxicity (Figure 2-20)

HEK cells were analyzed in 96-well plates. Transfected cells expressing ^{W37I}LplA and LAP2-Cys_4-mCherry-NLS were labeled 24 hours post-transfection using the protocols described above. Sodium azide treatment was performed at 37 °C. After two washes in fresh growth medium, cells were incubated with 0.5 mg/ml MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) for 40 minutes at 37 °C. Medium was removed and precipitate was resuspended in 200 µL DMSO. Absorbance measurements were taken at 570 nm (reference wavelength 670 nm) using a Tecan SAFIRE platereader. Measurements were performed in triplicate. Errors, ± 1 s.d.

Comparison of LplA lipoylation activity in ER and on cell surface (Figure 2-21)

HEK cells were transfected with 600 ng of E2p-CFP-TM and 600 ng of either ^{WT}LplA-NXN (for cis labeling) or ^{WT}LplA-ER (for ER labeling). 12 hours post-transfection cells were labeled in three different ways for comparison: exogenous lipoylation, cis lipoylation and ER lipoylation. For ER lipoylation, cells were incubated with 250 µM lipoic acid for 1 hour, relying on endogenous ATP for lipoylation. For cis lipoylation, cells were incubated with 250 µM lipoic acid, 5 mM $Mg(OAC)_2$, and 1 mM ATP for 1 hour. For exogenous lipoylation, cells were incubated with 250 µM lipoic acid, 5 mM $Mg(OAC)_2$, 1 mM ATP, and 5 µM ^{WT}LplA. For all samples, cells were washed 2 x 10 minutes with DMEM and then lipoylation was detected via live-cell antibody staining. Live-cell immunofluorescence staining was performed with 4 µg/ml anti-lipoic acid antibody (CalBiochem) in 1% BSA in DPBS for 15 minutes. After three quick washes, cells were incubated with 4 µg/ml goat anti-rabbit antibody conjugated to Alexa Fluor 568 (Invitrogen) in 1% BSA in DPBS for 15 minutes, then washed and imaged.

References

1. Fernandez-Suarez, M., *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **2007**, *25*, 1483-7.

2. Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2005**, *2*, 99-104.

3. Lin, C. W.; Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* **2006**, *128*, 4542-3.

4. Green, D. E.; Morris, T. W.; Green, J.; Cronan, J. E., Jr.; Guest, J. R. Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem J* **1995**, *309 (Pt 3)*, 853-62.

5. Cronan, J. E.; Zhao, X.; Jiang, Y. Function, attachment and synthesis of lipoic acid in Escherichia coli. *Adv Microb Physiol* **2005**, *50*, 103-46.

6. Fujiwara, K.; Okamura-Ikeda, K.; Motokawa, Y. Cloning and expression of a cDNA encoding bovine lipoyltransferase. *J Biol Chem* **1997**, *272*, 31974-8.

7. Fujiwara, K.; Takeuchi, S.; Okamura-Ikeda, K.; Motokawa, Y. Purification, characterization, and cDNA cloning of lipoate-activating enzyme from bovine liver. *J Biol Chem* **2001**, *276*, 28819-23.

8. Fujiwara, K., *et al.* Crystal structure of bovine lipoyltransferase in complex with lipoyl-AMP. *J Mol Biol* **2007**, *371*, 222-34.

9. Fujiwara, K., *et al.* Molecular cloning, structural characterization and chromosomal localization of human lipoyltransferase gene. *Eur J Biochem* **1999**, *260*, 761-7.

10. McManus, E.; Luisi, B. F.; Perham, R. N. 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*, 625-37.

11. 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*, 38081-9.

12. Posner, M. G.; Upadhyay, A.; Bagby, S.; Hough, D. W.; Danson, M. J. A unique lipoylation system in the Archaea. Lipoylation in Thermoplasma acidophilum requires two proteins. *FEBS J* **2009**, *276*, 4012-22.

13. Christensen, Q. H.; Cronan, J. E. The Thermoplasma acidophilum LplA-LplB complex defines a new class of bipartite lipoate-protein ligases. *J Biol Chem* **2009**, *284*, 21317-26.

14. Allary, M.; Lu, J. Z.; Zhu, L.; Prigge, S. T. Scavenging of the cofactor lipoate is essential for the survival of the malaria parasite Plasmodium falciparum. *Mol Microbiol* **2007**, *63*, 1331-44.

15. Gunther, S.; McMillan, P. J.; Wallace, L. J.; Muller, S. Plasmodium falciparum possesses organelle-specific alpha-keto acid dehydrogenase complexes and lipoylation pathways. *Biochem Soc Trans* **2005**, *33*, 977-80.

16. Kang, S. G.; Jeong, H. K.; Lee, E.; Natarajan, S. Characterization of a lipoate-protein ligase A gene of rice (Oryza sativa L.). *Gene* **2007**, *393*, 53-61.

17. Zhao, X.; Miller, J. R.; Jiang, Y.; Marletta, M. A.; Cronan, J. E. Assembly of the covalent linkage between lipoic acid and its cognate enzymes. *Chem Biol* **2003**, *10*, 1293-302.

18. Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* **2006**, *1*, 644-8.

19. Yao, J. Z., *et al.* Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J Am Chem Soc*, *134*, 3720-8.

20. Lavis, L. D.; Raines, R. T. Bright ideas for chemical biology. *ACS Chem Biol* **2008**, *3*, 142-55.

21. Baruah, H.; Puthenveetil, S.; Choi, Y. A.; Shah, S.; Ting, A. Y. An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* **2008**, *47*, 7018-21.

22. Sun, W. C.; Gee, K. R.; Haugland, R. P. Synthesis of novel fluorinated coumarins: excellent UV-light excitable fluorescent dyes. *Bioorg Med Chem Lett* **1998**, *8*, 3107-10.

23. 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*, 8212-6.

24. Puthenveetil, S.; Liu, D. S.; White, K. A.; Thompson, S.; Ting, A. Y. Yeast display evolution of a kinetically efficient 13-amino acid substrate for lipoic acid ligase. *J Am Chem Soc* **2009**, *131*, 16430-8.

25. Fujiwara, K., *et al.* Global conformational change associated with the two-step reaction catalyzed by Escherichia coli lipoate-protein ligase A. *J Biol Chem* **2010**, *285*, 9971-80.

26. 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*, 33645-51.

27. Fujiwara, K.; Okamura-Ikeda, K.; Motokawa, Y. Expression of mature bovine H-protein of the glycine cleavage system in Escherichia coli and in vitro lipoylation of the apoform. *J Biol Chem* **1992**, *267*, 20011-6.

28. Fujiwara, K.; Okamura-Ikeda, K.; Motokawa, Y. Lipoylation of acyltransferase components of alpha-ketoacid dehydrogenase complexes. *J Biol Chem* **1996**, *271*, 12932-6.

29. Tsien, R. Y. Fluorescent probes of cell signaling. *Annu Rev Neurosci* **1989**, *12*, 227-53.

30. Miller, L. W.; Sable, J.; Goelet, P.; Sheetz, M. P.; Cornish, V. W. Methotrexate conjugates: a molecular in vivo protein tag. *Angew Chem Int Ed Engl* **2004**, *43*, 1672-5.

31. Miller, L. W.; Cai, Y.; Sheetz, M. P.; Cornish, V. W. In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nat Methods* **2005**, *2*, 255-7.

32. Xu, Y.; Beckett, D. 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*, 7354-60.

33. 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*, 6063-76.

34. Martin, B. R.; Giepmans, B. N.; Adams, S. R.; Tsien, R. Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* **2005**, *23*, 1308-14.

Chapter 3. Application of coumarin PRIME to compartmentalized labeling of intracellular proteins.

The work presented in this chapter has been published in part in C. Uttamapinant, K.A. White, H. Baruah, S. Thompson, M. Fernández-Suárez, S. Puthenveetil, A.Y. Ting A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A*, **2005** *107*, 10914-9.

All MondoA, Mlx, and hexokinase plasmids were generously provided by D.E. Ayer (The University of Utah), who we also thank for helpful advice and troubleshooting tips for the MondoA shuttling assay.

The indicated data in Figure 3-3B is reproduced with permission from J. Dopie, K-P. Skarp, E. K. Rajakylä, K. Tanhuanpää, M. K. Vartiainen. Active maintenance of nuclear actin by importin 9 supports transcription. *Proc Natl Acad Sci U S A,* **2012**, *109,* E544-52.

Introduction

In the previous chapter, we described the development of the coumarin PRIME intracellular protein labeling methodology. This method uses an enzyme to site-specifically label peptide-tagged proteins with a coumarin fluorophore. We demonstrated that the method is fast, specific, and works in many subcellular compartments. One might argue that one drawback of the PRIME labeling method is that labeling requires two components: the enzyme and the LAP-tagged protein of interest. However, we see this feature as a distinct advantage of our method that allows for more sophisticated exploration of biology. A unique feature of PRIME labeling is the ability to use genetically targeted ligases to selectively highlight protein subpopulations.

Protein translocation and selective labeling methods

Temporally-controlled protein translocation is a dynamic process that is involved in many cellular functions[1-6]. Translocation of a single protein, such as an apoptotic factor[4,7], a transcription factor[8-12], or a signaling domain[13-15], can produce a cascade of downstream effects that have significant biological repercussions. The biology of some protein translocation events has been heavily studied; for example, a good deal is known about the effects of apoptotic factors shuttling between the mitochondria and nucleus and gene regulation following the translocation of transcription factors from the cytosol to the nucleus. However, recent results indicate that many proteins that were not previously thought to have compartment-specific functions may behave differently based on subcellular localization. Various labs have presented data that indicate cathepsins[16,17], kinases[18-20], and even cytoskeletal proteins[21-25] may each have nuanced biological functions that are linked to their subcellular localization.

Scientists who want to study the link between a protein's subcellular localization and function face significant challenges. One telling example is transglutaminase 2, which has been found in the nucleus, cytosol, mitochondria, plasma membrane, and the extracellular matrix[26,27]. The nuclear pool of transglutaminase 2 plays a role in gene regulation and histone modification, and these functions have interested groups for years[28-30]. Cell fractionization followed by biochemical assays have been used to

specifically study the nuclear population of transglutaminase 2. However, studying the nuclear biology of transglutaminase 2 inside living cells using traditional imaging methods is difficult, as the nuclear pool of transglutaminase 2 is estimated to be only about 5-7% of the total population. Exploring the specific function, activity, and distinct trafficking of such a small fraction of the entire protein population is extremely difficult using traditional labeling and fluorescence imaging techniques[26]. Since a simple genetically encoded tag such as GFP or SNAP/CLIP would label the whole population, researchers often combine these tools with advanced imaging techniques to study the behavior of proteins in one area of the cell[31].

The most frequently used imaging methods to probe protein exchange or protein trafficking are photobleaching techniques such as fluorescence loss in photobleaching (FLIP) or fluorescence recovery after photobleaching (FRAP)[31,32]. In FLIP, one area of a cell is illuminated with a high power laser, bleaching the fluorophores in that area; concomitant loss of fluorescence in a non-illuminated area of the cell indicates proteins are being rapidly exchanged between the two areas. In FRAP experiments, one area of a cell is selectively photobleached and then subsequent recovery of fluorescence in that area indicates exchange with another population outside the bleached area. However, the bleached population cannot be visualized with these methods, making it impossible to image all labeled proteins. To address this issue, several techniques employ the use of photo-activatable or photo-switchable dyes or FPs[33-37]. These photo-activation techniques allow for the simultaneous monitoring of fluorescence loss in one compartment and associated fluorescence gain in another. By using photo-activatable techniques, fluorophore populations can be specifically targeted and "turned-on" by a quick pulse of light and their movement can be monitored, which is particularly useful for studying proteins that traffic quickly and have short residences in multiple subcellular compartments[35,38].

While these advanced imaging methods allow for the study of a subpopulation of a protein of interest (the subpopulation illuminated by light) there are three principal limitations. First, each of these experiments is slow and time consuming and must be performed individually on a large number of cells in order to get meaningful ensemble

results about the target protein. Second, these methods require highly specific labeling techniques; consequently, most FLIP/FRAP experiments rely on perfectly specific FP fusions or highly specific protein-directed labeling methods. As previously discussed, fusions to large tags have been shown to disrupt protein localization, trafficking, and function and adverse effects have also been demonstrated in FLIP/FRAP studies[39]. In order to obtain biologically relevant results, careful control experiments must be performed to demonstrate proper behavior of the target protein-FP fusion. Finally, these experiments are limited to studying cellular compartments that can be easily and selectively illuminated by light. For example, it would be impossible using these methods to specifically highlight proteins at the neuronal synapse, the inner mitochondrial membrane, the ER, or the Golgi apparatus. This significant impediment is why most examples of these techniques involve proteins that undergo nucleocytoplasmic shuttling[32,40-45] or receptors that can be selectively labeled and bleached at the cell surface[46-49].

Compartmentalized labeling as a method to probe biological function

In order to specifically highlight spatially defined subpopulations of a protein of interest, we envisioned using genetically targeted ligases. Then we would be able to study the biology of the labeled pool independently of the rest of the protein population. In this chapter, we will summarize work towards this goal. We will first describe a proof-of-principle experiment that will demonstrate that spatially defined subpopulations of a protein of interest can be specifically labeled using targeted ligase. Then we will describe the use of this method to label spatially distinct subpopulations of three biologically relevant proteins: the cytoskeletal protein β -actin, the synaptic adhesion protein neurexin-1β, and the transcription factor MondoA. Before introducing the experimental results, we will provide a brief background on the physiological function of these proteins and the previous investigations of the relationship between their subcellular localizations and biological functions.

Results

Proof of principle: Compartmentalized coumarin labeling

Before investigating biologically relevant proteins, we wanted to demonstrate the concept of compartmentalized labeling in a controlled system. The first step is to demonstrate that coumarin labeling is observed only when coumarin ligase and substrate are co-localized, and not when they are expressed in separate compartments. We proposed using targeted ligases to label subpopulations of LAP2-YFP. The general scheme for this proof-of-principle experiment is shown in Figure 3-1. In this scheme, nuclear labeling should only label the nuclear population of LAP2-YFP, while cytosolic ligase should label only the cytosolic population of LAP2-YFP (Figure 3-1, top).

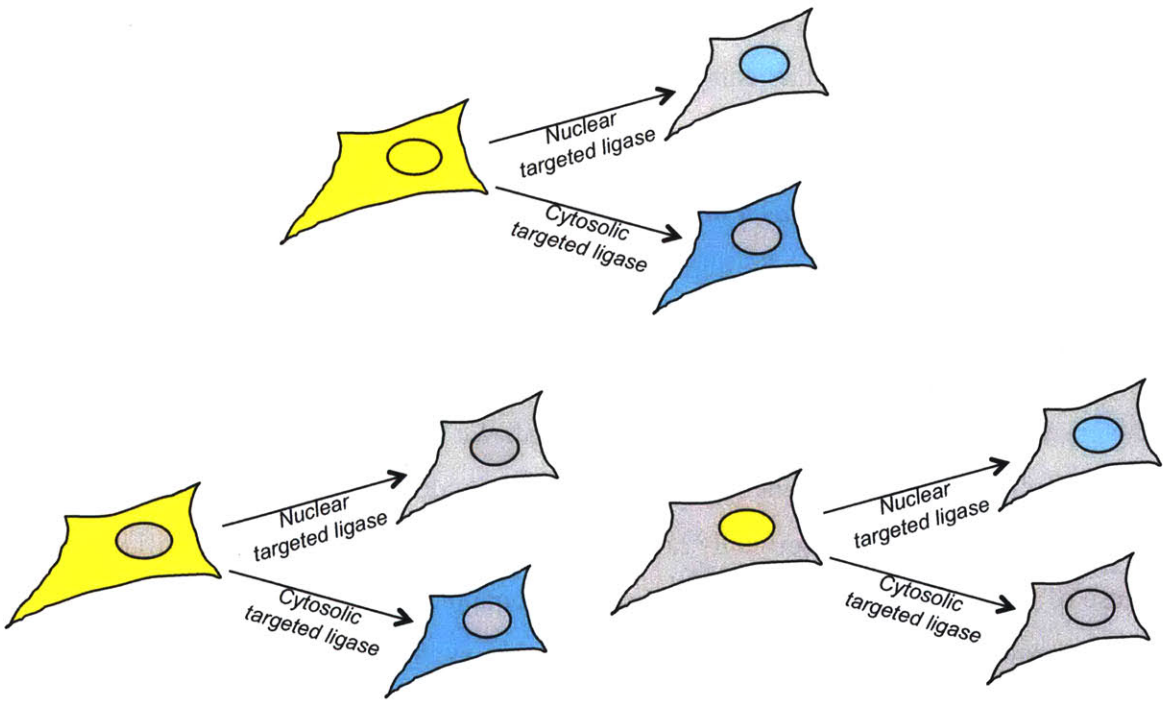


Figure 3-1. Scheme for proof-of-principle compartmentalized labeling experiment. HEK cells that are expressing LAP2-YFP across the whole cell, when labeled with nuclear- or cytosolic-targeted ligase should show only nuclear or cytosolic labeling, respectively (top). As control experiments, when using targeted substrate, we should only observe coumarin labeling when ligase and substrate are co-localized (bottom).

In practice, we transfected HEK cells with both nuclear LAP2-YFP and cytosolic LAP2-YFP. When using cytosolic ligase, we observed labeling of the cytosolic subpopulation (Figure 3-2A, left, top row), and when using nuclear ligase we observed specific labeling of the nuclear subpopulation of LAP2-YFP (Figure 3-2A, right, top

row). Positive controls are shown where individual enzyme and substrate are co-localized and strong labeling signal is observed (Figure 3-2A, middle rows). Negative controls show that when enzyme and substrate are not co-localized, no coumarin labeling was observed (Figure 3-2A, bottom rows). Antibody staining for enzyme in Figure 3-2B shows nuclear and cytosolic ^{W37V}LplA expression.

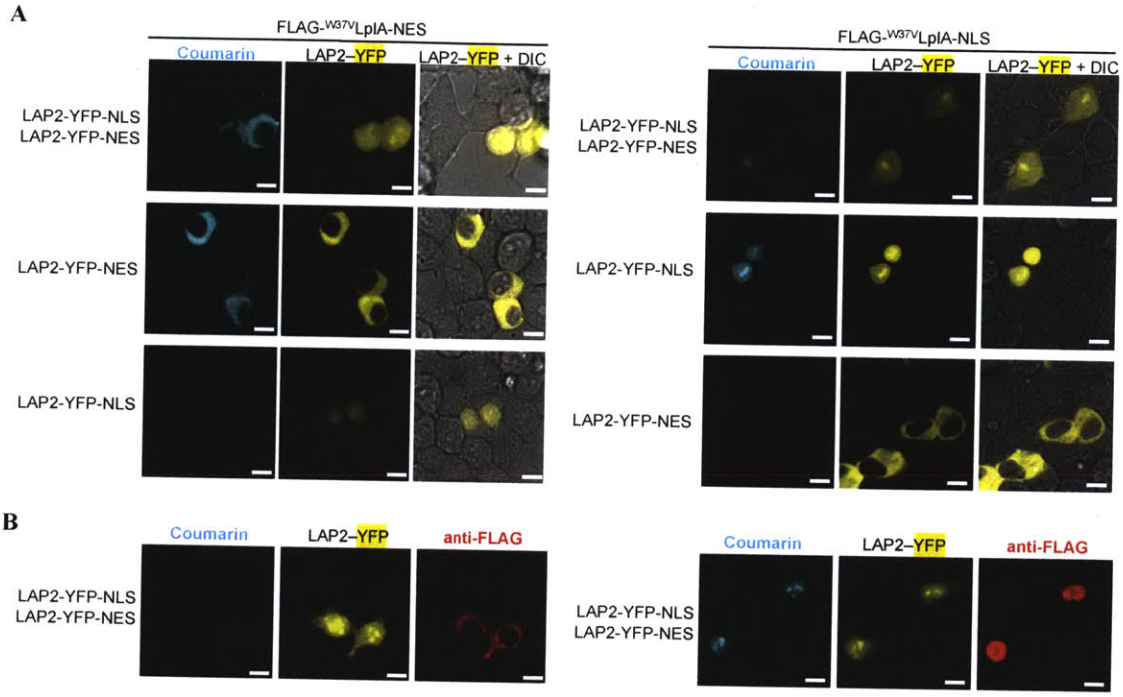


Figure 3-2. Proof-of-principle: compartmentalized PRIME labeling with genetically targeted ligase. **(A)** HEK cells that are expressing both LAP2-YFP-NLS (nuclear) and LAP2-YFP-NES (cytosolic) were labeled with either cytosolic (NES, left) or nuclear-targeted (NLS, right) ^{W37V}LplA. Controls where individual substrates are co-localized with enzyme (middle row) or not co-localized (bottom row) demonstrate the specificity of this method. **(B)** Anti-FLAG staining of fixed cells shows the expression pattern of the targeted ^{W37V}LplA enzymes.

This was an encouraging result, but in this experiment we used spatially defined proteins that are not actively exchanging. It was not clear whether this technique would allow the compartmentalized labeling of real proteins that may be undergoing constant dynamic trafficking and exchange. To probe the abilities of this technique, we moved forward in stages. We first investigated β-actin, whose sub-cellular localization and trafficking has been studied in a variety of ways, including fluorescence imaging techniques[40,50-52]. We then demonstrated compartmentalized labeling of the synaptic adhesion protein neurexin1β, which has not been extensively studied intracellularly but

has a well-defined trafficking pattern to the surface of neurons[53]. Finally, we attempted compartmentalized labeling of the transcription factor MondoA. Although MondoA is known to be present in multiple subcellular compartments under different conditions, its dynamic trafficking behavior is virtually unknown[54,55].

Biology of β-actin

Cytosolic β-actin is a major component of the cytoskeleton and its essential roles in cell migration, cell morphology, and intracellular trafficking have been extensively studied. Since it was proposed over 30 years ago, the existence of a nuclear pool of actin has been a topic of heated debate in the literature[50]. Primarily because of improvements in protein labeling techniques and fluorescence microscopy, the last ten years have produced an overwhelming consensus that β-actin is present in the nucleus[23,24]. The potentially distinct biological roles of nuclear actin are now an active area of research[21,25,56].

Nuclear actin has been identified as a component of chromatin remodeling complexes[24], associates with RNA polymerases[57], and binds to transcription factors[22] and proteins involved in mRNA export[58]. Thus, nuclear actin may have structural or even regulatory roles in the nucleus that are distinct from the traditional cytosolic functions of the protein. Since the roles of the spatially segregated actin populations seem to be very different, many groups have investigated whether the two pools of actin are segregated or actively exchanging. While the 42 kDa actin monomer is small enough to pass through the nuclear pore (generally accepted size-cutoff of ~50 kDa[59,60], although conflicting data exist[61]), it is not clear whether nuclear actin is actively imported or exported. Recent work by Dopie and colleagues has demonstrated using both FLIP and FRAP on GFP-tagged β-actin that the pools of nuclear and cytosolic actin seem to be exchanging quickly[40]. For the FLIP study, the authors performed continuous cytosolic photobleaching of GFP-β-actin, and observed concurrent photobleaching of nuclear actin, indicating that at least a portion of the nuclear pool of actin exchanges with the cytosolic population. In the FRAP study, the authors photobleached the nuclear GFP-β-actin and they observed gradual nuclear fluorescence recovery. This experiment suggests that the cytosolic pool of actin is either passively diffused or actively transported into the nucleus. Using RNAi

knockdowns of nuclear import/export proteins, Dopie and colleagues were able to demonstrate that, at least for the 69kDa GFP-β-actin chimera, active import is required.

While the study described above indicates that the two actin subpopulations are exchanging, there are several issues with the experimental design. First, GFP-β-actin almost doubles the size of the native actin, and likely prohibits passive diffusion through the nuclear pore. So while the GFP-β-actin chimera required active transport into and out of the nucleus, it is not prudent to conclude from this result that endogenous actin monomers also require active transport. Second, previous work has shown that nuclear export of β-actin requires actin binding factors that maintain the protein in its monomeric form[25,62]. This is another indication that the size and polymeric state of actin is critical to its transport and exchange. If even actin dimers cannot be exported from the nucleus, the export of a GFP-actin fusion is probably not the best model for biologically relevant export of monomeric actin. Third, previous studies[40,52,63] (including our own work) have demonstrated that fusions of actin to FPs such as mApple or GFP prohibit or decrease the nuclear accumulation of actin under basal conditions, a phenomenon which is not seen when using antibody staining for endogenous actin or live-cell imaging with small tags [2357](Figure 3-3A). Dopie and colleagues do not address this obvious issue in their paper, even though their GFP-actin control cells clearly show significant nuclear exclusion of their GFP-actin chimera (Figure 3-3B). We thought the PRIME technology would be a good fit for studying nuclear β-actin. The method uses a small tag that minimally perturbs actin monomers and hopefully closer mimics endogenous β-actin. In the subsequent section, we demonstrate how we have used genetically targeted ligases to label spatially defined subpopulations of LAP2-actin.

Demonstration of compartmentalized labeling of nuclear actin

We first made a LAP2-β-actin fusion and performed coumarin PRIME labeling using untargeted ligase. The coumarin labeling signal of LAP2-β-actin has overlapping localization to antibody staining or phalloidin staining, two traditional methods of imaging actin in fixed cells (Figure 3-3A). We therefore concluded that the LAP2-tagged β-actin has similar localization as endogenous actin and would likely be a good mimic of

the behavior of monomeric actin. We next proposed to label actin subpopulations using both nuclear and cytosolic coumarin ligase (^{W37V}LplA).

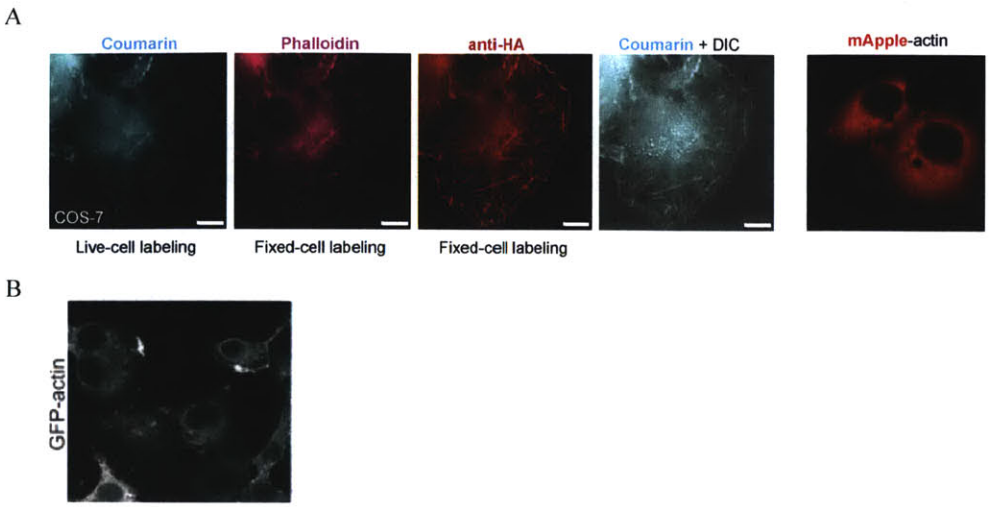


Figure 3-3. Actin localization pattern depends on detection method. (A) Comparison of methods for actin visualization. Coumarin PRIME labeling of HA-LAP2-β-actin was performed in live cells. Cells were then fixed and actin detected via two traditional methods: phalloidin staining for filamentous actin and antibody staining. These three methods all show a nuclear pool of β-actin. These methods are compared to the effect of a large FP tag (mApple), where significant nuclear exclusion is observed **(B)** Localization pattern of the GFP- β-actin fusion as reported in Dopie *et. al.* (reprinted here with permission). Note that the localization pattern of this FP chimera is also nuclear excluded.

Figure 3-4 shows coumarin labeling of LAP2-β-actin in COS-7 cells when using untargeted, cytosolic, and nuclear ligases. We initially did not see any labeling signal using the nuclear-targeted coumarin ligase. We hypothesized that this was due to nuclear actin being bound in complexes or not sterically accessible to the nuclear enzyme. This result also indicates that the nuclear actin we observe when using untargeted ligase is probably the result of coumarin labeling of actin monomers in the cytosol, followed by translocation to the nucleus. The observation that labeling with cytosolic coumarin ligase produces nuclear labeling signal supports this idea that nuclear translocation follows cytosolic labeling of accessible LAP2-β-actin monomers (Figure 3-4, middle panels).

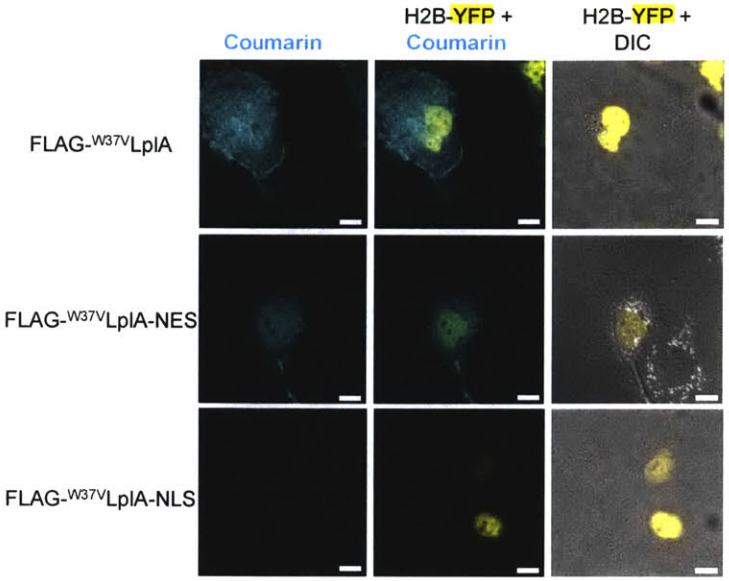


Figure 3-4. Compartmentalized labeling of LAP2-β-actin. COS-7 cells expressing LAP2-β-actin are labeled with untargeted, cytosolic (NES), or nuclear-targeted (NLS) coumarin ligase.

To test this hypothesis, we performed coumarin labeling using nuclear targeted ligase in the presence of the drug leptomycin-B (LMB), a selective inhibitor of Crm1 nuclear exporter, which has been previously demonstrated to cause nuclear accumulation of β-actin[23,62]. Under these conditions, we observed specific labeling of the nuclear pool of LAP2-β-actin using nuclear-targeted coumarin ligase (Figure 3-5). When labeling in the presence of leptomycin-B, we saw nuclear actin structures that included filaments and nucleolar clusters. We suspect that leptomycin-B may cause an increase in nuclear actin such that endogenous actin-binding proteins are saturated, producing free actin that is accessible to the nuclear coumarin ligase, which is not the case under basal conditions.

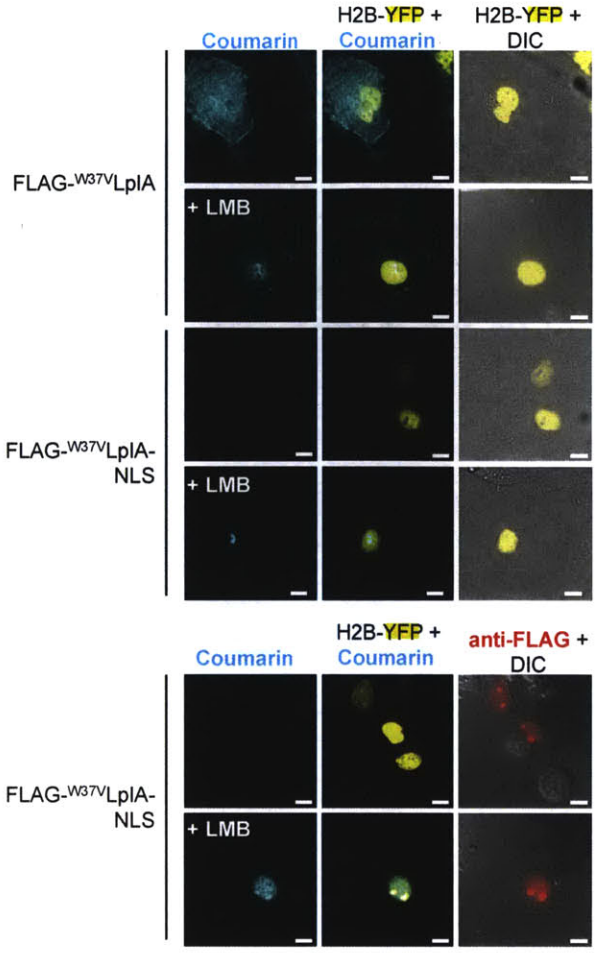


Figure 3-5. Compartmentalized labeling of nuclear actin in the presence of leptomycin B. COS-7 cells expressing LAP2-β-actin were labeled with coumarin using untargeted or nuclear-targeted ^{W37V}LplA. Where indicated, labeling was performed after treatment with 2 ng/ml leptomycin B (LMB) to inhibit nuclear export. LMB is required to see coumarin labeling when using nuclear-targeted ligase.

The behavior of LAP2-β-actin in the presence of LMB closely matches previous results using either antibody staining of endogenous actin post-fixation or live-cell imaging with microinjected fluorescently labeled actin, which all indicate nuclear accumulation of the actin[23,62]. However, these results stand in stark contrast to the observations from Dopie and colleagues, who found no difference in the level of nuclear GFP-actin in the presence or absence of LMB. While the authors claim that this indicates nuclear export of actin is not Crm1 dependent and is instead associated with another exporter (Exp6), it could simply be that GFP-β-actin does not behave, traffic, or interact with nuclear actin binding proteins in the same way as LAP2-β-actin or endogenous actin. The only other instance of reported β-actin insensitivity to LMB was seen with YFP-fused actin[52], which supports the theory that this insensitivity may be an artifact of the large FP fusions.

Biology of neurexin-1β

Neurexins are pre-synaptic transmembrane adhesion proteins that have been linked to synaptogenesis and synapse maturation. The broad class of neurexin proteins is extremely heterogeneous due to the large number of alternative splice sites, with the theoretical number of isoforms approaching that of the immunoglobulin family[14,64]. Neurexins were discovered in the early 1990s and quickly rose to the forefront of neurobiology after initial experiments indicated these proteins trigger post-synaptic differentiation and thus may be involved in synapse modulation[53]. Soon after neuroligin was identified as the post-synaptic binding partner of neurexin, these two proteins became the subject of intense study by the neuroscience community. Subsequent experiments indicated that neurexin-neuroligin trans-synaptic interactions appear to control the balance of GABAergic and glutamatergic inputs and thus may have huge implications in the understanding of not only the formation of the synapse but also signal transduction, synapse classification, and even clinical manifestations of neurological diseases such as autism[14,65,66].

The process of synapse formation is divided into several distinct steps. First, axons and dendrites must grow towards each other and form initial contacts. Second, a synaptic junction is established with subsequent (or perhaps concomitant) recruitment of pre-and post-synaptic scaffolding proteins and cellular machinery. Finally, the synapse matures and, through modification of protein-protein interactions at the synapse and within the pre and post-synaptic busons, synapse specification is achieved. Through various in vitro and in vivo experiments, neurexin and neuroligin have been implicated in every step of this process[14,67-69]. However, investigation into the biological relevance of these results has so far been limited by the live-cell techniques available. While all areas of synapse formation are being actively researched, groups studying synapse maturation are particularly interested in exploring the membrane implantation and subsequent turnover of neurexin and neuroligin[53,70-72]. The ability to specifically highlight the intracellular pool of neurexin before it is trafficked to the cell surface and then follow protein recruitment to maturing synapses would be extremely beneficial. We describe in the subsequent section the application of coumarin PRIME to the compartmentalized

labeling of LAP2-tagged neurexin-1β. While labeling in neurons is the end goal of neurexin imaging, here we demonstrate labeling of neurexin-1β in fibroblasts.

Demonstration of compartmentalized labeling of neurexin-1β

For these experiments, we cloned a double LAP2 fusion to neurexin-1β (LAP2-neurexin1β-LAP2). This construct geometry would allow us to perform labeling on both sides of the membrane. We performed three experiments to label distinct subpopulations of neurexin-1β as it is trafficked through the secretory pathway and implanted in the membrane. First, using exogenously applied enzyme, we labeled only the LAP2-neurexin1β-LAP2 that is accessible on the cell surface (Figure 3-6, left). This produced clearly defined rings with no labeling of intracellular LAP2-neurexin1β-LAP2. Second, using ligase targeted to the inner leaflet of the plasma membrane, we labeled the intracellular face of the implanted neurexin-1β (Figure 3-6, middle). This produced a clear ring at the plasma membrane, and also a low level of signal in the cell interior, corresponding to sites where the enzyme and LAP2-neurexin1β-LAP2 came into contact before membrane implantation. It is likely that shorter labeling times would eliminate this aberrant signal. Finally, we used untargeted ligase to label the entire intracellular pool of LAP2. This produced not only a clear ring at the plasma membrane from labeling of the intracellular face of the implanted neurexin1β but also strong signal corresponding to labeling of accessible LAP2 at the cytosolic face of the endoplasmic reticulum and Golgi apparatus as the protein is trafficked to the cell surface (Figure 3-6, right).

We also considered labeling of LAP2-tagged neurexin as it is trafficked through the endoplasmic reticulum using an ER-targeted ligase. However, as we discussed in Chapter 2, coumarin labeling using ER-targeted ligases is not possible. Chapter 7 summarizes the evolution of a cell-surface and ER-active probe ligase and includes a demonstration of ER labeling of neuroligin, neurexin's interacting partner.

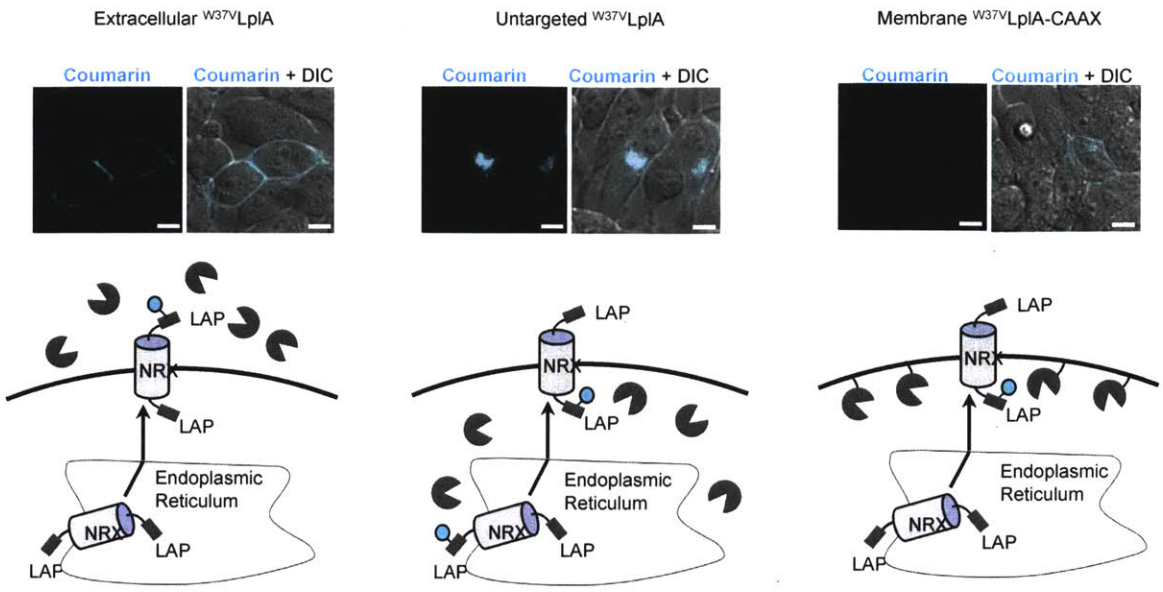


Figure 3-6. Compartmentalized labeling of neurexin-1β. The double-tagged neurexin-1β construct, LAP2-neurexin1β-LAP2 (LAP2-NRX-LAP2 in the diagram) was transfected into HEK cells. Coumarin labeling was performed using ^{W37V}LplA applied exogenously (left), ^{W37V}LplA targeted to the inner-leaflet of the plasma membrane (right), and untargeted ^{W37V}LplA (middle).

Biology of MondoA

Glucose is a fundamental metabolite and the biochemical pathways of glucose metabolism are among the best understood in biochemistry. For decades scientists have known that altered glucose homeostasis is a characteristic of human diseases such as cancer, diabetes, and Alzheimer's disease[73,74]. However, most of the molecular mechanisms for how cells sense and respond to changes in extracellular glucose concentrations are not well understood. Several proposed pathways for glucose sensing involve cytosolic proteins that sense intracellular glucose levels and then respond through regulation of glycolytic enzymes or by initiating insulin secretion[75-77]. About ten years ago, a new family of transcriptional regulators of glucose homeostasis was discovered[9]. These protein complexes sense glucose metabolism and then trigger a transcriptional response that induces expression of glycolytic target genes[78]. The most intriguing of these glucose sensors is the MondoA:Mlx dimeric transcription factor.

MondoA:Mlx is a functional heterodimer that is localized to the outer mitochondrial membrane when intracellular glucose concentrations are low. When intracellular levels of glucose increase, MondoA:Mlx translocates to the nucleus and

induces transcription of several key proteins involved in glucose regulation, including thioredoxin-interacting protein (TXNIP), which is a potent down regulator of glucose uptake. Previous work demonstrated that MondoA:Mlx is associated with hexokinase on the outer mitochondrial membrane and that the translocation is dependent on hexokinase activity (hexokinase phosphorylates glucose to glucose-6-phosphate). This result, coupled with the fact that MondoA does not sense glucose directly but instead responds to levels of glucose-6-phosphate (G6P), indicates that MondoA:Mlx translocation may respond to a conformational change upon catalysis of glucose by hexokinase[54].

Recent work on MondoA:Mlx translocation has revealed intriguing results. While it has been clearly established that high intracellular glucose concentrations result in translocation of MondoA and TXNIP transcription, little is known about how the system resets itself. In many cases, TXNIP transcription remains high even after glucose levels have returned to normal (personal communication, Vamsi Mootha). This is what we call "metabolic memory"; the cells remember that they were in a high glucose environment and need to be reset in order to return to the basal state. An influx of insulin will erase the metabolic memory of the cell and finally reduce TXNIP transcription after it has begun. The mechanism of the erasure is not known, but seems to be dependent on insulin as TXNIP levels remain high until an influx of insulin drops expression to basal levels (Vamsi Mootha, personal communication).

The mechanism by which insulin erases the metabolic memory of the cell is unknown. One hypothesis is that insulin directly affects MondoA:Mlx localization; by disrupting the MondoA:Mlx/DNA interaction, for example, insulin would result in the loss of TXNIP transcription and the return to the cell's basal state. Another theory is that perhaps insulin serves to activate export of MondoA:Mlx from the nucleus to the outer mitochondrial membrane where the heterodimer is able to re-associate with hexokinase; this would also "erase" metabolic memory as the translocation would result in the loss of TXNIP transcription. Our PRIME technology is uniquely suited to investigating these two hypothesis. By specifically labeling the nuclear pool of MondoA:Mlx, we would then be able to monitor the localization of this population and follow any trafficking events after the addition of insulin to erase metabolic memory. Although we were not able to

directly address these issues using our coumarin PRIME technology, what follows is a description of the considerable progress towards this goal.

Progress towards compartmentalized labeling of MondoA

MondoA-LAP2 and MondoA-E2p fusions exhibit proper translocation

The first goal of this project was to simply produce a fusion to MondoA that does not disrupt the protein's translocation behavior. Although MondoA is over 1000 amino acids long, it is exquisitely susceptible to disruption via genetic fusions. This sensitivity to modification is unsurprising considering that proper function of MondoA requires at least two known protein-protein interactions (with both Mlx and hexokinase) along with unimpeded DNA binding by the basic helix-loop-helix leucine zipper domain of MondoA. While small epitope tags (V5 and HA) had been fused to the C-terminus of MondoA without interfering with MondoA:Mlx translocation, FPs at the N-terminus, C-terminus or even internal fusion sites did not produce proper MondoA shuttling behavior (Donald Ayer, personal communication). These FP fusions all resulted in improper nuclear localization under low-glucose conditions, indicating that they likely disrupt binding to the mitochondrial membrane or perhaps hexokinase.

We made fusions of E2p and LAP2 to the both the N- and C-termini of MondoA. We then tested the behavior of these fusions in a standard MondoA glucose shuttling assay (for details, see Experimental). This assay reproduces both a glucose-starved (Low Glucose) and a high G6P (High Glucose) state. Under the glucose-starved (Low Glucose) conditions, there is very little G6P inside the cells, so the MondoA-Mlx heterodimer is localized to the mitochondria (Figure 3-7, left). Under high G6P (High Glucose) conditions, the MondoA:Mlx heterodimer is localized to the nucleus (Figure 3-7, right).

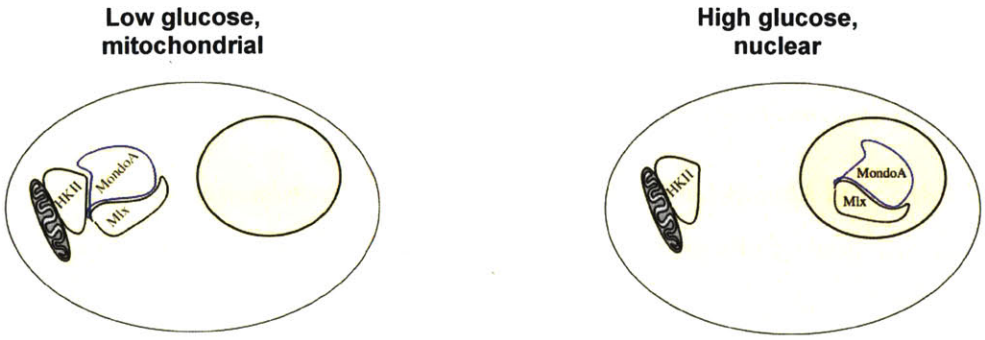


Figure 3-7. Scheme of the subcellular localization patterns for MondoA:Mlx under both high and low intracellular glucose concentrations.

We performed the glucose shuttling assay in live HeLa cells expressing the MondoA fusions and then fixed the cells and compared the subcellular localization pattern to that of MondoA-V5. The MondoA-V5 construct from the Ayer lab traffics properly in HeLa cells, as demonstrated in the top panel of Figure 3-8. The N-terminal fusions of E2p and LAP2 to MondoA localize primarily to the nucleus regardless of intracellular glucose levels. This indicates that even very small tags can disrupt the function of MondoA when fused at the N-terminus of the protein. The C-terminal fusions of E2p and LAP2 to MondoA both traffic normally in the shuttling assay (Figure 3-8). This was the first demonstration of the proper trafficking of a MondoA fusion to tags that can be used for live-cell protein labeling.

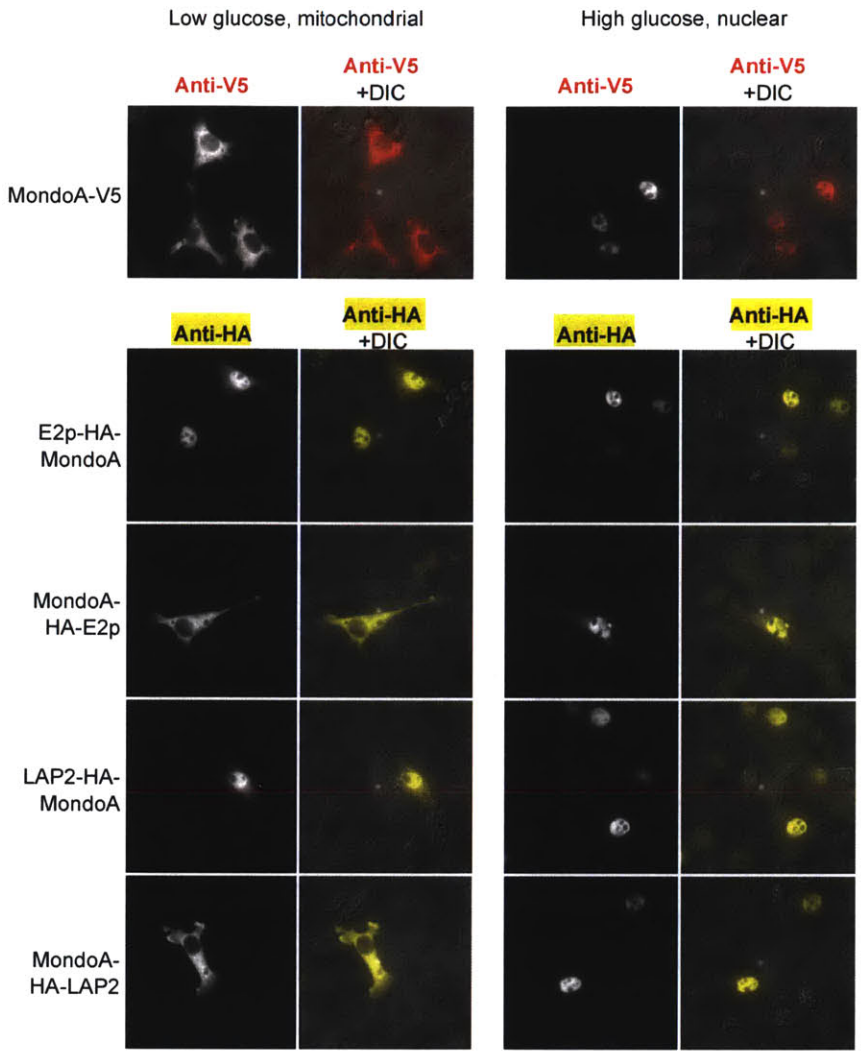


Figure 3-8. C-terminal LAP2 and E2p fusions to MondoA traffic as wild-type MondoA. The glucose shuttling assay (see experimental) was performed in HeLa cells expressing the indicated MondoA fusions. Fusion expression patterns were determined by antibody staining after cell-fixation. While N-terminal fusions were mis-localized under low glucose conditions, the C-terminal fusions of LAP2- and E2p- to MondoA traffic properly.

Expression level must be controlled for proper shuttling of MondoA-E2p and MondoA-LAP2

We quickly determined that MondoA localization is highly sensitive to not only to fusions, but also overexpression. In cells expressing MondoA-V5 construct, overexpression induced whole-cell localization of the protein when treated with high glucose, although we always observed mitochondrial MondoA under low glucose conditions regardless of expression level (data not shown). Similar mislocalization effects

were seen with E2p and LAP2 fusions when expressed at high concentrations (Figure 3-9). These data suggest that perhaps the nuclear import mechanism or unknown MondoA nuclear binding partners become saturated when MondoA is overexpressed, resulting in cytosolic background accumulation of the protein under high glucose conditions.

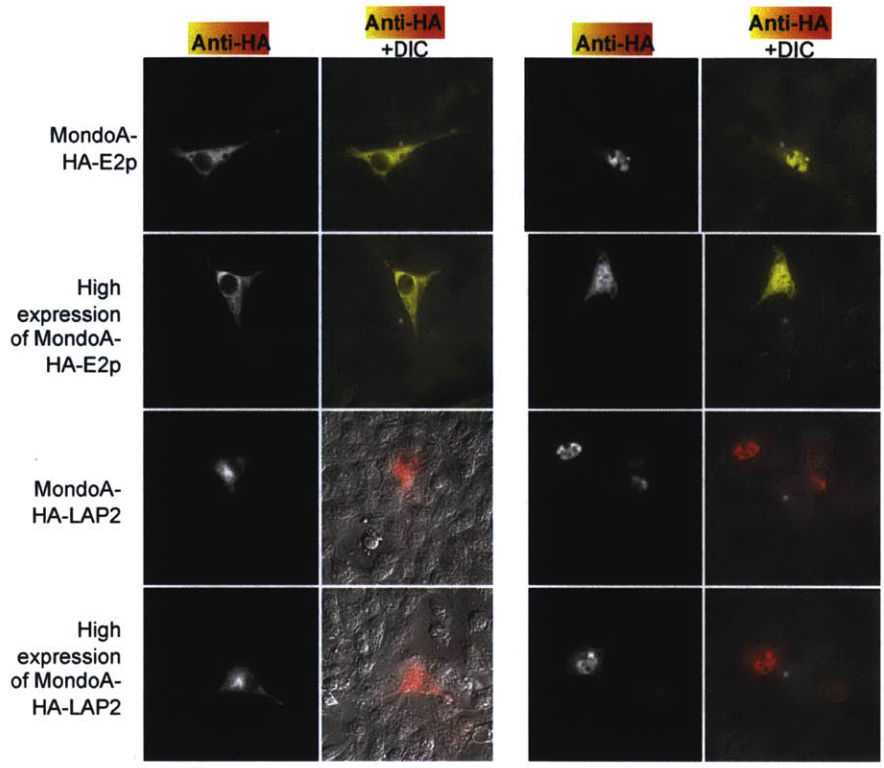


Figure 3-9. MondoA-LAP2 and MondoA-E2p nuclear localization is disrupted at high expression levels. HeLa cells were prepared with the indicated MondoA fusion at either high or low expression conditions (normal expression is identical to levels in Figure 3-8, high expression is twice that level).

After much optimization of MondoA expression conditions, attempts at co-transfecting additional hexokinase or Mlx, and even adding nuclear export inhibitors under high glucose conditions, we determined that MondoA fusions simply must be expressed at concentrations less than 15 μM in order to localize properly in the shuttling assay.

Coumarin labeling is not sensitive enough to achieve coumarin labeling of MondoA-E2p and MondoA-LAP2 at low expression levels

Since MondoA expression levels must be limited to less than 15 μM in order to exhibit proper localization, we were concerned that our coumarin PRIME technology (which has a sensitivity limit of about 10 μM) would not be applicable to labeling the MondoA in live cells. Unfortunately, even with untargeted coumarin ligase, we were not able to see any labeling of the MondoA at the mitochondria (Low Glucose) or in the nucleus (High Glucose) (Figure 3-10A). However, we did observe labeling of MondoA when it was over-expressed and improperly localized (data not shown).

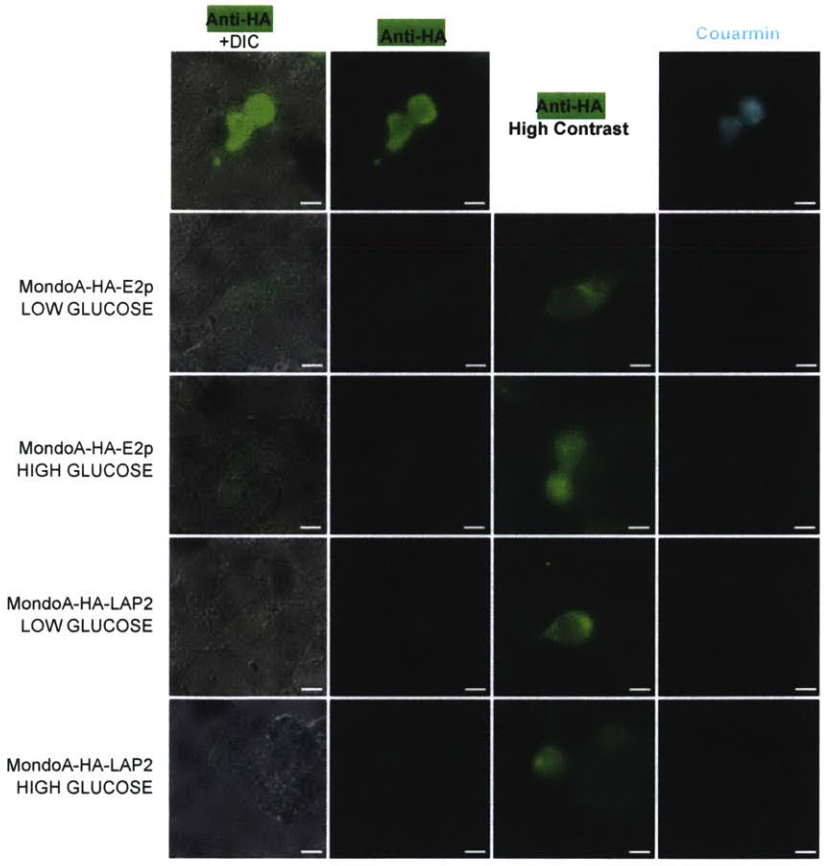


Figure 3-10. No coumarin labeling of MondoA fusions when properly localized. HEK cells expressing the indicated MondoA fusion and ^{W37V}LplA were subjected to the Mondo shuttling assay. Coumarin labeling was then performed under both high and low glucose conditions. No coumarin labeling was observed, although we did see coumarin labeling of the positive control plasmid (top panels).

At low expression, the MondoA is at the cusp of what we could detect using the PRIME methodology. Keeping that in mind, we were still concerned that perhaps the phenomenon we were seeing was not the result of low labeling sensitivity but actually the result of the LAP2 or E2p being sterically inaccessible when the MondoA fusions were properly localized. To test this theory, we decided to try intracellular lipoylation using targeted ^{WT}LplA. Since the wild-type lipoylation reaction is more sensitive than the coumarin ligation, we hoped that we could observe labeling of MondoA-LAP2 or MondoA-E2p under the low expression conditions required for proper localization.

Demonstration of compartmentalized lipoylation of MondoA-E2p and MondoA-LAP2

For lipoylation, we needed to perform tetracycline-controlled expression ^{WT}LplA, as described in Chapter 2. We first confirmed that all of our shuttling assays were functioning properly in the HeLa cell line stably expressing the tetracycline repressor (data not shown). We then performed lipoylation of MondoA-LAP2 and MondoA-E2p under low glucose conditions using untargeted ligase and under high glucose conditions using nuclear-targeted LplA. We detected lipoylation post-fixation and found that labeling of the properly-localized proteins using targeted ^{WT}LplA was successful(Figure 3-11).

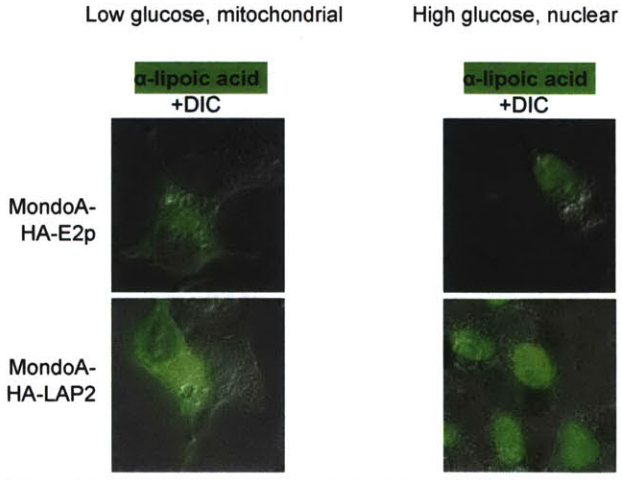


Figure 3-11. MondoA fusions can be lipoylated when properly localized. HeLa cells expressing the indicated MondoA fusion and ^{WT}LplA were subjected to the Mondo shuttling assay. The cells were then treated with lipoic acid, and lipoylation was detected post-fixation.

This allowed us to conclude that while the coumarin PRIME method was not sensitive enough to probe the biology of the MondoA subcellular shuttling, the MondoA

system is amenable to an enzyme mediated labeling method at both the mitochondria and the nucleus. This biological system can be investigated further after the development of a more sensitive probe ligase.

Conclusion

In this chapter, we have introduced a new method that uses genetically targeted coumarin ligase to achieve labeling of a subpopulation of a protein of interest. We first demonstrated this technique in a proof-of-principle experiment wherein we were able to selectively label either the nuclear or cytosolic population of a LAP2-YFP fusion. We then demonstrated the use of this method on three biologically relevant proteins: neurexin-1β, β-actin, and MondoA. All of these proteins have been demonstrated to have distinct functions depending on subcellular localization or dynamic shuttling events.

Using our compartmentalized labeling method, we were able to label nuclear β-actin and demonstrate that most of the nuclear β-actin labeling signal we see using untargeted ligase is the result of actin monomers being labeled in the cytosol and then translocated to the nucleus. We were also able to demonstrate that nuclear levels of actin rise in the presence of leptomycin B, which agrees with previous literature studying endogenous actin, but conflicts with recent results with large actin-FP fusions. Next, we demonstrated labeling of three subpopulations of neurexin-1β as it was trafficked through the secretory pathway. Finally, we demonstrated steps toward labeling of MondoA:Mlx heterodimer. We were able to demonstrate the first instance of proper MondoA trafficking when fused to a useful live-cell protein labeling tag (both E2p and LAP2). While coumarin labeling was not sensitive enough to label spatially distinct subpopulations, we were able to achieve compartmentalized lipoylation of MondoA-E2p and MondoA-LAP2 fusions.

Experimental

Fluorescence imaging.

Cells were imaged in Dulbecco's Phosphate Buffered Saline (DPBS) in epifluorescence or confocal modes. For epifluorescence imaging, we used a Zeiss AxioObserver inverted microscope with a 40x or 63x oil-immersion objective. Coumarin (400/20 excitation, 425 dichroic, 435/30 emission), YFP (493/16 excitation, 506 dichroic, 525/30 emission), mCherry (570/20 excitation, 585 dichroic, 605/30 emission) and differential interference contrast (DIC) images were collected and analyzed using Slidebook software (Intelligent Imaging Innovations). For confocal imaging, we used a Zeiss Axiovert 200M inverted microscope with a 40x oil-immersion objective. The microscope was equipped with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), and 561 nm (DPSS) lasers (all 50 mW). Coumarin (405 laser excitation, 445/40 emission), YFP (491 laser excitation, 528/38 emission), mCherry (561 laser excitation, 617/73 emission), and DIC images were collected using Slidebook software. Fluorescence images in each experiment were normalized to the same intensity ranges. Acquisition times ranged from 10–1000 milliseconds.

General protocol for coumarin PRIME labeling in living cells. (Figure 3-3)

HEK cells were transfected at ~70% confluency with expression plasmids for ^{W37V}LplA (20 ng for a 0.95 cm^2 dish) and the E2p, LAP4.x or LAP2 fusion protein of interest (400-600 ng) using Lipofectamine 2000 (Invitrogen). 8–24 hours after transfection, cells were incubated with 20 μM coumarin-AM_2 in serum-free DMEM (Dulbecco's modified Eagle medium, Cellgro) for 10 minutes at 37 °C. Note: 0.1% w/v Pluronic F-127 (Invitrogen) can be optionally added to the labeling solution to give more even coumarin distribution in cells. The media was then replaced 3–4 times over 30–60 min at 37 °C, with DMEM supplemented with 10% FBS (Fetal Bovine Serum), to wash out excess coumarin.

Intracellular coumarin labeling (Where protocol differs from that described above)

For Figure 3-2, to obtain equal expression of LAP2-YFP-NES and LAP2-YFP-NLS, a 2:1 (NES:NLS) ratio of plasmid DNA was used. Labeling was performed for 30 minutes when using ^{W37V}LplA-NLS (or ^{W37I}LplA-NLS). After labeling and live-cell imaging, samples were fixed using 3.7% formaldehyde in DPBS followed by cold precipitation with methanol. Subcellular localization of FLAG-tagged LplA was confirmed using immunofluorescence staining with 4 μg/ml anti-FLAG M2 antibody (Stratagene) in 1% BSA in DPBS for 30 minutes. After 15-minute washes, cells were incubated with 4 μg/ml goat anti-mouse antibody conjugated to Alexa Fluor 568 (Invitrogen) in 1% BSA in DPBS for 30 minutes, then washed and imaged.

For Figure 3-6, cell-surface labeling was performed for 30 minutes at room temperature with 10 μM ^{W37V}LplA or ^{W37I}LplA, 50 μM coumarin **4**, 1 mM ATP, and 2 mM $Mg(OAc)_2$ in 50 mM Na_2HPO_4 pH 7.2.

For Figure 3-4 and 3-5, COS-7 cells were labeled for 30 minutes. Where noted, 2 ng/ml leptomycin B was added to the labeling and washout media. Fixed samples were prepared as described above.

MondoA Shuttling Assay (Figure 3-8, 3-9, 3-10, 3-11)

HEK cells were transfected at ~70% confluency with 0.3 μg MondoA-V5 or the indicated LAP2 or E2p fusion to MondoA, 0.3 μg pME18 Flag-Mlx, and 0.2 μg Hexokinase II. Transfected duplicates for both high and low glucose states. After 24 hours, media was exchanged for glucose-free MEM (Cellgro). After 12 hours of incubation in glucose-free media, a "high glucose" state was induced where required by adding 20 mM 2-deoxyglucose to the glucose-free MEM and incubating with cells for an additional 6 hours. Note: Glucose-free media is *required* to observe strong nuclear shuttling of wt Mondo. Cells were then fixed and the MondoA was detected via V5 antibody while the MondoA fusions were detected via HA staining.

MondoA coumarin labeling (Figure 3-10)

MondoA glucose shuttling assay is performed as described above, except ^{W37V}LplA is co-transfected with the plasmids listed. Coumarin labeling was performed in the indicated glucose-free or 2-deoxy-glucose supplemented media by adding 20 μM coumarin-AM2 and incubating for 30 minutes followed by washing steps.

Cells were imaged live for coumarin. Where indicated in legends, cells were fixed (see protocol above) and then stained for MondoA-HA-E2p or MondoA-HA-LAP2 expression before imaging.

MondoA Lipoylation (Figure 3-11)

MondoA glucose shuttling assay is performed as described above but in TREX-HeLa cell lines and with pcDNA4/TO-^{WT}LplA untargeted or targeted to the nucleus. Lipoylation was performed in the indicated glucose-free or 2-deoxy-glucose supplemented media by adding 250 μM lipoic acid and incubating for 4 hours. The cells were then rinsed and fixed using 3.7% Paraformaldehyde in DPBS, followed by cold methanol permeabilization. Cells were blocked for 1 hour in 3% BSA in DPBS at room temperature. Lipoylation was detected by incubating with 1:300 dilution of anti-lipoic acid antibody (CalBiochem) in 1% BSA in DPBS for 20 minutes at room temperature. After two washes, a 1:300 dilution of anti-rabbit AlexaFluor488 antibody (Invitrogen) in 1% BSA in DPBS was added to cells for 20 minutes at room temperature. Cells were washed twice and then imaged.

Wedge-method approximation of MondoA expression levels

Mondo-HA-LAP2 was expressed in HEK cells at varying levels. Antibody staining for HA was performed post-fixation as described above with AlexaFluor488 secondary antibody used. In a side-by-side assay, HA-LAP2-mCherry was expressed and imaged live and then fixed and antibody staining for HA was performed with AlexaFluor488 secondary antibody used. Using the wedge method, we approximated the concentrations of HA-LAP2-mCherry and then correlated expression level to HA staining. Using this approximate ratio of HA staining to concentration, we calculated the approximate concentration of MondoA-HA-LAP2 at various transfection amounts based on HA staining levels.

References

1. Chu, C. T.; Plowey, E. D.; Wang, Y.; Patel, V.; Jordan-Sciutto, K. L. Location, location, location: altered transcription factor trafficking in neurodegeneration. *J Neuropathol Exp Neurol* **2007**, *66*, 873-83.

2. Lemberg, M. K. Intramembrane proteolysis in regulated protein trafficking. *Traffic*, *12*, 1109-18.

3. Szul, T.; Sztul, E. COPII and COPI traffic at the ER-Golgi interface. *Physiology (Bethesda)*, *26*, 348-64.

4. Yoshida, K. Nuclear trafficking of pro-apoptotic kinases in response to DNA damage. *Trends Mol Med* **2008**, *14*, 305-13.

5. Crabtree, G. R.; Graef, I. A. Bursting into the nucleus. *Sci Signal* **2008**, *1*, pe54.

6. Shahbabian, K.; Chartrand, P. Control of cytoplasmic mRNA localization. *Cell Mol Life Sci*, *69*, 535-52.

7. Lindenboim, L.; Borner, C.; Stein, R. Nuclear proteins acting on mitochondria. *Biochim Biophys Acta*, *1813*, 584-96.

8. Billin, A. N.; Eilers, A. L.; Queva, C.; Ayer, D. E. Mlx, a novel Max-like BHLHZip protein that interacts with the Max network of transcription factors. *J Biol Chem* **1999**, *274*, 36344-50.

9. Billin, A. N.; Eilers, A. L.; Coulter, K. L.; Logan, J. S.; Ayer, D. E. MondoA, a novel basic helix-loop-helix-leucine zipper transcriptional activator that constitutes a positive branch of a max-like network. *Mol Cell Biol* **2000**, *20*, 8845-54.

10. Chan, Y. M.; Jan, Y. N. Roles for proteolysis and trafficking in notch maturation and signal transduction. *Cell* **1998**, *94*, 423-6.

11. Gordon, W. R.; Arnett, K. L.; Blacklow, S. C. The molecular logic of Notch signaling--a structural and biochemical perspective. *J Cell Sci* **2008**, *121*, 3109-19.

12. Hill, C. S. Nucleocytoplasmic shuttling of Smad proteins. *Cell Res* **2009**, *19*, 36-46.

13. Ebinu, J. O.; Yankner, B. A. A RIP tide in neuronal signal transduction. *Neuron* **2002**, *34*, 499-502.

14. Craig, A. M.; Kang, Y. Neurexin-neuroligin signaling in synapse development. *Curr Opin Neurobiol* **2007**, *17*, 43-52.

15. Giese, B., *et al.* Long term association of the cytokine receptor gp130 and the Janus kinase Jak1 revealed by FRAP analysis. *J Biol Chem* **2003**, *278*, 39205-13.

16. Turk, V., *et al.* Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*, *1824*, 68-88.

17. Reiser, J.; Adair, B.; Reinheckel, T. Specialized roles for cysteine cathepsins in health and disease. *J Clin Invest*, *120*, 3421-31.

18. Gould, G. W. IKKepsilon: a kinase at the intersection of signaling and membrane traffic. *Sci Signal*, *4*, pe30.

19. Wortzel, I.; Seger, R. The ERK Cascade: Distinct Functions within Various Subcellular Organelles. *Genes Cancer*, *2*, 195-209.

20. Zehorai, E.; Yao, Z.; Plotnikov, A.; Seger, R. The subcellular localization of MEK and ERK--a novel nuclear translocation signal (NTS) paves a way to the nucleus. *Mol Cell Endocrinol*, *314*, 213-20.

21. Visa, N.; Percipalle, P. Nuclear functions of actin. *Cold Spring Harb Perspect Biol*, *2*, a000620.

22. Miralles, F.; Visa, N. Actin in transcription and transcription regulation. *Curr Opin Cell Biol* **2006**, *18*, 261-6.

23. Kwak, I. H.; Kim, H. S.; Choi, O. R.; Ryu, M. S.; Lim, I. K. Nuclear accumulation of globular actin as a cellular senescence marker. *Cancer Res* **2004**, *64*, 572-80.

24. Olave, I. A.; Reck-Peterson, S. L.; Crabtree, G. R. Nuclear actin and actin-related proteins in chromatin remodeling. *Annu Rev Biochem* **2002**, *71*, 755-81.

25. Vartiainen, M. K. Nuclear actin dynamics--from form to function. *FEBS Lett* **2008**, *582*, 2033-40.

26. Kuo, T. F.; Tatsukawa, H.; Kojima, S. New insights into the functions and localization of nuclear transglutaminase 2. *FEBS J*, *278*, 4756-67.

27. Belkin, A. M. Extracellular TG2: emerging functions and regulation. *FEBS J*, *278*, 4704-16.

28. Shimizu, T., *et al.* Structure of a covalently cross-linked form of core histones present in the starfish sperm. *Biochemistry* **1997**, *36*, 12071-9.

29. Mishra, S.; Saleh, A.; Espino, P. S.; Davie, J. R.; Murphy, L. J. Phosphorylation of histones by tissue transglutaminase. *J Biol Chem* **2006**, *281*, 5532-8.

30. Boehm, J. E.; Singh, U.; Combs, C.; Antonyak, M. A.; Cerione, R. A. Tissue transglutaminase protects against apoptosis by modifying the tumor suppressor protein p110 Rb. *J Biol Chem* **2002**, *277*, 20127-30.

31. Ishikawa-Ankerhold, H. C.; Ankerhold, R.; Drummen, G. P. Advanced Fluorescence Microscopy Techniques-FRAP, FLIP, FLAP, FRET and FLIM. *Molecules*, *17*, 4047-132.

32. Koster, M.; Frahm, T.; Hauser, H. Nucleocytoplasmic shuttling revealed by FRAP and FLIP technologies. *Curr Opin Biotechnol* **2005**, *16*, 28-34.

33. Chudakov, D. M., *et al.* Photoswitchable cyan fluorescent protein for protein tracking. *Nat Biotechnol* **2004**, *22*, 1435-9.

34. Lummer, M., *et al.* Reversible photoswitchable DRONPA-s monitors nucleocytoplasmic transport of an RNA-binding protein in transgenic plants. *Traffic*, *12*, 693-702.

35. Sturgill, J. F.; Steiner, P.; Czervionke, B. L.; Sabatini, B. L. Distinct domains within PSD-95 mediate synaptic incorporation, stabilization, and activity-dependent trafficking. *J Neurosci* **2009**, *29*, 12845-54.

36. Diaspro, A., *et al.* 3D localized photoactivation of pa-GFP in living cells using two-photon interactions. *Conf Proc IEEE Eng Med Biol Soc* **2006**, *1*, 389-91.

37. Lippincott-Schwartz, J.; Patterson, G. H. Fluorescent proteins for photoactivation experiments. *Methods Cell Biol* **2008**, *85*, 45-61.

38. Lippincott-Schwartz, J.; Altan-Bonnet, N.; Patterson, G. H. Photobleaching and photoactivation: following protein dynamics in living cells. *Nat Cell Biol* **2003**, *Suppl*, S7-14.

39. Meyer, T.; Begitt, A.; Vinkemeier, U. Green fluorescent protein-tagging reduces the nucleocytoplasmic shuttling specifically of unphosphorylated STAT1. *FEBS J* **2007**, *274*, 815-26.

40. Dopie, J.; Skarp, K. P.; Rajakyla, E. K.; Tanhuanpaa, K.; Vartiainen, M. K. Active maintenance of nuclear actin by importin 9 supports transcription. *Proc Natl Acad Sci U S A*, *109*, E544-52.

41. Howell, J. L.; Truant, R. Live-cell nucleocytoplasmic protein shuttle assay utilizing laser confocal microscopy and FRAP. *Biotechniques* **2002**, *32*, 80-2, 84, 86-7.

42. Birbach, A.; Bailey, S. T.; Ghosh, S.; Schmid, J. A. Cytosolic, nuclear and nucleolar localization signals determine subcellular distribution and activity of the NF-kappaB inducing kinase NIK. *J Cell Sci* **2004**, *117*, 3615-24.

43. Akoumianaki, T.; Kardassis, D.; Polioudaki, H.; Georgatos, S. D.; Theodoropoulos, P. A. Nucleocytoplasmic shuttling of soluble tubulin in mammalian cells. *J Cell Sci* **2009**, *122*, 1111-8.

44. Roth, D. M.; Harper, I.; Pouton, C. W.; Jans, D. A. Modulation of nucleocytoplasmic trafficking by retention in cytoplasm or nucleus. *J Cell Biochem* **2009**, *107*, 1160-7.

45. Drake, K. R.; Kang, M.; Kenworthy, A. K. Nucleocytoplasmic distribution and dynamics of the autophagosome marker EGFP-LC3. *PLoS One*, *5*, e9806.

46. Hildick, K. L.; Gonzalez-Gonzalez, I. M.; Jaskolski, F.; Henley, J. M. Lateral diffusion and exocytosis of membrane proteins in cultured neurons assessed using fluorescence recovery and fluorescence-loss photobleaching. *J Vis Exp.*

47. van Royen, M. E.; Dinant, C.; Farla, P.; Trapman, J.; Houtsmuller, A. B. FRAP and FRET methods to study nuclear receptors in living cells. *Methods Mol Biol* **2009**, *505*, 69-96.

48. Lillemeier, B. F.; Koster, M.; Kerr, I. M. STAT1 from the cell membrane to the DNA. *EMBO J* **2001**, *20*, 2508-17.

49. Yu, W.; Kawasaki, F.; Ordway, R. W. Activity-dependent interactions of NSF and SNAP at living synapses. *Mol Cell Neurosci*, *47*, 19-27.

50. Schindler, M.; Jiang, L. W. Nuclear actin and myosin as control elements in nucleocytoplasmic transport. *J Cell Biol* **1986**, *102*, 859-62.

51. Oma, Y.; Harata, M. Actin-related proteins localized in the nucleus: from discovery to novel roles in nuclear organization. *Nucleus*, *2*, 38-46.

52. Hofmann, W. A., *et al.* SUMOylation of nuclear actin. *J Cell Biol* **2009**, *186*, 193-200.

53. Yamagata, M.; Sanes, J. R.; Weiner, J. A. Synaptic adhesion molecules. *Curr Opin Cell Biol* **2003**, *15*, 621-32.

54. Stoltzman, C. A.; Kaadige, M. R.; Peterson, C. W.; Ayer, D. E. MondoA senses non-glucose sugars: regulation of thioredoxin-interacting protein (TXNIP) and the hexose transport curb. *J Biol Chem* **2011**, *286*, 38027-34.

55. Sans, C. L.; Satterwhite, D. J.; Stoltzman, C. A.; Breen, K. T.; Ayer, D. E. MondoA-Mlx heterodimers are candidate sensors of cellular energy status: mitochondrial localization and direct regulation of glycolysis. *Mol Cell Biol* **2006**, *26*, 4863-71.

56. Percipalle, P.; Visa, N. Molecular functions of nuclear actin in transcription. *J Cell Biol* **2006**, *172*, 967-71.

57. Hofmann, W. A., *et al.* Actin is part of pre-initiation complexes and is necessary for transcription by RNA polymerase II. *Nat Cell Biol* **2004**, *6*, 1094-101.

58. Percipalle, P.; Raju, C. S.; Fukuda, N. Actin-associated hnRNP proteins as transacting factors in the control of mRNA transport and localization. *RNA Biol* **2009**, *6*, 171-4.

59. Elad, N.; Maimon, T.; Frenkiel-Krispin, D.; Lim, R. Y.; Medalia, O. Structural analysis of the nuclear pore complex by integrated approaches. *Curr Opin Struct Biol* **2009**, *19*, 226-32.

60. Allen, T. D.; Cronshaw, J. M.; Bagley, S.; Kiseleva, E.; Goldberg, M. W. The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. *J Cell Sci* **2000**, *113 (Pt 10)*, 1651-9.

61. Wang, R.; Brattain, M. G. The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. *FEBS Lett* **2007**, *581*, 3164-70.

62. Wada, A.; Fukuda, M.; Mishima, M.; Nishida, E. Nuclear export of actin: a novel mechanism regulating the subcellular localization of a major cytoskeletal protein. *EMBO J* **1998**, *17*, 1635-41.

63. Uttamapinant, C., *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A*, *107*, 10914-9.

64. Graf, E. R.; Kang, Y.; Hauner, A. M.; Craig, A. M. Structure function and splice site analysis of the synaptogenic activity of the neurexin-1 beta LNS domain. *J Neurosci* **2006**, *26*, 4256-65.

65. Graf, E. R.; Zhang, X.; Jin, S. X.; Linhoff, M. W.; Craig, A. M. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell* **2004**, *119*, 1013-26.

66. Cline, H. Synaptogenesis: a balancing act between excitation and inhibition. *Curr Biol* **2005**, *15*, R203-5.

67. Scheiffele, P.; Fan, J.; Choih, J.; Fetter, R.; Serafini, T. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* **2000**, *101*, 657-69.

68. Graff, J. M.; Stumpo, D. J.; Blackshear, P. J. Molecular cloning, sequence, and expression of a cDNA encoding the chicken myristoylated alanine-rich C kinase substrate (MARCKS). *Mol Endocrinol* **1989**, *3*, 1903-6.

69. Chen, S. X.; Tari, P. K.; She, K.; Haas, K. Neurexin-neuroligin cell adhesion complexes contribute to synaptotropic dendritogenesis via growth stabilization mechanisms in vivo. *Neuron*, *67*, 967-83.

70. Paraoanu, L. E.; Becker-Roeck, M.; Christ, E.; Layer, P. G. Expression patterns of neurexin-1 and neuroligins in brain and retina of the chick embryo: Neuroligin-3 is absent in retina. *Neurosci Lett* **2006**, *395*, 114-7.

71. Levinson, J. N.; El-Husseini, A. A crystal-clear interaction: relating neuroligin/neurexin complex structure to function at the synapse. *Neuron* **2007**, *56*, 937-9.

72. Chen, X.; Liu, H.; Shim, A. H.; Focia, P. J.; He, X. Structural basis for synaptic adhesion mediated by neuroligin-neurexin interactions. *Nat Struct Mol Biol* **2008**, *15*, 50-6.

73. Peterson, C. W.; Ayer, D. E. An extended Myc network contributes to glucose homeostasis in cancer and diabetes. *Front Biosci* **2011**, *17*, 2206-23.

74. Cardoso, S., *et al.* Insulin is a two-edged knife on the brain. *J Alzheimers Dis* **2009**, *18*, 483-507.

75. Fridlyand, L. E.; Ma, L.; Philipson, L. H. Adenine nucleotide regulation in pancreatic beta-cells: modeling of ATP/ADP-Ca2+ interactions. *Am J Physiol Endocrinol Metab* **2005**, *289*, E839-48.

76. Fridlyand, L. E.; Philipson, L. H. Glucose sensing in the pancreatic beta cell: a computational systems analysis. *Theor Biol Med Model*, *7*, 15.

77. Gao, Z. Y.; Li, G.; Najafi, H.; Wolf, B. A.; Matschinsky, F. M. Glucose regulation of glutaminolysis and its role in insulin secretion. *Diabetes* **1999**, *48*, 1535-42.

78. Stoltzman, C. A., *et al.* Glucose sensing by MondoA:Mlx complexes: a role for hexokinases and direct regulation of thioredoxin-interacting protein expression. *Proc Natl Acad Sci U S A* **2008**, *105*, 6912-7.

Chapter 4. Considerations for the directed evolution of LplA and introduction to evolution methods

The work described in this chapter is unpublished.

Introduction

Using LplA as a platform for labeling both intracellular and cell surface proteins has proven very advantageous, and the extension of the PRIME methodology to interaction-dependent labeling (ID-PRIME) has increased the range of biological problems that can be studied with this method. However, a major limitation of PRIME is that LplA and its mutants have low activity in the secretory pathway[1]. The remaining chapters of this thesis focus on the use of directed evolution techniques to produce LplA variants with improved activity in the secretory pathway and cell surface.

Before beginning the directed evolution of LplA, considerable effort was made to rationally engineer the enzyme. While these efforts were not successful in producing LplA variants with improved catalytic activity in the endoplasmic reticulum, they did inform the approaches used for the directed evolution of the enzyme. In this chapter, we first present a condensed summary of the rational design efforts and outline the goals for the directed evolution of LplA. Next, we introduce common methods of directed evolution and the criteria used to select methods for LplA evolution. Finally, we will present the initial experiments that were performed before focusing on the yeast-display directed evolution strategies summarized in Chapters 5, 6, and 7.

Rational engineering of LplA

LplA oxidation and stability

One of the first valuable observations we made about LplA is that storage in reducing buffer improves and preserves the activity of the purified enzyme. These results are supported by data from Green and colleagues, who noted that LplA activity declined during storage to ~15% of original activity after 3 months[2]. LplA has six surface cysteines, none of which are oxidized in the native form of the protein[2]. Green and colleagues hypothesized that this inactivation was due to oxidation of the protein, creating an intramolecular disulfide. This was supported by the determination that, while active LplA had six reactive thiol groups per molecule, the storage-inactivated LplA had just four reactive thiol groups per molecule. Interestingly, we found that some mutants of LplA are more susceptible to storage inactivation, but that storage in buffer containing

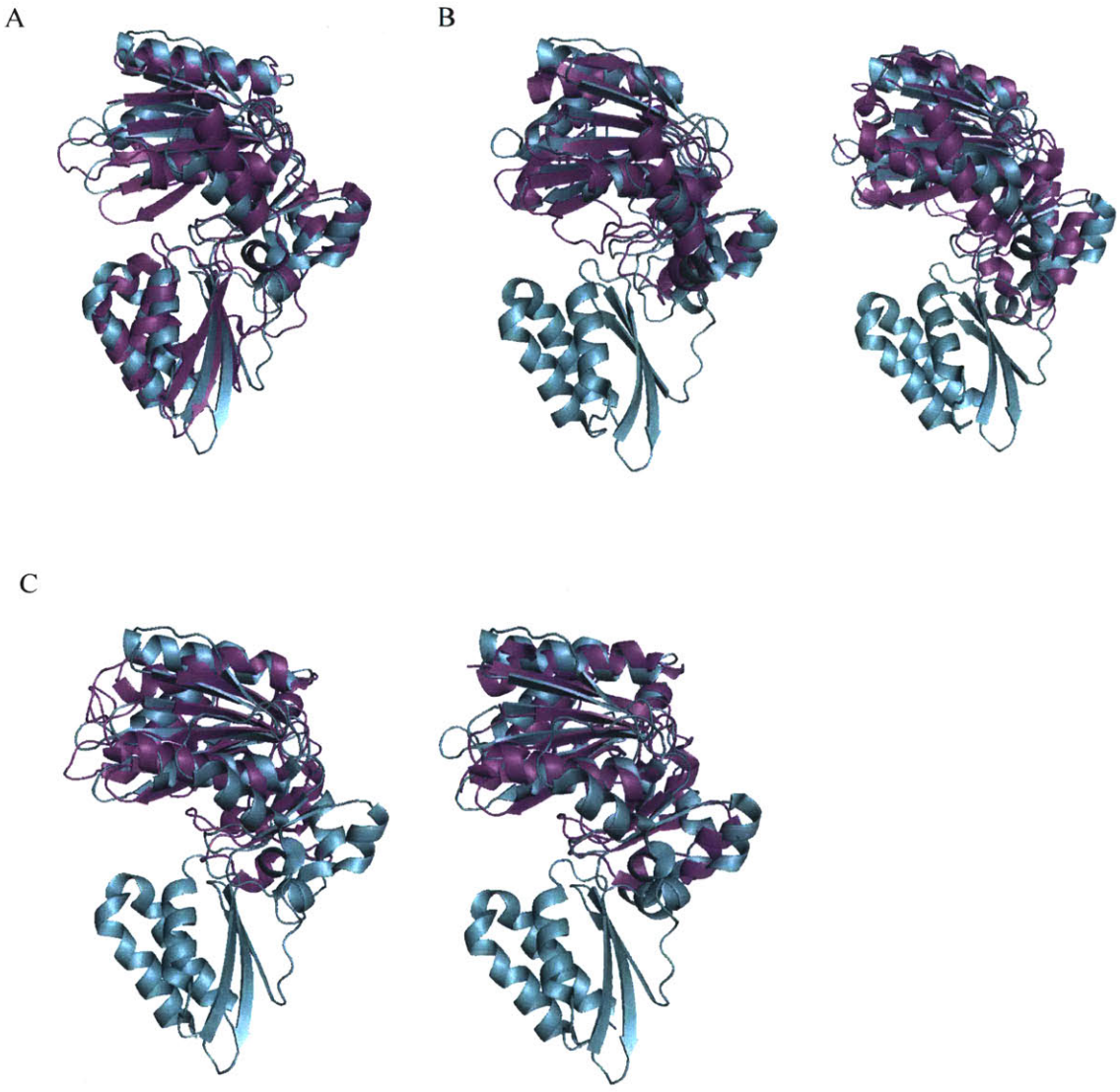


Figure 4-1. Structural homology of Lpls. (A) Crystal structure of apo *E. coli* LplA (cyan, PDB 1X2G) aligned with crystal structure of apo Lpl from *S. Pneumoniae* (magenta, PDB 1VQZ). Note that both enzymes have a large N-terminal and a smaller C-terminal domain. **(B)** Crystal structure of apo *E. coli* LplA (cyan) aligned with apo *T. acidophilum* LplA (magenta PDB 2ARS, left) and with apo *S. agalactiae* (magenta, PDB 2P0L, right). **(C)** Crystal structures of lipoyltransferases aligned with LplA. Crystal structure of apo *E. coli* LplA (cyan) aligned with apo *M. tuberculosis* lipoyltransferase (magenta, PDB 1W66, left) and with apo *T. thermophilus* (magenta, PDB 2HQV, right). All structure figures and alignments were generated in PyMol.

Another interesting result from the attempts to rationally engineer LplA is that C-terminal truncations of LplA increase enzyme stability and solubility. When we made truncations of LplA at Pro249 (removing the entire C-terminal domain, *vide infra*) and expressed the protein in *E. coli*, the soluble fraction of the protein increased; this effect was even more pronounced for mutants of LplA that were difficult to express (data not shown). When we assayed the activity of the Pro249 LplA truncation, we observed that, while lipoyltransferase activity was unaffected, the truncated enzyme had severely decreased adenylation activity. Intrigued by this, we wanted to explore the potential for engineering the activity of the N-terminal domain of LplA to achieve site-specific labeling with this smaller enzyme.

As previously discussed, LplA catalyzes the ligation of lipoic acid in a two-step reaction: an adenylation step, where lipoic acid is activated with ATP to form lipoyl-AMP, then a lipoyltransferase step, where the lipoate is transferred to the protein substrate. While *E. coli* LplA can perform both steps of the lipoylation reaction, some species require a separate enzyme for each step. Many bacterial Lpls have two domains, one large N-terminal domain and a small C-terminal domain. Examples of this are *E. coli* LplA (PDB 1X2G)[3] and Lpl from *Streptococcus pneumoniae*, (PDB 1VQZ) which are structurally homologous (Figure 4-1). These enzymes are capable of catalyzing both steps of the reaction. On the other hand, several species have Lpls with just one domain that is structurally homologous to the N-terminal domain of *E. coli* LplA. Examples of this are Lpls from *Thermoplasma acidophilum* (PDB 2ART and 2C8M)[4,5] and *Streptococcus agalactiae* (PDB 2P0L) (Figure 4-1B). Only the Lpl from *T. acidophilum* has been characterized in terms of lipoylation activity and can catalyze only the adenylation step of the reaction[5]. This information leads to the natural hypothesis that the N-terminal domain of *E. coli* LplA is responsible for the adenylation step and the C-terminal domain is responsible for the transfer step. However, when we consider the structures of various bacterial lipoyltransferases that have been solved (*Mycobacterium tuberculosis*, PDB 1W66; *T. thermophilus,* PDB 2QHV)[6,7], these enzymes do not have adenylation activity, but their structures also align with just the N-terminal domain of LplA (Figure 4-1C). Based on the structures available, the relationship between the multi-domain LplA enzyme and its two-step enzymatic activity was not clear.

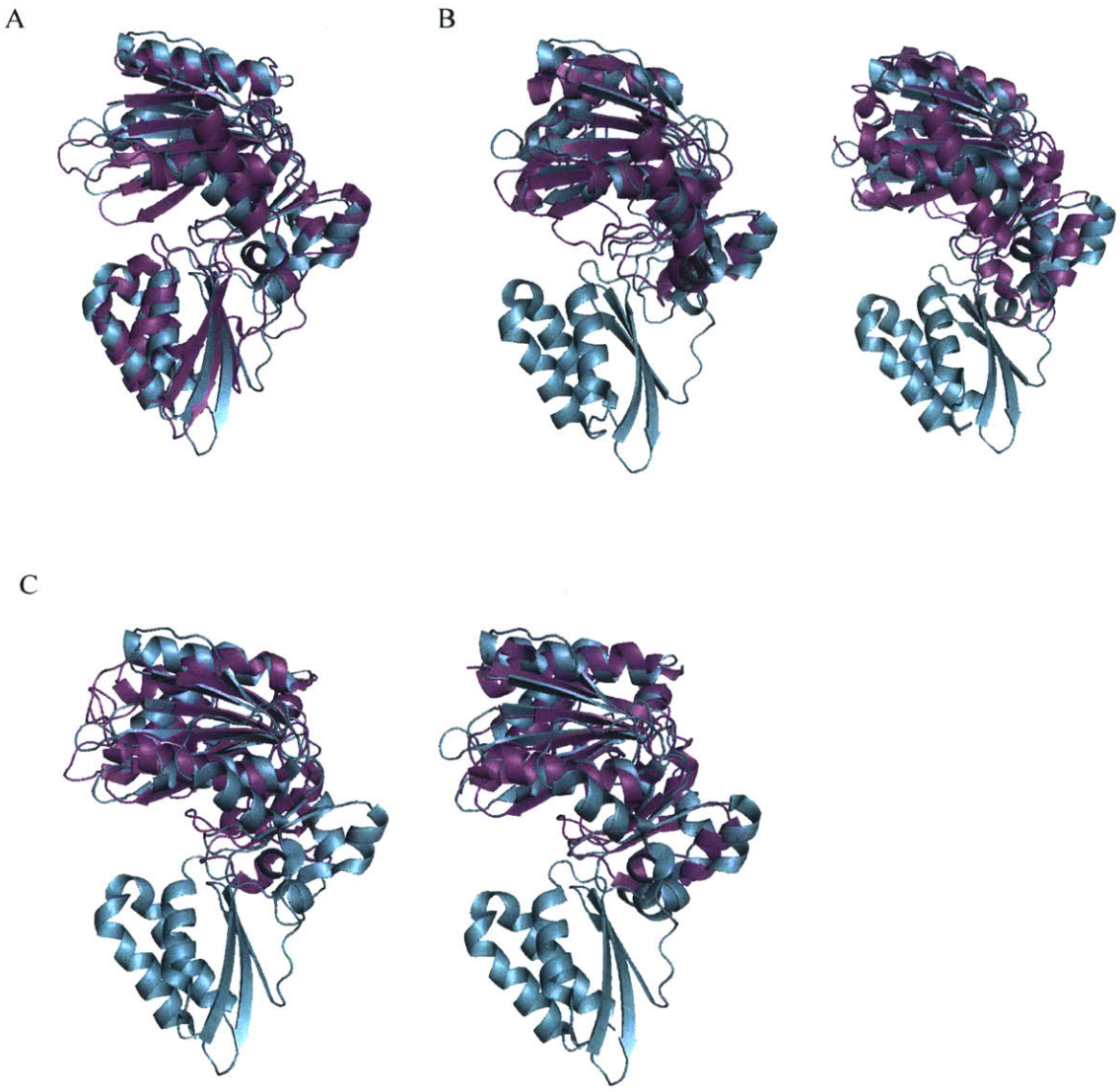


Figure 4-1. Structural homology of Lpls. (A) Crystal structure of apo *E. coli* LplA (cyan, PDB 1X2G) aligned with crystal structure of apo Lpl from *S. Pneumoniae* (magenta, PDB 1VQZ). Note that both enzymes have a large N-terminal and a smaller C-terminal domain. **(B)** Crystal structure of apo *E. coli* LplA (cyan) aligned with apo *T. acidophilum* LplA (magenta PDB 2ARS, left) and with apo *S. agalactiae* (magenta, PDB 2P0L, right). **(C)** Crystal structures of lipoyltransferases aligned with LplA. Crystal structure of apo *E. coli* LplA (cyan) aligned with apo *M. tuberculosis* lipoyltransferase (magenta, PDB 1W66, left) and with apo *T. thermophilus* (magenta, PDB 2HQV, right). All structure figures and alignments were generated in PyMol.

Soon after starting this project, a crystal structure of *E. coli* LplA with the bound lipoyl-AMP intermediate was published that provided insight into the interaction and functions of the N- and C-terminal domains[8]. This crystal structure demonstrated that LplA undergoes a global conformational change between the apo form of the protein and the lipoyl-AMP-bound form of the protein (Figure 4-2). In the apo form of the protein, the N-terminal and C-terminal domains are linked by a short loop (red). Upon formation of the adenylate ester, LplA undergoes three conformational changes. The adenylate binding loop (magenta) and the lipoate binding loop (yellow) close down on the lipoyl-AMP (cyan), shielding the activated intermediate from solvent. The third conformational change is a huge 180 rotation of the C-terminal domain (green). The rotation of this domain is thought to be required for proper protein substrate binding and lipoyltransfer. This hypothesis is supported by the observation that bovine lipoyltransferase is structurally homologous to the activated stretched conformation of *E. coli* LplA (Figure 4-3). However, it is not known if flipping of the C-terminal domain is required for LAP2 binding.

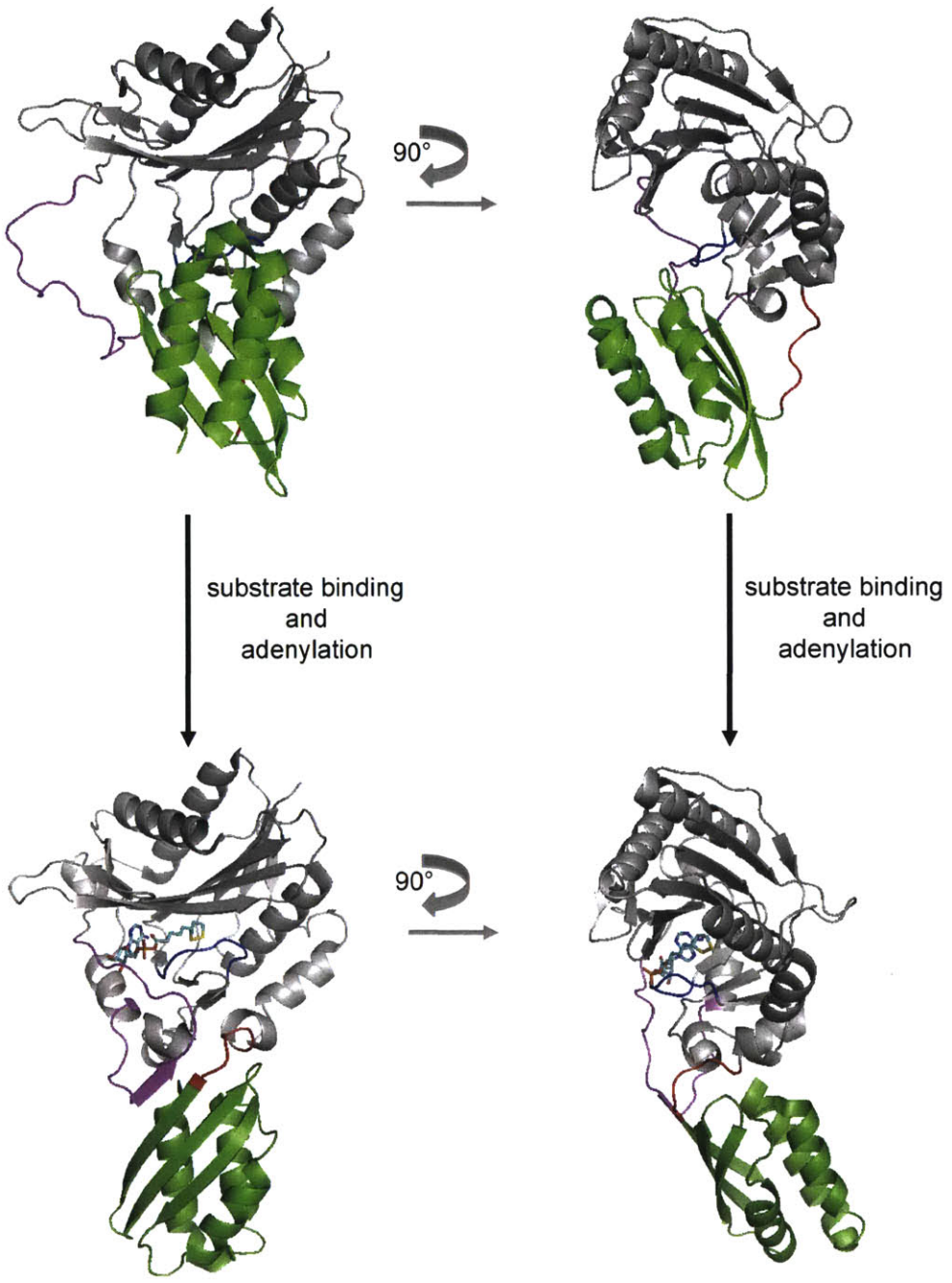


Figure 4-2. *E. coli* LplA undergoes a global conformational change. Top structure is of apo *E. coli* LplA. The N-terminal domain (grey) and C-terminal domain (green) are connected by a loop (red). The lipoate-binding loop (blue) and adenylate binding loop (magenta) both close over the lipoyl-AMP (cyan) upon adenylate formation (bottom structures, PDB 3A7R) and the C-terminal domain rotates 180 degrees. All structure views created using PyMol.

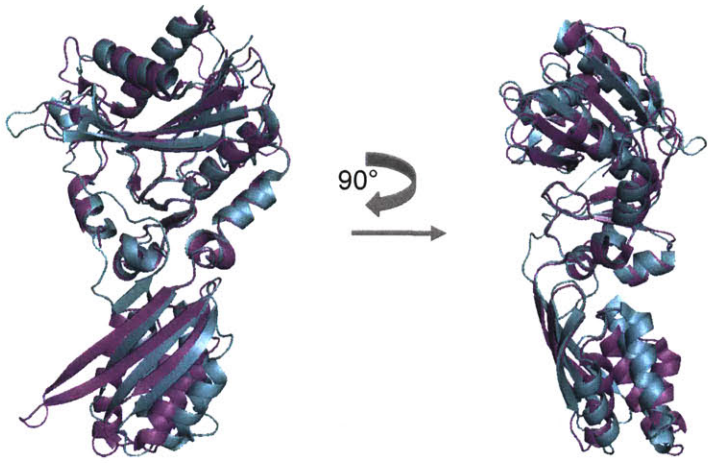


Figure 4-3. Comparison of bovine lipoyltransferase and lipoyl-AMP bound *E. coli* LplA. *E. coli* LplA with lipoyl-AMP bound (PDB 3A7R) is aligned to the apo form of bovine lipoyltransferase (PDB 3A7U). All structure views were created using PyMol.

This data helps explain the results we observed with truncated LplA. With the truncated LplA, we observed lipoyltransferase activity similar to wild type. We hypothesize that since the C-terminal domain is deleted, and therefore not blocking or controlling protein binding, the transfer step is not affected. The decrease in adenylation activity may have several sources. First, it could be that the 180° rotation of the C-terminal domain comes before the chemistry step for lipoyl-AMP formation and is required for efficient adenylation. Little data exists to support this hypothesis, and the truncated LplA does have some adenylation activity; however, as we discuss in Chapters 6 and 7, full-length LplA variants with improved catalytic activity have highly mutated C-terminal regions, which may indicate a relationship between adenylation and the C-terminal domain. Second, we noticed that in the lipoyl-AMP bound form of LplA, part of the adenylation loop interacts with residues 250-255 in the C-terminal domain (Figure 4-4). It may be that when these residues are not present, the adenylation loop has fewer interactions to induce closure over the small-molecule binding pocket and lipoyl-AMP formation is inhibited by solvent exposure. This hypothesis is supported by the fact that a truncation of LplA at Glu256 has increased adenylation activity compared to the Pro249 truncation, although is not as robust as full-length LplA (see Chapter 5). Attempts at improving adenylation activity by adding mutations or further adjusting the extent of LplA truncation were not successful (data not shown).

Figure 4-4. Residues 250-256 of LplA may be necessary for closing adenylation loop. Crystal structure of *E. coli* LplA with lipoyl-AMP bound (PDB 3A7R). The region depicted in red corresponds to residues 1-249, and represents the first truncated LplA tested. The lipoate-binding loop (blue) and adenylate-binding loop (magenta) both close over the lipoyl-AMP (cyan) upon adenylate formation and the adenylate-binding loop interacts with residues 250-256 (yellow) in the C-terminal domain. These residues were added on to the second truncated LplA tested (LplA-NTD, see Chapter 5). Structure view created in PyMol.

We determined that directed evolution of LplA might produce more promising results than the moderate successes obtained from rational design. Based on the exciting results obtained using the N-terminal domain of LplA, we decided to proceed with a two-pronged directed evolution approach. One arm of the project focused on the evolution of LplA-NTD and the other arm was geared towards the evolution of full-length LplA. The overarching goal for the evolution projects was to produce LplA variants with improved activity in the endoplasmic reticulum and cell surface. However, there were additional preliminary goals for the evolution of LplA-NTD. The first goal was to simply develop an LplA-NTD variant with adenylation activity as good or better than full-length LplA. The second goal was to then select for LplA-NTD mutants that were capable of ligating unnatural probes. With these goals in mind, we explored potential avenues for directed evolution of enzymatic activity.

Methods of directed protein evolution

Directed protein evolution is a powerful technique for improving existing properties of proteins and has even been used to produce new protein function. While rational design can be a valuable tool, it requires extensive knowledge about the

functional, structural, and mechanistic properties of an enzyme. Directed enzyme evolution, on the other hand, does not require *a priori* knowledge about the protein's properties in order to produce the desired characteristics. The general strategy for directed evolution is to first create a diverse library of genes that encode the expression of a collection of protein variants. Then, using a phenotype selection strategy, the proteins with the desired attributes are isolated from the library and the genes encoding that population of proteins are amplified. Through iterative rounds of gene expression, phenotype selection, and gene amplification, the library becomes enriched in genes that encode for proteins with the desired properties. The crux of directed evolution is that there must be a stable link between phenotype and genotype.

In Darwinian evolution, the membranes of the living cell serve to encapsulate proteins and thus connect genotype and phenotype. There are various *in vivo* directed evolution methods that mimic nature and use cells to preserve this link. With *in vivo* directed evolution, the gene diversification, protein selection, and gene amplification steps all occur inside living cells[9]. In addition, there are methods that we will call semi-*in vitro* that perform the gene diversification step outside of living cells, but then the diversified library is introduced into living cells via transformation or transfection. All subsequent selection and gene amplification steps occur inside cells. Finally, there are also several *in vitro* directed evolution methods that work outside of the confines of the living cell and thus require different methods of linking genotype and phenotype. We will briefly discuss existing *in vivo* directed evolution methods and then present the semi-*in vitro* and *in vitro* directed evolution methods that are most frequently used in the literature.

In vivo directed evolution

In vivo methods require mutation and gene diversification to occur within the cell, which is usually achieved by using *E. coli* strains that have introduced error-prone polymerases or have knocked-down DNA repair pathways[10]. Controlling the error rate in the target gene is very difficult using these methods. When the repair pathways are knocked down, only low mutation rates (~1 mutation per 2kb) are achievable[10,11] and endogenous genes will be mutated as well. The use of error-prone polymerases increases

the target gene error rate, but still presents the issue of introducing mutations into endogenous genes as well as target genes. To circumvent this problem, the Loeb lab developed a method that uses an error-prone Pol1 polymerase[12]. Since PolI is primarily involved in replication of the ColE1 plasmid, this technique leaves most endogenous proteins unaffected. This method produces error rates of approximately 2 mutations per 2kb, but because PolI is primarily involved in transcription initiation, these mutations are maximal in the first 700bp from the origin of replication[9,12].

After library diversification, *E. coli* can be selected using either drug resistance markers or screened for engineered manifestations of the desired property. In general, prokaryotic evolution methods are desirable because genome engineering is straightforward, culturing is simple, doubling time is fast, and libraries can be made and selected in a short time[9]. Recent work has even demonstrated that entire prokaryotic pathways can be adjusted using *in vivo* directed evolution[13]; including pathways to produce chiral drug intermediates[14] and useful alcohol-based fuels[15]. However, the prokaryotic methods are limited to proteins that express well inside prokaryotic cells. Any targets that require eukaryotic-specific post-translational modifications or folding pathways will not be amenable to this method. Since the goal for the directed evolution is to produce an LplA variant with improved activity in the secretory pathway, prokaryotic directed evolution is not an appropriate platform for LplA evolution.

When evolving eukaryotic proteins, an *in vivo* eukaryotic method is ideal since the protein can evolve in a natural environment, with all of the folding and post-translational modification machinery in place. Some eukaryotic *in vivo* evolution methods use modified yeast strains where the methods of gene diversification are similar to those used in prokaryotic methods, such as using repair-deficient yeast strains[9]. These methods have the same limitations in terms of genetic diversity as their prokaryotic counterparts. To address this issue of genetic diversity, researchers can take advantage of the somatic hypermutation that occurs in B-cells[16-18]. Somatic hypermutation is used in the immune response to randomize the variable regions of immunoglobulin (Ig) light and heavy chains. By inserting the target gene into regions flanking the hypermutable Ig genes, mutation rates of up to 2 mutations per 1kb can be achieved[9]. However, delivery at

the target locus flanking the Ig site is difficult to achieve specifically; most examples use random integration using viruses and then subsequent screening for proper transgene integration at the desired locus. To address this issue of low ratio of targeted to random transgene delivery, labs have turned to B cells from chicken or swine, which are able to undergo homologous recombination and thus can be used to achieve targeted transgene integration[19,20]. Eukaryotic *in vivo* methods have been used to evolve antibodies[20,21] and fluorescent proteins[16]. However, when working with living cells it is difficult to control selection conditions for improving enzyme activity, and even then, the organism may accommodate selective pressure through the effects of off-target proteins or pathways. While this is a problem for the evolution of specific activity in a single protein, this feature is one of the reasons these *in vivo* methods are perfect platforms for evolving metabolic pathways or complex protein networks[9].

Semi-in vitro directed evolution

A significant drawback to using an *in vivo* method is the difficulty in controlling the mutation rate and diversification of the target genes. This is why many directed evolution methods utilize a semi-*in vitro* approach where genetic library design and diversification are performed *in vitro*, but cells are used for the selection and gene amplification steps. For example, powerful error-prone PCR or DNA shuffling techniques can be used to generate a large library of gene variants, which can then be transformed into *E. coli* or yeast or transfected into mammalian cells for intracellular expression and subsequent downstream selections and amplification. Unlike *in vivo* methods, where gene diversification occurs simultaneously with gene amplification, when using semi-*in vitro* methods the gene library must be isolated from the host and re-diversified *in vitro*, if desired. One significant limitation to most semi-*in vitro* methods is that the selection strategy must be cell-compatible (selections using phage display are the exception).

One benefit of performing the gene diversification process *in vitro* is that the mutation rate can be exquisitely controlled. *In vitro* library generation techniques also allow researchers to focus mutations in specific areas of the protein of interest, which can be extremely valuable when evolving enzyme activity[22]. Moreover, techniques such as

error-prone PCR can produce very high mutation rates: up to 15 mutations per 1 kb when only altered salt concentrations are used[10], and rates as high as 200 mutations per 1kb when dNTP analogues are included[23]. While gene libraries generated by *in vitro* techniques can easily have $>10^{11}$ unique members[10], the overall library size using a semi-*in vitro* method is limited by the delivery efficiencies into the cells being used (bacteria, $<10^{10}$; yeast, ($<10^{8}$) ; mammalian cells ($<10^{8}$)[24-27].

In addition to intracellular methods, there are various semi-*in vitro* directed evolution methods that use displayed protein for selections. Existing display strategies are phage display, bacterial display, yeast-display, and mammalian display. The most frequently used of these methods is M13 phage display[28,29]. In this method, the protein to be evolved is displayed on the surface of the M13 filamentous phage particle, and the gene encoding the protein is packaged within the phage particle. Phage-display is semi-*in vitro* because the phage must be amplified within infected *E. coli* and there is an *in vivo* selection against proteins that are unstable or poorly exported. Phage-display has most frequently been used to evolve new antibodies[30,31] and binding or substrate peptides[32-34]. Unfortunately, the use of phage-display to select for enzyme catalysis has been limited. Some groups have used phage to display enzyme variants and select for binding to transition-state analogues or known mechanism-based inhibitors[29], with limited success[35]. Others have used a strategy where the substrate is covalently attached to the phage and product is monitored by intra-molecular reaction of the enzyme and substrate on the phage particle itself[36,37]. However, reported enrichment factors in model selections are very low (<10-fold), indicating this strategy has limited dynamic range and does not adequately discriminate between active and inactive library members.

Bacterial-display, yeast-display and mammalian cell display are all related methods that have been used for protein evolution. These methods display the proteins on the cell surface by making library fusions to surface proteins or to transmembrane domains[38-40]. Through technology developments, the issue of library delivery and expression control has been solved for both bacterial[39] and yeast surface display[41] through technology developments. However, library delivery is still problematic for mammalian cell surface display. For example, stable transfection of foreign DNA

produces multiple integrations into the host cell genome, often resulting in multiple library members expressed in the same cell[42]. To address this issue, viral infection has frequently been used in an attempt to control multiplicity of plasmid delivery[43,44]. Recently, Amgen has published a method that uses homologous recombination to achieve defined integration at a single site in the genome[47]. This is a considerable technological advance, which should make mammalian surface display more amenable to enzyme evolution. Still, there have been several reports of directed enzyme evolution using bacterial display[45] and yeast display[25,46], yet there have been no such examples of enzyme evolution using mammalian cell display.

In-vitro directed evolution

When using *in vitro* directed evolution methods, all steps of the evolution process: gene diversification, protein translation and selection, and gene amplification, are performed *in vitro*. These methods require specific techniques to link genotype and phenotype. Two of the most frequently used *in vitro* methods for providing this link are mRNA display and ribosome display[24,48]. Ribosome display uses mRNA sequences that have been engineered without a stop codon. When performing translation of these engineered sequences, the ribosome stalls on the mRNA producing a complex of translated protein and mRNA linked via the halted ribosome. Ribosome display has been used to evolve peptide binders[49,50], antibodies[51-53], and proteins with improved stability[54]. The Szostak group developed a conceptually similar approach called mRNA display[55]. In mRNA display, the ribosomal inhibitor puromycin is conjugated to the mRNA; when proteins are translated from the mRNA and when the puromycin terminator is reached, a covalent link between the newly translated protein and the mRNA is formed.

In vitro compartmentalization (IVC) is a technique that mimics the cell by encapsulating the genetic library in aqueous droplets of an oil-water emulsion[48,56]. This is done at concentrations such that, at most, only one DNA copy exists per compartment. This method allows very large libraries to be screened; for example, a library of 10^{10} members can be effectively compartmentalized in just 50 μL of an emulsion[48]. This method is often combined with mRNA or ribosome display to improve control of reaction conditions, concentrations, and protein expression. Compartmentalization serves

to prevent product or protein diffusion and to standardize protein expression by ensuring equal distribution of *in vitro* translation components[56]. This technique has been used to evolve several binding proteins[57], including an RNA-binding protein with altered specificity[58], and p53 variant with improved recognition of low-response elements[59]. Enzyme activity is more difficult to evolve; but *in vitro* compartmentalization has been used to change and improve the substrate specificity of a methyltransferase[60] and change the substrate recognition of endonucleases[61]. It is telling that the majority of these successes are for enzymes or proteins that directly bind DNA or RNA. This is due to the limited methods to link product to DNA using a completely *in vitro* system. Recently Griffiths and colleagues have combined IVC with FACS to circumvent this issue and have been able to evolve improved activity of a phophotriesterase[62]. This result and other similar approaches are discussed in detail in Chapter 7.

One benefit of using a purely *in vitro* directed evolution method is that larger libraries can be created and selected than what are feasible using semi-in vitro methods. In theory, library sizes as large as 10^{15} can be generated and screened, although in practice library sizes usually range from 10^{11}-10^{13} [48,63-65]. Therefore, these libraries can be 10^2-10^4 times larger than those practically obtained using the previously described methods. Moreover, since *in vitro* evolution methods utilize *in vitro* protein translation, difficult targets can be evolved, such as extremely unstable or even toxic proteins. *In vitro* methods are also advantageous because selection conditions can be easily modified and selection strategies that are incompatible with living cells are possible. One drawback, however, is that without the encapsulating membranes of a cell, the method of linking genotype-phenotype must be carefully considered. More importantly, if enzyme activity is the goal of the selection, the enzymatic product must somehow be associated with the resulting DNA or encapsulated using IVC.

Criteria for a directed evolution method

The goal for these projects is to evolve an LplA variant that has increased activity in the secretory pathway of the living cell. Using eukaryotic cells would allow us to evolve LplA in the environment where we will later be applying the enzyme. The ability to use powerful *in vitro* library generation methods such as error-prone PCR or DNA

shuffling is also desirable, particularly because the mechanism for LplA activity loss in the secretory pathway is not understood. For these two reasons, a semi-*in vitro* eukaryotic cell-based method would be ideal for LplA evolution. The final question is whether to use a method that displays the enzyme on the surface of a cell or to rely on intracellular labeling and detection. We decided to investigate the use of yeast-display and mammalian cell-display since a display method permits us to monitor enzyme expression level by antibody staining while using a variety of methods to detect labeling signal. Moreover, since the enzymes displayed on the cell surface have been trafficked through the endoplasmic reticulum, these methods allow for the selection of activity in the secretory pathway.

In the results section of this chapter, we describe our initial attempts at using mammalian-cell based FACS to evolve improved full-length LplA activity. While model selections in mammalian cells looked promising, we will discuss the various issues that made this evolution route untenable.

Results

Demonstration of cis labeling of E2p by LplA when co-expressed on HEK cell surface.

For this evolution strategy, we decided to use HEK cells since these cells are small, have high transfection efficiencies, and fast doubling times. We proposed using a selection platform where the ligase and peptide substrate are co-displayed on the mammalian cell surface. Using this platform, enzymatic activity can be assayed by incubating cells with probe and ATP and monitoring the cis labeling of the substrate by the co-expressed enzyme.

Our lab has previously demonstrated that LplA substrates can be highly expressed on the mammalian cell surface as fusions to CFP with the inclusion of a single transmembrane domain (LAP2-CFP-TM and E2p-CFP-TM). Unfortunately, when ^{WT}LplA was fused to an FP-TM in an identical geometry, the fusion had extremely poor surface delivery and formed aggregates along the secretory pathway (data not shown). To address this issue, we tested the surface delivery of ^{WT}LplA fusions to other surface proteins (neurexin-1β, LDLR, and neuroligin-1). Of all the constructs tested, ^{WT}LplA-

neurexin1β fusion gave the best surface expression (data not shown). For this reason, all subsequent mammalian cell-surface optimizations and testing use LplA-neurexin-1β fusions. While the downstream goal of this evolution is to evolve LplA activity for the labeling of LAP2 fusions, LplA has higher cis lipoylation activity with E2p on the cell surface (data not shown). For this reason, model selections and initial optimizations of mammalian cell labeling use co-displayed E2p-CFP-TM

The first step in the development of a HEK cell evolution platform was to determine whether the cis labeling strategy described above could effectively distinguish between highly active and moderately active or inactive ligases. As proof-of-principle, we first performed the simplest experiment using a cis lipoylation assay (Figure 4-5A). In this assay, we determined the cis lipoylation activity of ^{WT}LplA and compared this to the cis lipoylation activity of two less active variants (^{W37V}LplA or ^{W37A}LplA) or an inactive variant (^{K133R}LplA). After labeling and performing antibody staining for lipoylation and for enzyme expression level, we analyzed the samples by fluorescence imaging and fluorescence activated cell sorting (FACS) (Figure 4-5B-C). From these data it is clear that the proposed selection scheme can effectively distinguish between active and less active ligases using both fluorescence imaging and fluorescence activated cell sorting. Interestingly we were even able to distinguish between the cis lipoylation activity of ^{W37V}LplA and the slightly more active ^{W37A}LplA mutant.

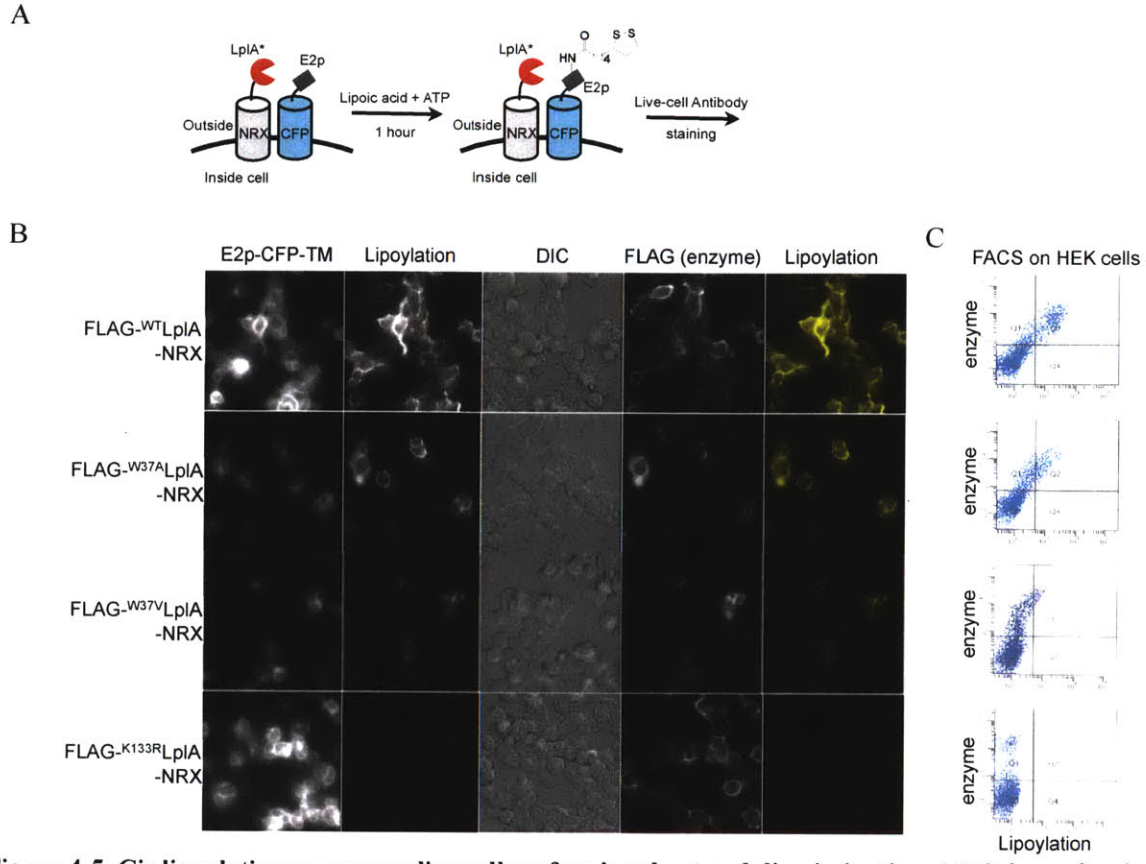


Figure 4-5. Cis lipoylation on mammalian cell surface is robust and discriminating. (A) Scheme for cis labeling on the mammalian cell surface. E2p-CFP-TM and LplA-neurexin-1β (NRX) are co-expressed on the mammalian cell surface. Cells are incubated with lipoic acid and ATP and the enzyme is capable of lipoylating the E2p substrate. The lipoylation is detected by antibody staining (yellow) and FLAG staining reports on the enzyme expression level. **(B)** Fluorescence imaging of cis lipoylation on HEK cell surface. Cells were prepared and labeled as described in (A). Top row shows extent of lipoylation using ^{WT}LplA. Middle rows show extent of lipoylation with less-active mutants, ^{W37A}LplA and ^{W37V}LplA. Bottom row shows that when a dead mutant of LplA is used (^{K133R}LplA) there is no lipoylation observed. **(C)** HEK cells prepared identically to those described in (B) were analyzed by FACS. Each dot in the scatter plot corresponds to an individual cell. Lipoylation level is on x-axis, while enzyme expression is on the y-axis. Row designations correspond to rows in (B).

Model selection of LplAs displayed on HEK cell surface

The next step was to determine whether active ligases could be enriched from a heterogeneous population of cells using FACS. To do this, we performed one round of model selection on mixtures of ^{WT}LplA and ^{W37V}LplA HEK. For the model selections, transfected HEK cells were mixed in 1:1, 1:10, and 1:100 ratios of ^{WT}LplA: ^{W37V}LplA yeast and then labeled with lipoic acid. After fluorophore staining, HEK cells that exhibited a high ratio of labeling to enzyme expression were sorted out (Figure 4-6). The FACS plots demonstrate that the population of highly labeled HEK cells decreased with

increasing dilution of ^{WT}LplA. DNA was recovered from both the unsorted mixtures and the sorted cells for subsequent analysis of the enrichment factor of this selection strategy.

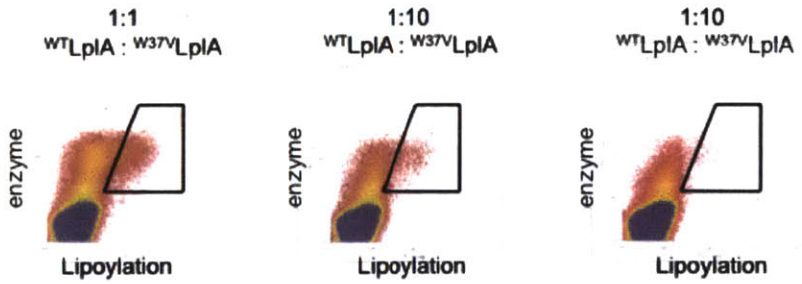


Figure 4-6. Model selection for enrichment of lipoylation activity. HEK cells expressing E2p-CFP-TM and either ^{WT}LplA-neurexin-1β or ^{W37V}LplA-neurexin-1β were labeled in cis as in Figure 4-5 and then mixed in ratios of 1:1, 1:10, or 1:100 ^{WT}LplA:^{W37V}LplA expressing cells. The cells were then sorted via FACS using the gate shown. This gate was designed to select out highly labeled HEK cells.

Two strategies to determine enrichment factor were tested. First we attempted to analyze the DNA directly using a PCR assay. We had previously modified the ^{WT}LplA-neurexin-1β DNA by adding 200 base pairs to the non-coding region of the plasmid. This modification allowed the use of identical primers to amplify the plasmid mixture and distinguish between the ^{WT}LplA and ^{W37V}LplA populations based on amplicon size. Unfortunately, it became clear that the PCR assay was introducing a bias, selectively amplifying the ^{W37V}LplA insert over the ^{WT}LplA insert (Figure 4-7A). This bias was present even when using known ratios of purified plasmid DNA. This approach was therefore abandoned in favor of direct sequencing analysis of the isolated DNA samples. By analyzing the sequencing chromatogram at the W37 residue, the ratio of ^{WT}LplA and ^{W37V}LplA can be determined accurately (Figure 4-7B). From this data, we calculated that this selection strategy gives 50-fold enrichment of ^{WT}LplA over the less-active W37V mutant.

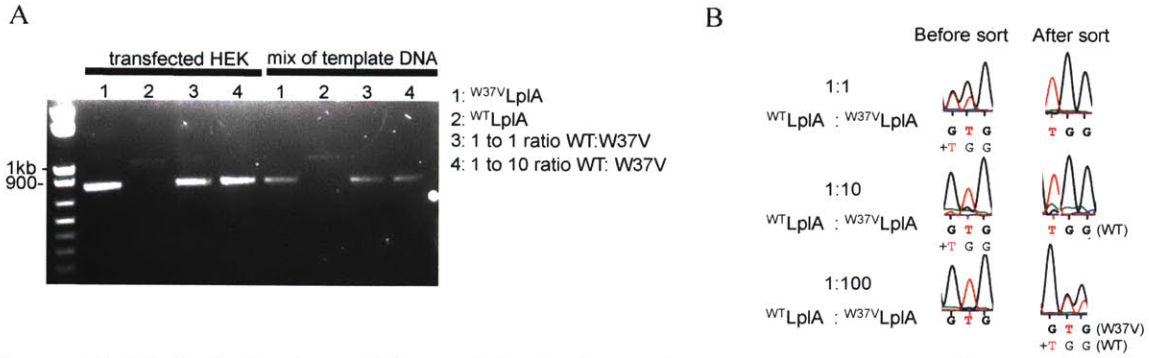


Figure 4-7. Methods for determining model selection enrichment. (A) PCR amplification assay for enrichment. ^{WT}LplA-neurexin-1b has an additional 200 base pairs added to the non-coding region of the DNA, so that it can be distinguished from ^{W37V}LplA-neurexin-1b using identical primers. The PCR reactions were performed and products run on a gel. When using either DNA isolated from HEK cells (Lanes 1-4, left), or simply template DNA mixed *in vitro* (Lanes 1-4, right), there is a bias in the PCR amplification. **(B)** Sequencing assay for enrichment. HEK cells were lysed after sorting and the plasmid DNA isolated and sent for sequencing. By analyzing the sequencing chromatograms at residue 37 of LplA, the ratio of ^{WT}LplA: ^{W37V}LplA can be determined. After comparison of DNA mixtures before and after sorting it was determined using this assay that one round of selection can produce ~50x enrichment.

These results demonstrate that the mammalian cell based selection strategy can give good enrichment using transiently transfected cells. Before we began actual library selections, however, two points must be addressed. First, since the FACS selections gate on substrate labeling and enzyme expression only, the amount of substrate needs to be consistent across all library members. For this, we created HEK cell lines stably expressing LAP2-CFP-TM and E2p-CFP-TM (data not shown). Second, the issue of library delivery needs to be addressed. We determined that we would use error-prone PCR to create a library of LplA mutants (see Chapter 5 for protocol details). However, as previously mentioned, one of the problems with using mammalian cell-based evolution platforms is the difficulty of achieving delivery of ≤1 library member per cell. At the time we were working on this project, the best option was to use viral delivery of library members[47]. Since then, a group has reported use of a FLIP-IN system which stably integrates plasmid DNA at a single locus using CHO cells stably expressing Flp recombinase[38]. This is a fantastic development, because stable integration at a single locus eliminates potential variations in expression levels of library members from gene silencing that are common when using random genomic integration.

Viral infection gives poor LplA surface expression in HEK cells

Since viral delivery was the most promising technique available at the time of these experiments, we proceeded to test viral transduction and expression of ^{WT}LplA-neurexin-1β. The commercial RetroMax viral delivery system was used to prepare viruses and expression plasmids. Using this system, we were able to achieve fantastic viral delivery and expression of a cytosolic ^{WT}LplA-GFP control plasmid (Figure 4-8, top). However, no surface expression was observed when using ^{WT}LplA-neurexin-1β (Figure 4-8, middle). Antibody staining after cell fixation demonstrated that viral delivery of the plasmid was successful and the protein was being expressed, but it was not trafficked properly to the cell surface under these conditions (Figure 4-8, bottom). Further optimization of the viral plasmid (such as adjusting the linker between ^{WT}LplA and the neurexin) did not recover the level of surface expression we observe when the plasmid is transiently transfected.

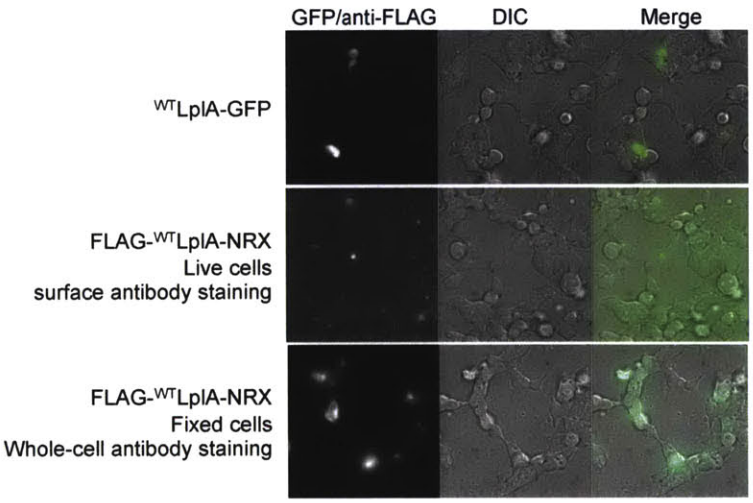


Figure 4-8. Viral infection of ^{WT}LplA-neurexin-1β gives poor surface expression. HEK cells were infected using packaged retrovirus prepared using the RetroMax kit (Imgenex). When infecting using ^{WT}LplA fused to cytosolic GFP (^{WT}LplA-GFP) expression is good and transduction efficiency is ~10% (top row). When infecting using ^{WT}LplA-neurexin-1β (NRX), there is no labeling when performing surface antibody staining for enzyme expression (FLAG, middle row). However, when the infected HEK cells are fixed and then stained for enzyme expression, signal is observed (bottom row), indicating good expression but poor surface delivery.

When this result was obtained, it had just been determined that HEK cell viability after FACS sorting was less than 20%. With viability this low, the HEK cells could not be propagated through rounds without drastically reducing library diversity due to cell death. Instead the cells would need to be lysed after sorting followed by DNA recovery,

new virus preparation, and infection of fresh HEK cells for the next round of selections. As the technological difficulties associated this approach became increasingly apparent, we abandoned mammalian cell surface display in favor of yeast-display. While mammalian display is in its infancy with no examples of evolved catalytic activity, there are several examples of using yeast-display to evolve enzymatic activity. The literature precedence for enzyme evolution[25,46], combined with the improved recovery of yeast cells from FACS were the most important factors in our decision to switch to a yeast-display platform. The yeast-display evolution of LplA will be discussed in Chapters 5-7 of this thesis.

Conclusions

This chapter summarizes the results of the rational engineering of LplA. While we were able to determine that LplA activity in the secretory pathway and cell surface is affected by enzyme oxidation state, there are additional pathways of enzyme inactivation that we were not able to identify directly. These factors could be general enzyme stability or enzyme misfolding. We also demonstrated that truncation of LplA improves enzyme stability and LplA-NTD retains full lipoyltransferase activity and partial adenylation activity. Since rational engineering of LplA was largely unsuccessful in restoring activity in the secretory pathway, and the source of enzyme inactivation is not entirely clear, we determined that directed evolution of the enzyme might produce better results.

We decided that a semi-*in vitro* directed evolution method be ideal for the selection of LplA variants. It would simultaneously allow the use of powerful *in vitro* library generation techniques while permitting selection for activity along the secretory pathway of living cells. We identified two routes of enzyme evolution, one focused on evolving LplA-NTD for protein labeling and one focused on evolving full-length LplA for improved activity in the secretory pathway.

We first tested a mammalian cell surface display evolution platform on full-length LplA. While this platform produced ~50x enrichment of active ligases in a model selection, there were two significant problems that led to the abandonment of this approach. First, the recovery of mammalian cells after sorting was very poor. As a result, preventing the loss of library diversity due to cell death would require that the library

DNA be isolated between each round of selection, which would complicate this strategy. Second, the viral delivery of our target gene produced extremely poor surface expression and infection was difficult to achieve reproducibly over several batches of virus. This issue, combined with the need to perform viral delivery at each round to deal with poor cell recovery, made continuing with this directed evolution platform untenable. We therefore turned to yeast-display evolution which is a much more established method of directed evolution. The remaining chapters of this thesis focus on the efforts to use yeast-display evolution to improve various properties of LplA. The results of the yeast-display evolution of LplA-NTD are summarized in Chapter 5 and the results of the yeast-display evolution of full-length LplA are discussed in Chapter 6 and Chapter 7.

Experimental

Cis lipoylation on mammalian cell surface (Figure 4-5, Table 4-1)

HEK cells were transfected with 600 ng of E2p-CFP-TM and 600 ng of ^{WT}LplA-NRX or the indicated LplA mutant using Lipofectamine 2000 (Invitrogen) per manufacturer's protocol. 12 hours post-transfection cells were labeled with 500 μM lipoic acid, 5 mM $Mg(OAc)_2$, and 1 mM ATP for 6 hours at 30C. For all samples, cells were washed 2 x 10 minutes with DMEM and then lipoylation was detected via live-cell antibody staining. Live-cell immunofluorescence staining was performed with 4 μg/ml anti-lipoic acid antibody (CalBiochem) and 4 μg/ml anti-FLAG antibody (Stratagene) in 1% BSA in DPBS for 15 minutes. After three quick washes, cells were incubated with 4 μg/ml goat anti-rabbit antibody conjugated to AlexaFluor 488 and 4 μg/ml goat anti-mouse AlexaFluor 568 (both Invitrogen) in 1% BSA in DPBS for 15 minutes, then washed and imaged.

For FACS analysis, HEK cells plated into T25 flasks were transfected with 4 μg of E2p-CFP-TM and 4 μg of ^{WT}LplA-NRX or the indicated LplA mutant. Labeling was performed identically to above and then the adherent HEK cells were mechanically lifted from the flasks and resuspended in DMEM at 1 x 10^6 cells/ml for FACS analysis.

Model selection for lipoylation on HEK cell surface (Figure 4-6, 4-7)

For FACS analysis, HEK cells plated into T25 flasks were transfected with 4 μg of LAP2-CFP-TM and 4 μg of ^{WT}LplA-NRX or ^{W37A}LplA-NRX using Lipofectamine 2000 (Invitrogen) per manufacturer's protocol. Labeling was performed as described above in **cis lipoylation on mammalian cell surface** and then the adherent HEK cells were mechanically lifted from the flasks and resuspended in DMEM at 1 x 10^6 cells/ml. The two pools of HEK cells were then combined in 1:1, 1:10, or 1:100 ratios of ^{WT}LplA: ^{W37V}LplA. Approximately 250 μL of each mixed sample was removed and saved for subsequent DNA isolation and sequencing, the remaining sample was sorted via FACS and using the indicated sort gates.

Pre-mixed HEK samples and sorted HEK cells were pelleted, lysed and the plasmid DNA isolated. Isolated DNA was then sequenced and sequencing chromatograms were analyzed to determine enrichment factors for the model selection.

Viral transduction of LplA-NRX (Figure 4-8)

Viral transduction was performed according to the RetroMax Retroviral system (Imgenex). HEK cells were transfected with the viral expression gene: pCLNCX ^{WT}LplA-NRX or pCLNCX ^{WT}LplA-GFP and the viral packaging gene: pCLAmpho. 2 hours post-transfection, change media. 24 hours after media change, collect virus-containing supernatant from the cells. Virus was filter-sterilized and stored in aliquots at -80C. Virus was used in 1:10 or 1:20 titres in conditioned media with 8 μg/ml Polybrene and incubated with cells for 24 hours.

References

1. Uttamapinant, C., *et al.* A fluorophore ligase for site-specific protein labeling inside living cells. *Proc Natl Acad Sci U S A*, *107*, 10914-9.

2. Green, D. E.; Morris, T. W.; Green, J.; Cronan, J. E., Jr.; Guest, J. R. Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem J* **1995**, *309 (Pt 3)*, 853-62.

3. 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*, 33645-51.

4. McManus, E.; Luisi, B. F.; Perham, R. N. 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*, 625-37.

5. 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*, 38081-9.

6. Ma, Q., *et al.* The Mycobacterium tuberculosis LipB enzyme functions as a cysteine/lysine dyad acyltransferase. *Proc Natl Acad Sci U S A* **2006**, *103*, 8662-7.

7. Kim do, J., *et al.* Structural basis of octanoic acid recognition by lipoate-protein ligase B. *Proteins* **2008**, *70*, 1620-5.

8. Fujiwara, K., *et al.* Global conformational change associated with the two-step reaction catalyzed by Escherichia coli lipoate-protein ligase A. *J Biol Chem* **2010**, *285*, 9971-80.

9. Blagodatski, A.; Katanaev, V. L. Technologies of directed protein evolution in vivo. *Cell Mol Life Sci* **2010**, *68*, 1207-14.

10. Labrou, N. E. Random mutagenesis methods for in vitro directed enzyme evolution. *Curr Protein Pept Sci* **2010**, *11*, 91-100.

11. Greener, A.; Callahan, M.; Jerpseth, B. An efficient random mutagenesis technique using an E. coli mutator strain. *Mol Biotechnol* **1997**, *7*, 189-95.

12. Camps, M.; Naukkarinen, J.; Johnson, B. P.; Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc Natl Acad Sci U S A* **2003**, *100*, 9727-32.

13. Cobb, R. E.; Sun, N.; Zhao, H. Directed evolution as a powerful synthetic biology tool. *Methods*.

14. Zhang, K.; Li, H.; Cho, K. M.; Liao, J. C. Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc Natl Acad Sci U S A*, *107*, 6234-9.

15. Atsumi, S.; Liao, J. C. Directed evolution of Methanococcus jannaschii citramalate synthase for biosynthesis of 1-propanol and 1-butanol by Escherichia coli. *Appl Environ Microbiol* **2008**, *74*, 7802-8.

16. Campbell, R. E., *et al.* A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002**, *99*, 7877-82.

17. Wang, L.; Jackson, W. C.; Steinbach, P. A.; Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc Natl Acad Sci U S A* **2004**, *101*, 16745-9.

18. Arakawa, H.; Hauschild, J.; Buerstedde, J. M. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* **2002**, *295*, 1301-6.

19. Buerstedde, J. M.; Takeda, S. Increased ratio of targeted to random integration after transfection of chicken B cell lines. *Cell* **1991**, *67*, 179-88.

20. Seo, H., *et al.* Rapid generation of specific antibodies by enhanced homologous recombination. *Nat Biotechnol* **2005**, *23*, 731-5.

21. Seo, H.; Yamada, T.; Hashimoto, S.; Lin, W.; Ohta, K. Modulation of immunoglobulin gene conversion in chicken DT40 by enhancing histone acetylation, and its application to antibody engineering. *Biotechnol Genet Eng Rev* **2007**, *24*, 179-93.

22. Goldsmith, M.; Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol*.

23. Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol* **1996**, *255*, 589-603.

24. Leemhuis, H.; Stein, V.; Griffiths, A. D.; Hollfelder, F. New genotype-phenotype linkages for directed evolution of functional proteins. *Curr Opin Struct Biol* **2005**, *15*, 472-8.

25. Agresti, J. J., *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* **2010**, *107*, 4004-9.

26. Colby, D. W., *et al.* Engineering antibody affinity by yeast surface display. *Methods Enzymol* **2004**, *388*, 348-58.

27. Evran, S.; Telefoncu, A.; Sterner, R. Directed evolution of (betaalpha)8-barrel enzymes: establishing phosphoribosylanthranilate isomerisation activity on the scaffold of the tryptophan synthase alpha-subunit. *Protein Eng Des Sel* **2012**.

28. Bratkovic, T. Progress in phage display: evolution of the technique and its applications. *Cell Mol Life Sci* **2009**.

29. Fernandez-Gacio, A.; Uguen, M.; Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol* **2003**, *21*, 408-14.

30. Wu, G., *et al.* Bioscreening of phage display antibody library and expression of a humanized single-chain variable fragment antibody against human connective tissue growth factor (CTGF/CCN2). *Biotechnol Appl Biochem*, *56*, 95-102.

31. Kosako, H., *et al.* Isolation and characterization of neutralizing single-chain antibodies against Xenopus mitogen-activated protein kinase kinase from phage display libraries. *Biochemistry* **1996**, *35*, 13212-21.

32. Chen, I.; Choi, Y. A.; Ting, A. Y. 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*, 6619-25.

33. Marks, K. M.; Rosinov, M.; Nolan, G. P. In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* **2004**, *11*, 347-56.

34. Lower, B. H., *et al.* In vitro evolution of a peptide with a hematite binding motif that may constitute a natural metal-oxide binding archetype. *Environ Sci Technol* **2008**, *42*, 3821-7.

35. Janda, K. D., *et al.* Chemical selection for catalysis in combinatorial antibody libraries. *Science* **1997**, *275*, 945-8.

36. Demartis, S., *et al.* A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. *J Mol Biol* **1999**, *286*, 617-33.

37. Cesaro-Tadic, S., *et al.* Turnover-based in vitro selection and evolution of biocatalysts from a fully synthetic antibody library. *Nat Biotechnol* **2003**, *21*, 679-85.

38. Zhou, C.; Jacobsen, F. W.; Cai, L.; Chen, Q.; Shen, W. D. Development of a novel mammalian cell surface antibody display platform. *MAbs*, *2*, 508-18.

39. Olsen, M. J., *et al.* Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* **2000**, *18*, 1071-4.

40. Gai, S. A.; Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Curr Opin Struct Biol* **2007**, *17*, 467-73.

41. Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **1997**, *15*, 553-7.

42. Akamatsu, Y.; Pakabunto, K.; Xu, Z.; Zhang, Y.; Tsurushita, N. Whole IgG surface display on mammalian cells: Application to isolation of neutralizing chicken monoclonal anti-IL-12 antibodies. *J Immunol Methods* **2007**, *327*, 40-52.

43. Urban, J. H., *et al.* Selection of functional human antibodies from retroviral display libraries. *Nucleic Acids Res* **2005**, *33*, e35.

44. Wolkowicz, R.; Jager, G. C.; Nolan, G. P. A random peptide library fused to CCR5 for selection of mimetopes expressed on the mammalian cell surface via retroviral vectors. *J Biol Chem* **2005**, *280*, 15195-201.

45. Varadarajan, N.; Rodriguez, S.; Hwang, B. Y.; Georgiou, G.; Iverson, B. L. Highly active and selective endopeptidases with programmed substrate specificities. *Nat Chem Biol* **2008**, *4*, 290-4.

46. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011**, *108*, 11399-404.

47. Li, F., *et al.* Human anti-EGFL7 recombinant full-length antibodies selected from a mammalian cell-based antibody display library. *Mol Cell Biochem.*

48. Lu, W. C.; Ellington, A. D. In vitro selection of proteins via emulsion compartments. *Methods.*

49. Sawata, S. Y.; Taira, K. Modified peptide selection in vitro by introduction of a protein-RNA interaction. *Protein Eng* **2003**, *16*, 1115-24.

50. Lamla, T.; Erdmann, V. A. Searching sequence space for high-affinity binding peptides using ribosome display. *J Mol Biol* **2003**, *329*, 381-8.

51. Zahnd, C.; Amstutz, P.; Pluckthun, A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. *Nat Methods* **2007**, *4*, 269-79.

52. Hanes, J.; Schaffitzel, C.; Knappik, A.; Pluckthun, A. Picomolar affinity antibodies from a fully synthetic naive library selected and evolved by ribosome display. *Nat Biotechnol* **2000**, *18*, 1287-92.

53. Zahnd, C., *et al.* Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem* **2004**, *279*, 18870-7.

54. Matsuura, T.; Pluckthun, A. Selection based on the folding properties of proteins with ribosome display. *FEBS Lett* **2003**, *539*, 24-8.

55. Roberts, R. W.; Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A* **1997**, *94*, 12297-302.

56. Miller, O. J., *et al.* Directed evolution by in vitro compartmentalization. *Nat Methods* **2006**, *3*, 561-70.

57. Horisawa, K., *et al.* In vitro selection of Jun-associated proteins using mRNA display. *Nucleic Acids Res* **2004**, *32*, e169.

58. Chen, Y.; Mandic, J.; Varani, G. Cell-free selection of RNA-binding proteins using in vitro compartmentalization. *Nucleic Acids Res* **2008**, *36*, e128.

59. Fen, C. X.; Coomber, D. W.; Lane, D. P.; Ghadessy, F. J. Directed evolution of p53 variants with altered DNA-binding specificities by in vitro compartmentalization. *J Mol Biol* **2007**, *371*, 1238-48.

60. Cohen, H. M.; Tawfik, D. S.; Griffiths, A. D. Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Eng Des Sel* **2004**, *17*, 3-11.

61. Doi, N.; Kumadaki, S.; Oishi, Y.; Matsumura, N.; Yanagawa, H. In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Res* **2004**, *32*, e95.

62. Griffiths, A. D.; Tawfik, D. S. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J* **2003**, *22*, 24-35.

63. Mastrobattista, E., *et al.* High-throughput screening of enzyme libraries: in vitro evolution of a beta-galactosidase by fluorescence-activated sorting of double emulsions. *Chem Biol* **2005**, *12*, 1291-300.

64. Wilson, D. S.; Keefe, A. D.; Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc Natl Acad Sci U S A* **2001**, *98*, 3750-5.

65. Seelig, B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat Protoc*, *6*, 540-52.

Chapter 5. Yeast-display evolution of a truncated LplA for site-specific protein labeling

The work discussed in this chapter is unpublished. The first model selections and first-generation selections for LplA-NTD variants with both adenylation and lipoyltransferase activity were performed by Dr. Sujiet Puthenveetil. Philip Zegelbone assisted with the *in vitro* characterization of ^{WT}LplA-NTD and ^{T57I}LplA-NTD.

Introduction

We describe in this chapter the efforts to use yeast-display to evolve the N-terminal domain (NTD) of *E. coli* LplA. The first goal of this evolution project was to improve the ability of LplA-NTD to catalyze the adenylation step of the natural reaction. The second and more ambitious goal is to evolve LplA-NTD variants that can catalyze the adenylation and ligation of unnatural probes. The development of such an enzyme would have two far-reaching benefits for the PRIME methodologies. First, reducing the size of the ligase by almost 30% (~36kDa to ~26kDa) would make LplA labeling more attractive for applications such as ID-PRIME. When performing intracellular PRIME labeling, the large size of LplA is not a factor. However, enzyme size is a significant issue in ID-PRIME. Because the enzyme is fused to a protein of interest in ID-PRIME applications, fusions to full-length LplA may interfere with biology as is observed with GFP and other protein-based labeling methods. Second, since yeast-display will be used to evolve the truncated LplA-NTDs, any evolved variants selected for increased catalytic activity would be concomitantly selected for improved activity when expressed on the cell surface. Therefore, the ligases would also address the original overarching goal of these LplA-based selections: produce a ligase with better activity in oxidizing compartments of the cell.

This chapter is divided into two sections. In the first section, we present the directed evolution of an LplA-NTD variant with better lipoylation activity than ^{WT}LplA. We first compare the lipoylation activity of the original ^{WT}LplA-NTD and full-length ^{WT}LplA on the mammalian cell surface and *in vitro*. Then we briefly describe the general yeast-display platform and the selection strategy that produced LplA-NTD variants capable of catalyzing both the adenylation and transfer steps of the lipoylation reaction. The lipoylation activities of the resulting ligases are characterized on mammalian cells and *in vitro* and the contributions of the consensus mutations are investigated. These consensus mutations informed and focused the second generation of LplA-NTD selections, which we describe in the second section. The second-generation libraries and selections were directed at the evolution of LplA-NTD variants with unnatural probe ligation activity.

Results

Evolution of LplA-NTD variants capable of catalyzing both steps of lipoylation reaction

Truncation of LplA and cis lipoylation tests

As introduced in Chapter 4, we had discovered that LplA retained some catalytic activity when the C-terminal domain was truncated. Based on the apo *E. coli* structure, we had initially performed truncation of the enzyme at position 249 (Figure 5-1A)[1]. This enzyme had full lipoyltransferase activity but drastically reduced adenylation activity. When we considered the structure of the activated lipoyl-AMP bound form of LplA, we noticed that a previously disordered adenylation loop becomes ordered and the resulting beta sheet interacts with the first beta-sheet of the C-terminal domain, corresponding to residues 250-255 (Figure 5-1, right)[2]. We therefore also tested a truncation of the enzyme at Glu256. We hoped that this enzyme might have better adenylation activity compared to the Pro249 truncation. We found that LplA truncated at Glu256 had adenylation activity that was approximately 70% higher than the Pro249 truncation in an in vitro assay (data not shown). Therefore, it was determined that the Glu256 truncation (hereafter referred to as LplA-NTD) would be the basis for future experiments and evolution.

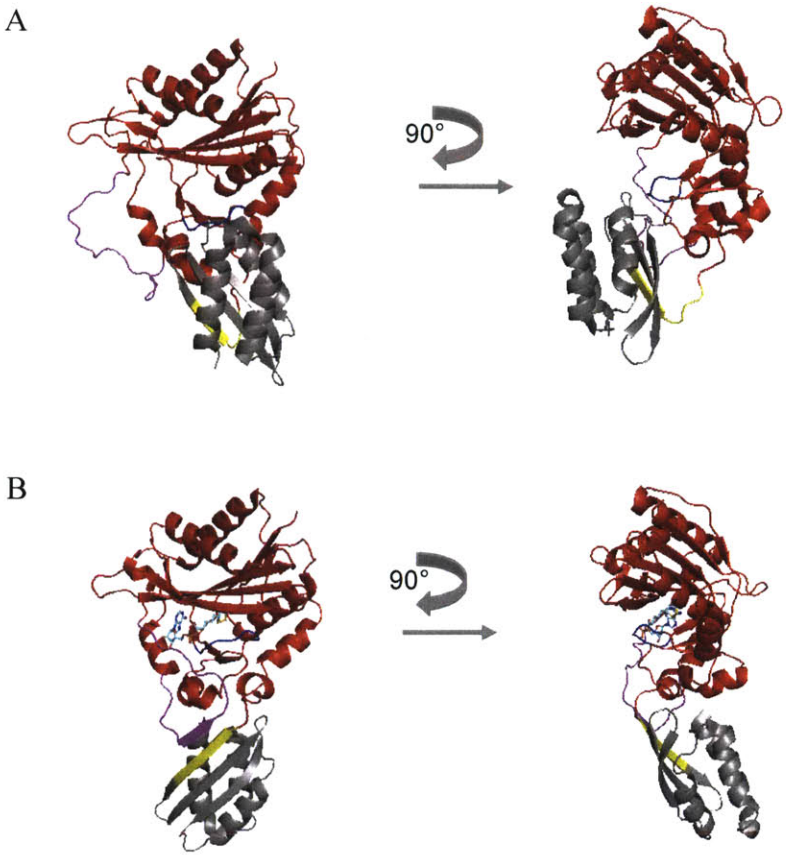


Figure 5-1. Structural comparison of the two truncated LplA variants. (A) Crystal structure from apo *E. coli* LplA (PDB 1X2G). Adenylate-binding loop and lipoate binding loop are in pink and blue, respectively. C-terminal domain is colored grey. Residues 1-249, corresponding to Pro249 truncate, are colored in red. Residues 250-255 are colored in yellow and represent the extension to make Glu256 truncate. **(B)** Crystal structure from Lipoyl-AMP bound *E. coli* LplA (PDB 3A7R). All colors are maintained as in (A).

We next wanted to test the lipoylation activity of LplA-NTD in a mammalian cell-surface assay, to do this we made a fusion of LplA-NTD to neurexin-1β. This is a fusion that, as previously described in Chapter 4, gives good cell-surface expression of the enzyme. We then expressed this fusion in HEK cells along with a surface LAP2 fusion (LAP2-CFP-TM). The cells were incubated with lipoic acid and ATP and the activity of the enzyme was determined by monitoring cis lipoylation of the LAP2-CFP-TM. Figure 5-2 shows that when lipoyl-AMP is added to the cells, the cis lipoylation of LAP2 by LplA-NTD is equivalent to that observed with full-length ^{WT}LplA. However, when the cells are incubated with lipoic acid and ATP, the cis lipoylation of LAP2 by ^{WT}LplA-NTD is barely detectable, while full-length ^{WT}LplA gives very strong cis signal. These results suggest that the lipoyltransferase activity of ^{WT}LplA-NTD is unaffected by

the truncation of the C-terminal domain, but that the adenylation activity of ^{WT}LplA-NTD is markedly lower than that of ^{WT}LplA.

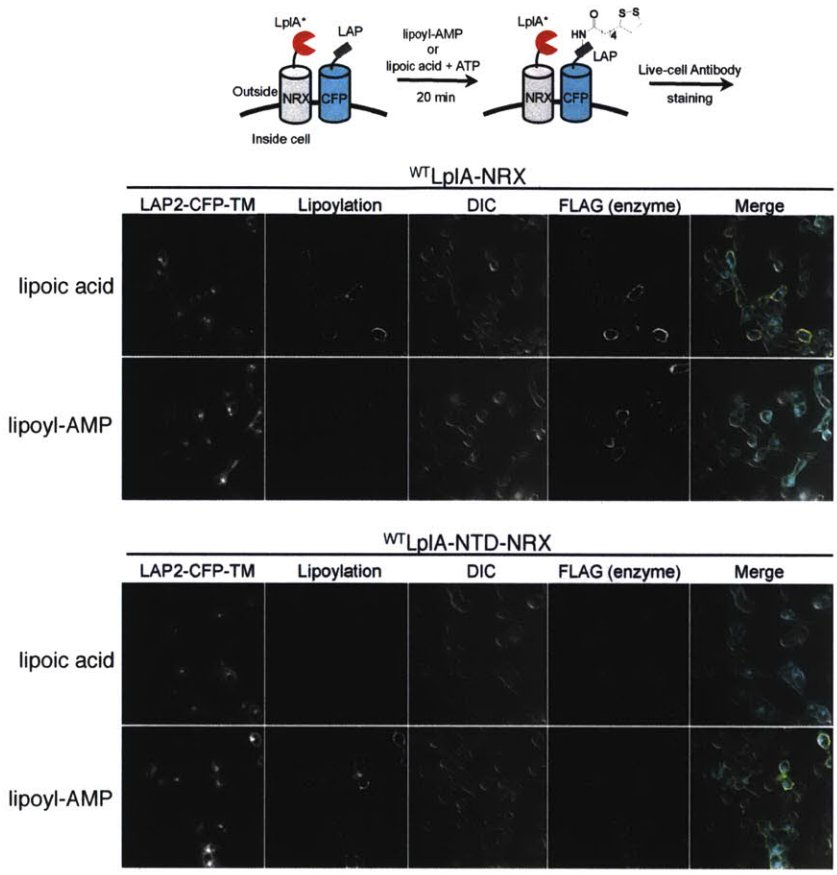


Figure 5-2. Comparison of adenylation and lipoyltransferase activity using LplA-NTD and full length LplA. Scheme shows cis lipoylation assay. The indicated enzyme and LAP2-CFP-TM are co-expressed on the surface of HEK cells. Transfected cells are incubated with lipoyl-AMP or lipoic acid and ATP for 1 hour. Cis lipoylation is detected by antibody staining and cells are imaged live.

Evolution of LplA-NTD variants capable of performing adenylation and ligation of lipoic acid

Yeast-display evolution schemes generally express the protein libraries fused to a yeast mating protein, Aga2p[3]. Aga2p is attached to the yeast cell surface via a disulfide interaction with the Aga1p mating protein. This scaffold produces consistently high expression of target libraries, reducing variability in selections[4]. In order to use yeast-display to evolve improved LplA activity, we need to also express the LAP2 peptide on the surface of the yeast (Figure 5-3). Chen and colleagues recently introduced a method for yeast-display evolution where enzyme and substrate are co-displayed on the yeast cell surface[5]. We proposed using a similar technique where the LAP2 peptide is constitutively

expressed on the yeast cell surface as a fusion to Aga1p.The enzyme library can then be induced in a subset of these cells as a fusion to Aga2p.

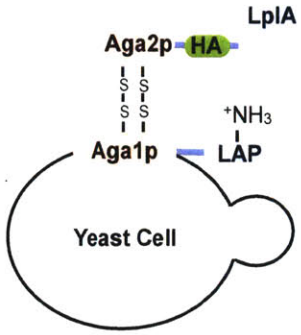


Figure 5-3. Yeast-display platform. Yeast cells constitutively express LAP2 as a fusion to the mating protein Aga1p. A subset of the cells inducibly express LplA as a fusion to Aga2p. Enzyme expression level can be monitored by HA epitope staining. LplA can catalyze the enzymatic ligation of probes onto the LAP2.

The general scheme for yeast-display evolution can be found in Figure 5-4A. Sujiet Puthenveetil made a first-generation library of LplA-NTD variants using error-prone PCR with dNTP analogues and Taq polymerase[6]. The resulting library was transformed into yeast cells constitutively expressing LAP2. The enzyme library is induced, and each yeast cell expresses a unique library member on the surface. The yeast pool is then labeled using lipoic acid and ATP and the lipoylation level of LAP2 can be detected using antibody staining (Figure 5-4B). Active ligases produced highly labeled yeast cells that were sorted out by FACS and then propagated to the next round. This scheme is then repeated iteratively with more stringent rounds of labeling and selection.

Sujiet performed the selection as outlined above for lipoylation activity, by the fourth round the best clone was identified as ^{T57I}LplA-NTD. To verify that the increase in labeling signal obtained with ^{T57I}LplA on the yeast cell surface is not context dependent, we tested the lipoylation activity of these clones on the mammalian cell surface in a cis labeling assay using lipoic acid and ATP (Figure 5-5). We compared this activity to ^{F147L}LplA-NTD and ^{WT}LplA-NTD as well as full-length ^{WT}LplA. The ^{F147L}LplA-NTD mutant originated from a first-generation LplA-NTD library that was subjected to selection for the ability to ligate an azidoalkanoic probe, although it didn't possess this activity on the mammalian cell surface (data not shown).

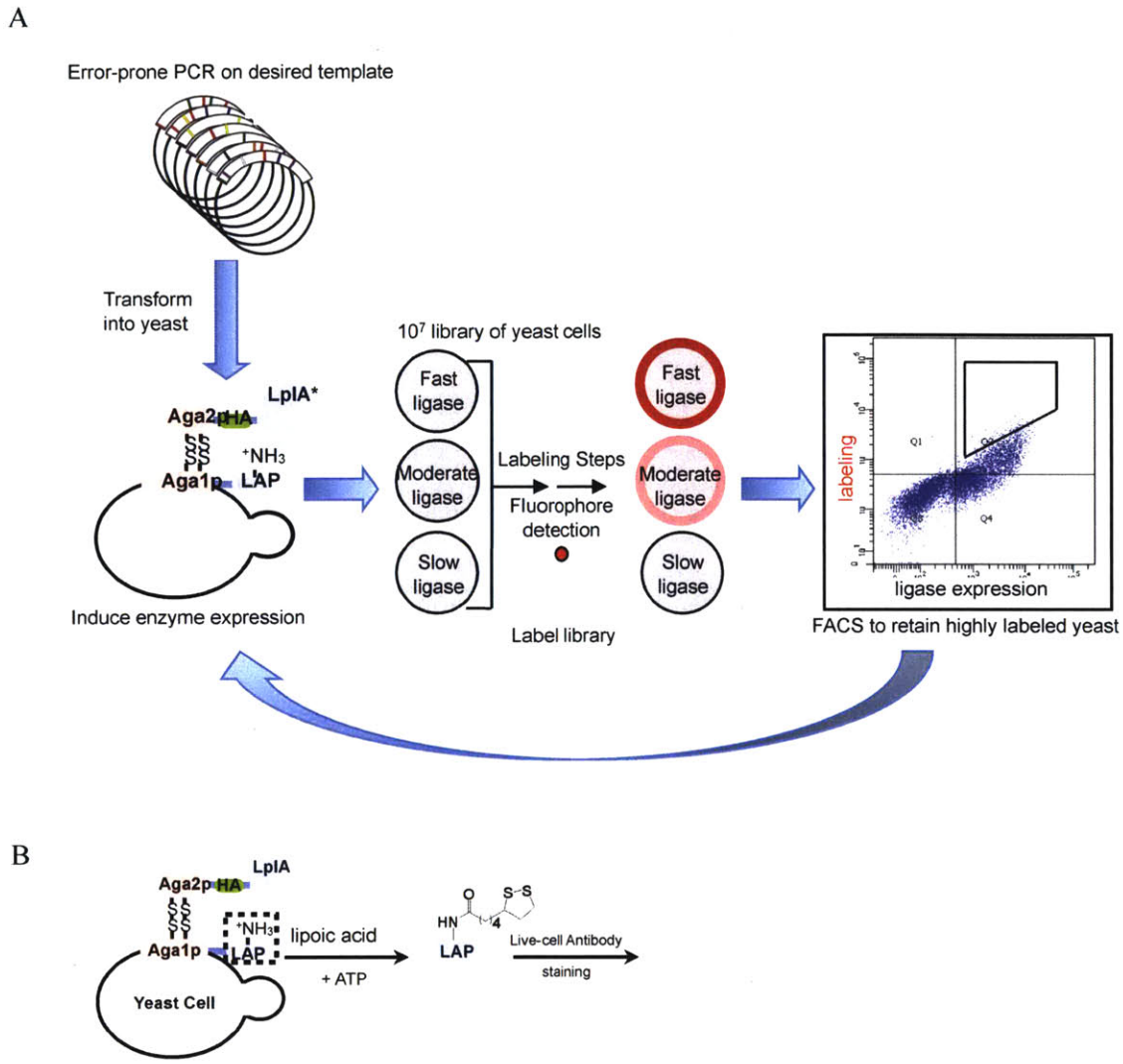


Figure 5-4. Yeast display selection strategy. (A) A library of LplA mutants is made using error-prone PCR. The library is transformed into yeast cells constitutively expressing LAP2-Aga1p. Each yeast cell is expressing a unique library member. The library of yeast cells is labeled and then highly labeled yeast cells are selected out via fluorescence activated cell sorting (FACS). This process is then repeated in increasingly stringent iterations of enzyme induction, labeling, and sorting. **(B)** Yeast cells are incubated with lipoic acid and ATP, allowing enzymatic lipoylation of the LAP2 by the LplA. Lipoylation is detected by live-cell antibody staining on the yeast cell surface. Enzyme expression level is also monitored by HA staining.

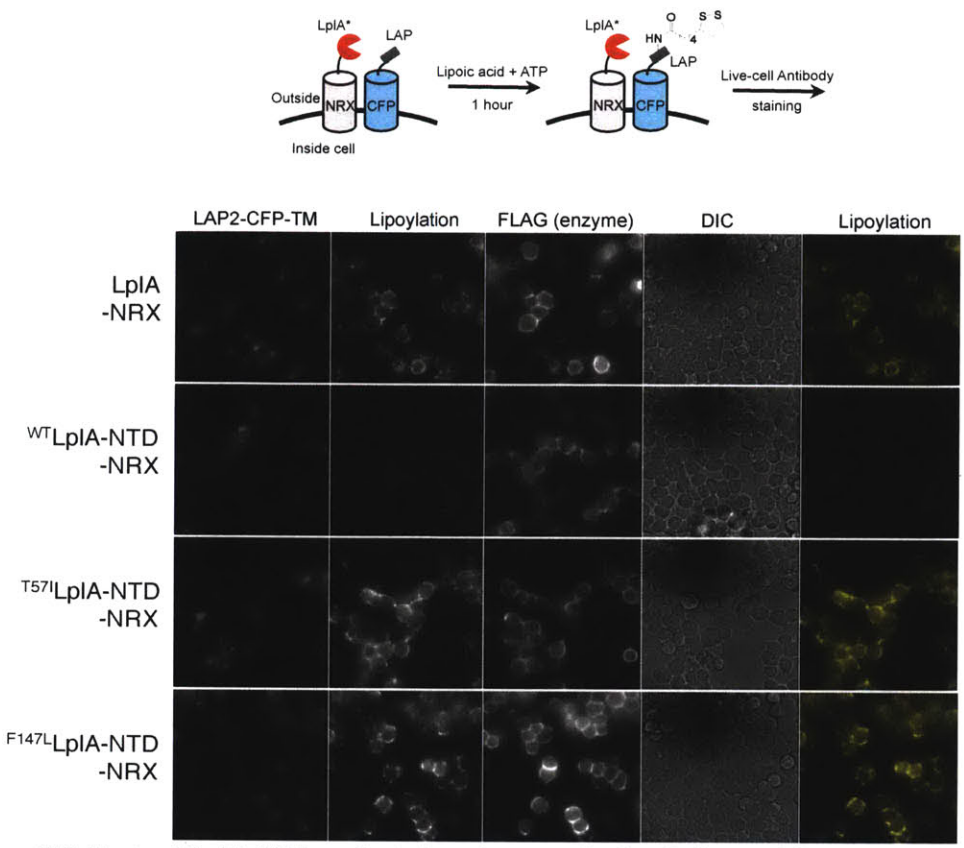


Figure 5-5. Evolved LplA-NTD variants have improved adenylation activity. HEK cells co-expressing LAP2-CFP-TM and the indicated enzyme fused to neurexin-1β (NRX) were incubated with lipoic acid and ATP. Cis lipoylation of the LAP2 was detected by antibody staining. High signal indicates both adenylation and lipoyltransferase activity.

From the cis lipoylation assay, it is clear that both ^{T57I}LplA-NTD and ^{F147L}LplA-NTD have improved adenylation activity compared to ^{WT}LplA-NTD. More importantly, the lipoylation activity of the evolved LplA-NTD variants rivals that observed with full-length ^{WT}LplA under these conditions. While the lipoylation activity was increased, no cis labeling was observed on the mammalian cell surface when using the azidoalkanoic probes (data not shown). Given this result, we were concerned that perhaps the original error-prone library based on the ^{WT}LplA-NTD template did not produce many active clones. We therefore decided to perform a second round of directed evolution with two fundamental changes. First, we biased the library towards improved adenylation activity by using the lipoylation-improving mutations in the template for error-prone PCR. Second, instead of selecting for lipoylation activity, the second-generation evolution scheme would select for unnatural probe ligation.

Evolution of LplA-NTD variants capable of performing adenylation and ligation of an unnatural probe

Selection scheme for LplA-NTD directed evolution for bromoalkanoic probe ligation

The first step in the second-generation evolution was to determine the target probe for the selection platform. The first option is to select for coumarin fluorophore ligation. However, this was not feasible for three reasons. First, coumarin is very sticky on the yeast cell and the resulting background is high. Second, the cell sorters we have access to do not have optimized lasers and filter sets for coumarin fluorescence, and the suboptimal settings available did not produce very high signal, even when yeast were labeled exogenously with purified enzyme (data not shown). Finally, while a direct fluorophore ligation is ideal for intracellular applications, a two-step labeling scheme would be best for enzyme evolution. A two-step scheme would evolve an LplA variant to ligate the functional group handle and then detect that label with a variety of fluorophore or probe conjugates (Figure 5-6).

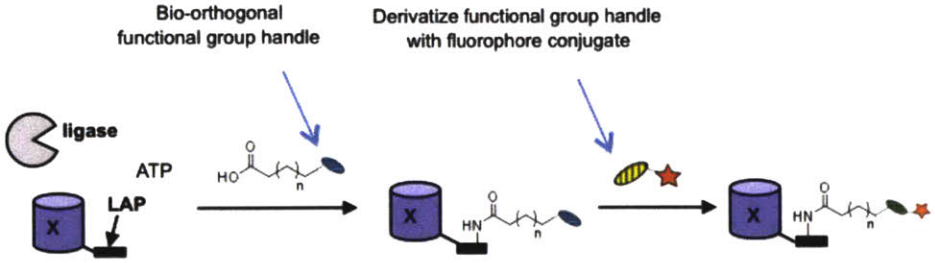


Figure 5-6. Two-step labeling strategy using LplA. In the first step, LplA is used to perform enzymatic ligation of a functional group handle or affinity handle. In the second step, the functional group or affinity handle is detected by chemoselective derivatization or affinity probe.

There was already precedence in the lab for performing two-step labeling on the yeast cell surface. For the evolution of the LAP2 peptide, Sujiet Puthenveetil had successfully performed labeling and enrichment using a two-step labeling method, and we proposed using a similar selection scheme in the second-generation selection. In the proposed scheme we use LplA to ligate an 11-bromoundecanoic acid probe (11-Br) onto the LAP2 peptide. The ligated probe is detected using a two-step method with HaloTag[7] conjugated to biotin followed by streptavidin fluorophore conjugate (Figure 5-7A). This scheme gives an effective fluorescent readout of the enzymatic activity of LplA-NTD library members.

Once we had determined a selection scheme, the next step was to generate a library of LplA-NTD mutants. The library was generated using error-prone PCR with dNTP analogues. Both 8-oxo-deoxyguanosine-5'-triphosphate (8-oxo-GTP) and 2'-deoxy-P-nucleoside-5'-triphosphate (dPTP) were used to induce point mutations in the template DNA[6,8]. The 8-oxo-GTP analogue induces A to C and T to G mutations in the DNA with T to G mutations being slightly over-represented (1:1.5 ratios). The dPTP analogue induces A to G, T to C, G to A, and C to T mutations, with A-G mutations being very over-represented (5:4:1:1 ratios). By controlling the concentrations of dNTP analogues added to the PCR and by controlling the number of PCR cycles performed in the presence of the analogues, the mutation rate of the library can be exquisitely controlled. The libraries were made using a 1:1:1 ratio of ^{WT}LplA:^{T57I}LplA-NTD:^{F147L}LplA-NTD. Using this mix of DNA templates would incorporate additional diversity and we hoped bias the library towards incorporation of mutations that increased the lipoylation activity of ^{WT}LplA-NTD would improve the initial activity of the library members.

Because we were not sure whether a highly mutated library would be beneficial or deleterious to LplA-NTD activity, three libraries were made with various mutation rates (see Experimental). Error rates ranged from just one amino acid mutation per library member to eight amino acid mutations per library member. The libraries were transformed into yeast using electroporation[9] and library sizes ranged from 6.0×10^6 to 3.0×10^7 members (see Experimental). The resulting yeast libraries were pooled and selections were performed on the pooled library, which had a combined diversity of 5×10^7.

The pooled library was subjected to four rounds of selection for ligation of the 11-bromoundecanoic acid (11-Br); yeast pools were amplified between rounds but no further library diversification was performed. Evolutionary pressure for improved catalytic activity was introduced by decreasing the labeling times (from 5 hours to 1 hour) and by collecting a smaller fraction of cells in each successive round (Figure 5-7B). By round 4, the pooled library of 5×10^7 unique members had been reduced to just four consensus members (Table 5-1). All consensus clones contained the T57I mutation, indicating that

this mutation increases labeling activity when combined with a variety of additional mutations. We hypothesize that T57I may be improving LplA-NTD surface stabilization. T57I is located at the end of a short alpha helix that is more solvent exposed in the apo form of LplA compared to the lipoyl-AMP bound form of the protein (Figure 5-8). We wonder whether the mutation of a polar threonine residue to a more hydrophobic isoleucine residue may help bury the loop and stabilize the N-terminal domain when it is expressed alone.

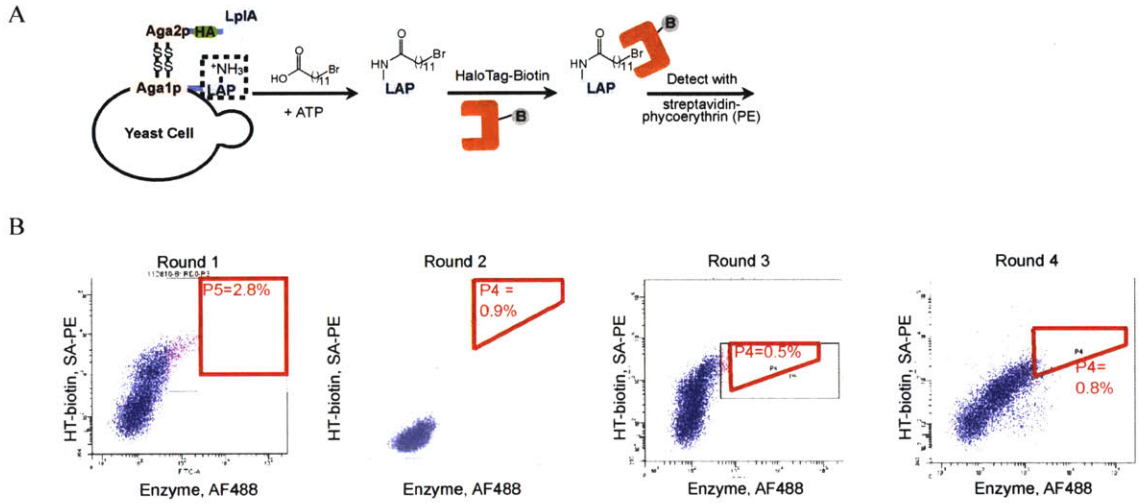


Figure 5-7. Selection for 11-Br probe ligation using LplA-NTD library. (A) Selection strategy. The prepared yeast library is incubated with 11-Br probe and ATP, active enzymes are able to catalyze the covalent attachment of the 11-Br probe to the LAP2 peptide. The ligated 11-Br can then be detected using HaloTag-biotin conjugate followed by streptavidin-PE. This gives an effective fluorescent readout of enzymatic activity. **(B)** 11-Br activity selection on LplA-NTD library. Sort windows from increasingly stringent labeling and selection rounds. All rounds used 500 μM 11-Br probe. Round 1- 5 hour incubation with 11-Br and ATP. Round 2- 2.5 hour incubation with 11-Br and ATP. Round 3-1.5 hour incubation with 11-Br and ATP. Round 4- 1 hour incubation with 11-Br and ATP.

	Frequency in R4
$^{T57I/Y139H/F147L}$LplA-NTD	10%
$^{T57I/A76S}$LplA-NTD	30%
$^{T57I/S159G}$LplA-NTD	40%
$^{T57I/Q247R}$LplA-NTD	20%

Table 5-1. Consensus clones from round 4 of 11-Br selection on LplA-NTD library. 20 clones from round 4 were randomly selected and sent for sequencing. The table depicts the relative frequencies of each consensus clone.

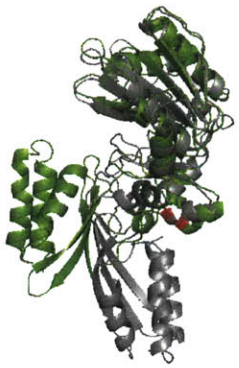


Figure 5-8. Position of T57I on LplA crystal structures. Apo *E. coli* LplA is shown in green, lipoyl-AMP bound LplA is shown in grey. Position of T57I at the end of a helix is shown in red. Structure view created in PyMol.

The next step was to test the 11-Br ligation activity of the consensus clones on the mammalian cell surface. To do this, the various consensus clones were expressed as fusions to neurexin-1β on the surface of HEK cells co-expressing LAP2-CFP-TM. These cells were incubated with 11-Br and ATP for 1 hour. After labeling, the 11-Br ligation onto LAP2 was detected by HaloTag-fluorophore conjugate (HT-AlexaFluor647) (Figure 5-9). In this assay, three of the consensus clones produced high cis labeling that was almost equivalent to labeling with exogenous enzyme (Figure 5-9, top row).

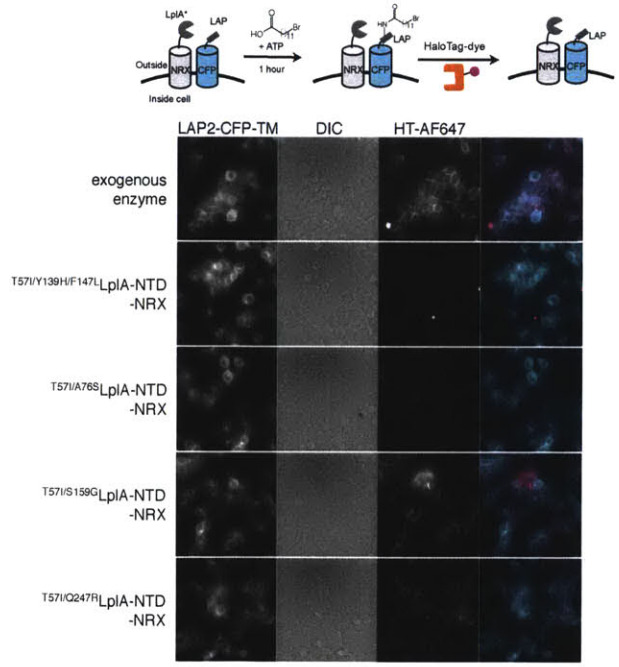


Figure 5-9. Evolved LplA-NTDs have increased 11-Br cis labeling activity on mammalian cell surface. Consensus clones from 11-Br selections on LplA-NTD library were co-expressed on the HEK cell surface with LAP2-CFP-TM. Cis 11-Br ligation is detected using a HaloTag-AlexaFluor647 conjugate.

To test the specificity of these active clones, we performed a side-by-side assay where specific labeling (determined using cis ligation of 11-Br onto LAP2-CFP-TM) was compared to the nonspecific labeling (determined using cis ligation of 11-Br onto LAP2(ala)-CFP-TM) (Figure 5-10). For each enzyme, the ratio of specific to nonspecific signal was approximated by determining the average labeling signal to noise for cells expressing LAP2 and comparing to that obtained when labeling cells expressing LAP2(ala) (Table 5-2). We hereafter refer to the ratios of specific to nonspecific labeling signal as "specificity ratios"; a non-specific enzyme will have a lower specificity ratio.

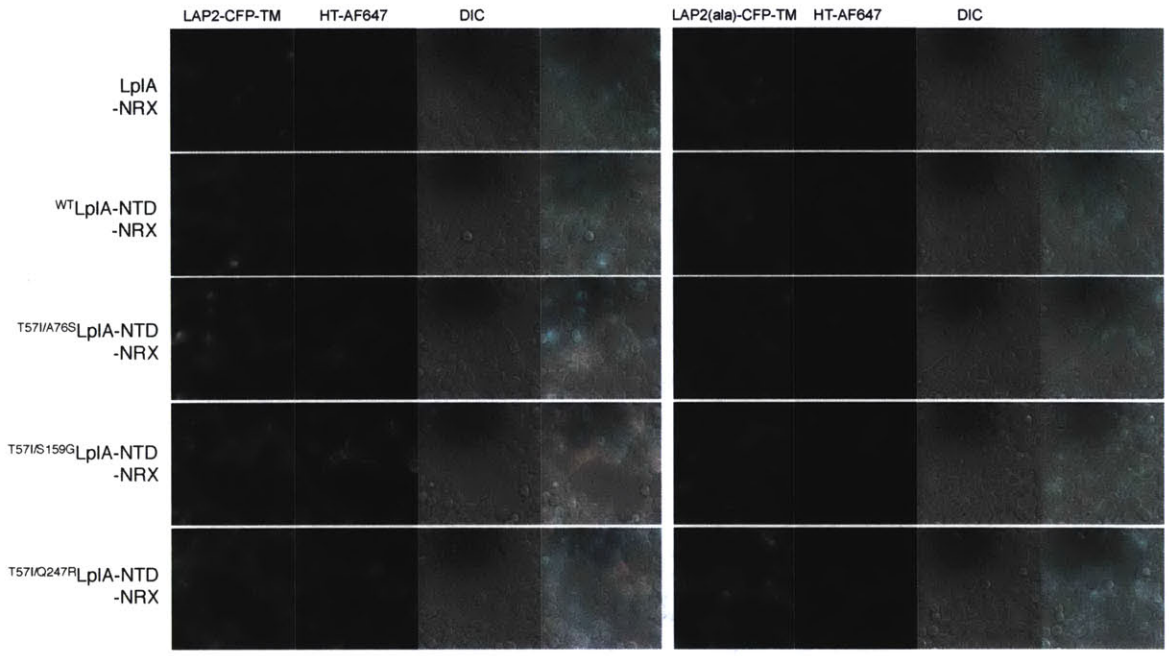


Figure 5-10. Evolved LplA-NTDs exhibit various degrees of non-specific 11-Br labeling. Consensus clones that were active in initial cis 11-Br ligation assays (see Figure 5-9) were co-expressed as fusions to neurexin-1B (NRX) on HEK cells expressing either LAP2-CFP-TM or LAP2(ala)-CFP-TM. Any labeling of LAP2(ala)-CFP-TM is nonspecific. Evolved ligases are compared to ^{WT}LplA-NTD and full-length ^{WT}LplA. Cis 11-Br ligation is detected using a HaloTag-AlexaFluor647 conjugate.

	11-Br-OH ligation
	Specificity ratio
^{WT}LplA-NTD	1.5
$^{T57I/Y139H/F147L}$LplA-NTD	ND
$^{T57I/A76S}$LplA-NTD	4
$^{T57I/S159G}$LplA-NTD	6
$^{T57I/Q247R}$LplA-NTD	5.5

Table 5-2. Specificity ratios for cis ligation of 11-Br using LplA-NTD variants. For the imaging data in Figure 5-10, AlexaFluor647 fluorescence intensities were quantified for ~50 transfected cells for each condition. The ratio of labeling to background fluorescence was calculated for each cell. Then, for each enzyme, specificity ratios were calculated by dividing the average signal to noise observed with LAP2 by the average signal to noise observed with LAP2(ala). Values are rounded to the nearest 0.5.

There were two observations that were surprising in this experiment. First, while full-length ^{WT}LplA produces no cis 11-Br ligation labeling, ^{WT}LplA-NTD produces a low level of cis 11-Br labeling using this assay; and this signal is mostly nonspecific as the specificity ratio is approaching 1. Second, all of the evolved consensus clones exhibit some nonspecific labeling signal. However, in contrast to ^{WT}LplA-NTD, the evolved clones have improved specific 11-Br ligation activity, as all clones have specificity ratios >1. Because of the multi-step detection method, the nonspecific signal on the cell surface could have several sources: nonspecific probe sticking, nonspecific enzymatic labeling, or nonspecific HaloTag fluorophore conjugate. Each of these issues was explored individually.

First, the 11-Br probe could be sticky to the cell surface, producing nonspecific signal that isn't enzyme-dependent. The long hydrophobic probe may stick in membranes, producing background that is detected using HT-fluorophore conjugates. To address the issue of probe sticking we performed cis labeling specificity comparison identically to that described above, but using a 10-Br-AMP probe (Figure 5-11, quantification Table 5-3). We hypothesized that if probe sticking is the source of the problem, the more hydrophilic 10-Br-AMP would have reduced nonspecific background sticking. We discovered that a more hydrophilic probe decreased the background we observed, but did not completely eliminate the nonspecific signal. Also, the reduction in background was not consistent across all mutants.

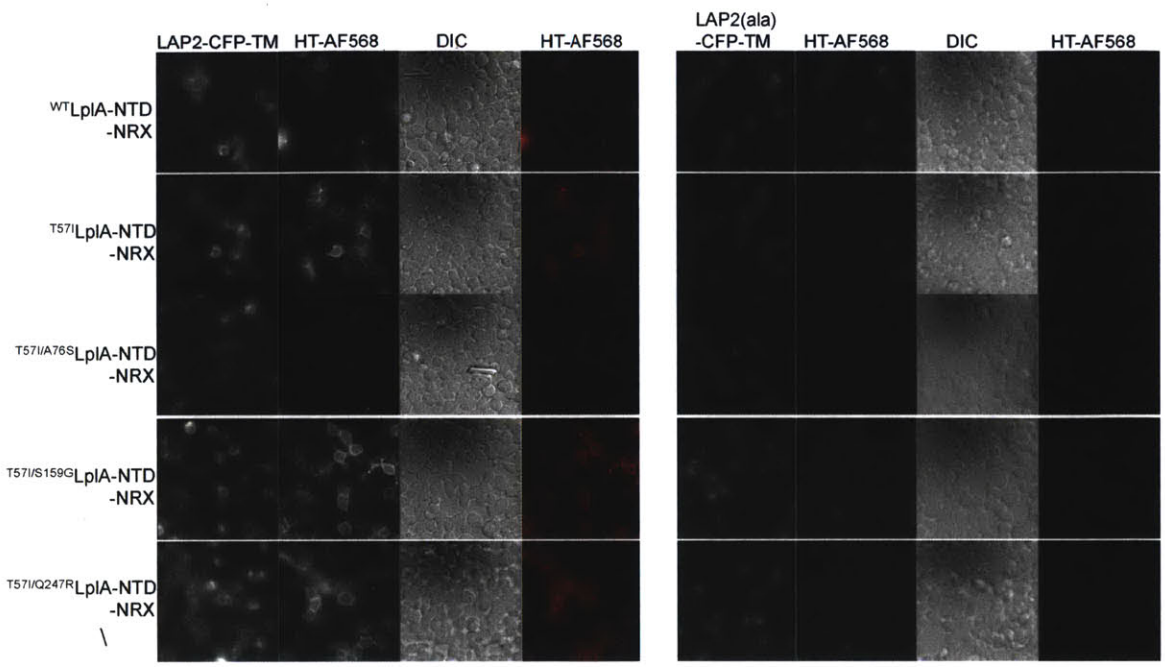


Figure 5-11. Evolved LplA-NTDs exhibit better specificity with 10-Br-AMP labeling. Consensus clones that were active in initial cis 11-Br ligation assays (see Figure 5-9) were co-expressed as fusions to neurexin-1β (NRX) on HEK cells expressing either LAP2-CFP-TM or LAP2(ala)-CFP-TM. Cells were incubated with 10 μM 10-Br-AMP for 20 minutes. Any labeling of LAP2(ala)-CFP-TM is nonspecific. Evolved ligases are compared to ^{WT}LplA-NTD and full-length ^{WT}LplA. Cis 10-Br ligation is detected using a HaloTag-AlexaFluor568 conjugate.

	10-Br-AMP ligation
	Specificity ratio
^{WT}LplA-NTD	2
^{T57I}LplA-NTD	10
$^{T57I/A76S}$LplA-NTD	4
$^{T57I/S159G}$LplA-NTD	8
$^{T57I/Q247R}$LplA-NTD	15

Table 5-3. Specificity ratios for cis ligation of 10-Br-AMP using LplA-NTD variants. For the imaging data in Figure 5-10, AlexaFluor647 fluorescence intensities were quantified for ~50 transfected cells for each condition. The ratio of labeling to background fluorescence was calculated for each cell. Then, for each enzyme, specificity ratios were calculated by dividing the average signal to noise observed with LAP2 by the average signal to noise observed with LAP2(ala). Values are rounded to the nearest 0.5.

Second, the enzymatic ligation of the evolved LplA-NTDs may be nonspecific, producing covalently labeled proteins that are not LAP-tagged. To address this, we used two independent assays to judge specificity of the enzymatic ligation. First, we tested cis lipoylation on the mammalian cell surface, all of the clones exhibited low levels of specific lipoylation of LAP2, and no observable lipoylation of LAP2(ala) (data not shown). However, because the specific lipoylation activity was so low with these evolved mutants, we were not sure if lack of observable LAP2(ala) labeling was due to true

specificity or an artifact of low lipoylation activity. We therefore attempted an additional assay testing for specificity on the yeast cell surface. In this assay we prepared yeast that were expressing the LplA-NTD variant without any LAP2 co-expressed. These yeast cells were incubated with 11-Br probe and ATP and any non-specific labeling was detected using the previously described detection protocol (Figure 5-12).

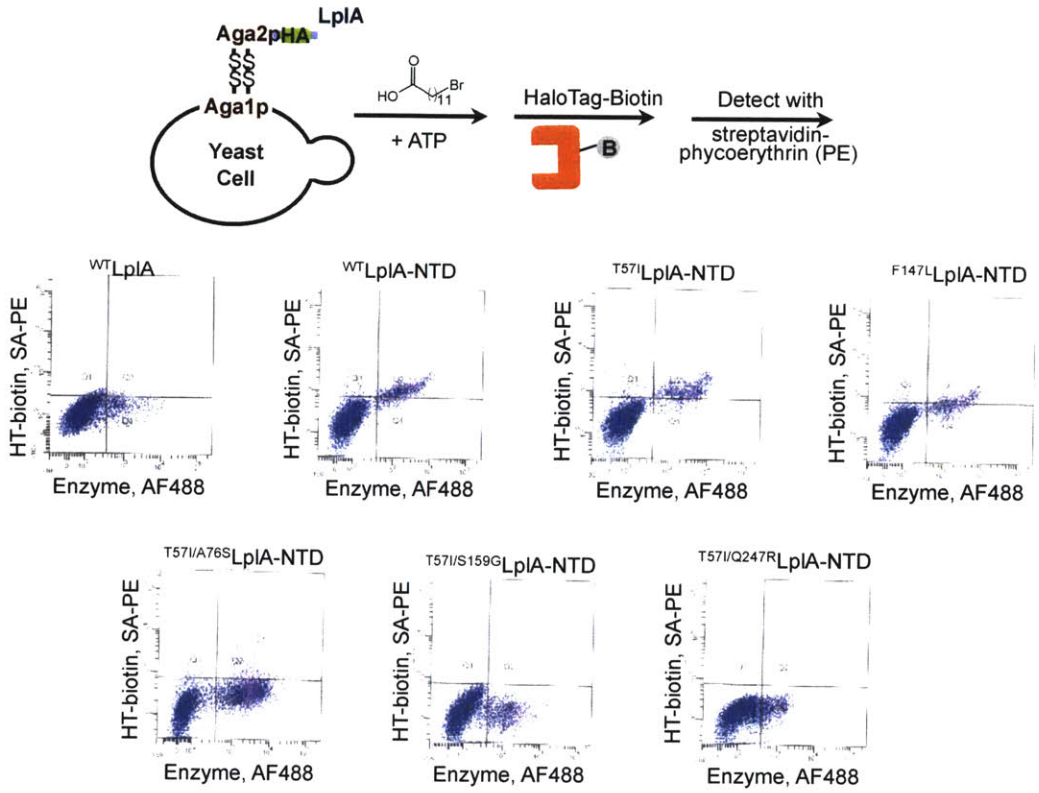


Figure 5-12. Some LplA-NTD enzymes exhibit nonspecific signal on the yeast cell surface. The indicated LplA variant is expressed as a fusion to Aga2p in yeast with no substrate present. The yeast cells are then incubated with 11-Br probe and ATP and any non-specific ligation is detected with HaloTag-biotin conjugate followed by streptavidin-PE. Yeast populations were analyzed by FACS. Any SA-PE signal indicates nonspecific labeling.

Using this assay, two of the evolved LplA-NTD variants demonstrated no non-specific labeling. However, ^{WT}LplA-NTD, ^{T57I}LplA-NTD, ^{F147L}LplA-NTD, and $^{T57I/A76S}$LplA-NTD all produced some non-specific labeling signal. While suggestive that some enzymes may be nonspecifically ligating the probe, this result is not conclusive since the same potential issues with hydrophobic 11-Br probe sticking to the displayed proteins could be an issue in this assay.

Finally, it could be that the HaloTag protein detection is nonspecific due to HaloTag sticking independent of the presence of the probe. To test this, we expressed the LplA-NTD variants in HEK and then applied HaloTag-AlexaFluor647 directly to the cells without any probe incubation (Figure 5-13). What we found was surprising. When HEK cells were expressing full-length ^{WT}LplA, there was almost no sticking of the HaloTag fluorophore conjugate; however, with ^{WT}LplA-NTD and ^{T57I}LplA-NTD there was a significant amount of background sticking.

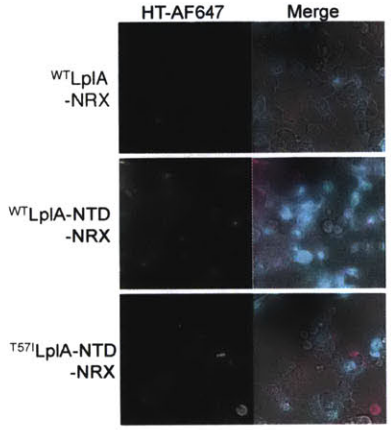


Figure 5-13. HaloTag-fluorophore conjugates give higher background on cells expressing LplA-NTDs. HEK cells were transfected with LAP2-CFP-TM and the indicated enzyme fused to neurexin-1B (NRX). Cells were treated with HaloTag-AlexaFluor647 without any prior probe incubation. Any AF647 signal represents non-specific sticking.

What we hypothesize is that the LplA-NTD variants have more exposed hydrophobic surfaces due to the truncation. These exposed hydrophobic surfaces may then be more sticky to the HaloTag protein, resulting in background. Analysis of the crystal structure shows that a truncation at Glu256 will expose an extensive hydrophobic patch on the protein (Figure 5-14, in green). However, this hypothesis does not explain why the level of HaloTag sticking varies depending on the NTD variant being used.

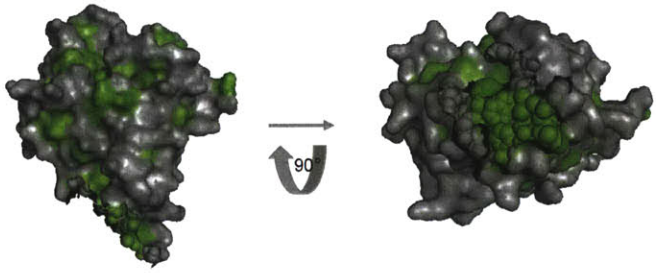


Figure 5-14. Analysis of hydrophobic surfaces of LplA-NTD. Crystal structure of LplA (PDB 3A7R) with truncation at Glu256. Hydrophobic surfaces are colored in green, hydrophilic surfaces in grey. Truncation exposes the hydrophobic core of the protein. All structure views created in PyMol.

Based on these experiments, we concluded that performing 11-Br ligation with the LplA-NTD variants followed by HaloTag-fluorophore conjugate does not produce clean labeling signal on the cell surface. Because ligation of the 11-Br probe was so problematic on the mammalian cell surface, and the evolved LplA-NTD mutants also had very low ligation activity with lipoic acid and no ligation activity with the alkyl azide probes (data not shown), we did not continue with further characterization of these mutants.

Conclusion

This chapter describes efforts towards the directed evolution of the N-terminal domain of LplA (LplA-NTD). In first-generation library selections, we were able to evolve two LplA-NTD variants with improved adenylation activity (^{T57I}LplA-NTD and ^{F147L}LplA-NTD). When lipoic acid and ATP were applied to the cell surface, these LplA-NTD variants demonstrated better cis lipoylation activity onto LAP2 fusions than full-length ^{WT}LplA. In an in vitro assay, we determined that both ^{T57I}LplA-NTD and ^{F147L}LplA-NTD had improved activity compared to ^{WT}LplA-NTD, although the in vitro activity of these enzymes was not equivalent to ^{WT}LplA.

We describe second-generation library selections for cis ligation activity with an unnatural 11-Br probe. These selections produced four consensus clones, three of which had improved activity over ^{WT}LplA-NTD on the mammalian cell surface. However, further investigation demonstrated that the cis 11-Br ligation assay produced moderate levels of nonspecific background labeling. A series of experiments were performed to ascertain the source of the nonspecific signal that we observed. Based on these experiments, we concluded that the nonspecific labeling on the mammalian cell surface is partially explained by probe sticking, but sticking of the HaloTag-fluorophore conjugates or nonspecific enzymatic ligation cannot be completely ruled out.

While we were successful in producing LplA-NTD variants with improved lipoylation activity on the cell surface, we were not able to evolve LplA-NTD variants capable of site-specific labeling using unnatural probes. The reason for this is likely due to the detection method we used. The use of a sticky probe (11-Br) and a protein-based

detection method (HaloTag-fluorophore) likely both contribute to the background observed on the cell surface. From the work in this chapter, we concluded that future evolution schemes and cell-surface ligation assays should rely on more specific detection chemistries or methods and employ small molecules whenever possible.

Experimental

Cis lipoylation on mammalian cell surface (Figure 5-2, 5-5)

HEK cells were transfected with 600 ng of LAP2-CFP-TM and 600 ng of either ^{WT}LplA-NRX, ^{WT}LplA-NTD-NRX, or indicated mutants using Lipofectamine 2000 (Invitrogen) per manufacturer's protocol. 12 hours post-transfection cells were labeled. For labeling with lipoic acid: cells were incubated with 500 μM lipoic acid, 5 mM $Mg(OAc)_2$, and 1 mM ATP for 1 hour at 30°C. For labeling with lipoyl-AMP: cells were incubated with 100 nM lipoyl-AMP and 5 mM $Mg(OAc)_2$ for 1 hour at 30°C. For all samples, cells were washed 2 x 10 minutes with DMEM and then lipoylation was detected via live-cell antibody staining.

Live-cell immunofluorescence staining was performed with 4 μg/ml anti-lipoic acid antibody (CalBiochem) and 4 μg/ml anti-FLAG antibody (Stratagene) in 1% BSA in DPBS for 15 minutes. After three quick washes, cells were incubated with 4 μg/ml goat anti-rabbit antibody conjugated to AlexaFluor 488 and 4 μg/ml goat anti-mouse AlexaFluor 568 (both Invitrogen) in 1% BSA in DPBS for 15 minutes, then washed and imaged.

General Yeast methods

Growth and induction

Unless otherwise noted, cells were grown in SDCAA growth media [2% dextrose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.86% Monosodium phosphate] at 30°C with shaking for 20 hours. To induce protein expression, the cultures were diluted to a cell density of ~10^6 cells/ml in SGCAA media (identical to SDCAA above but with 2%

galactose as the carbon source). Induction in SGCAA was performed at 30°C for 20 hours.

Transformation

For single clones, L2-YIP yeast cells were transformed with DNA using the Frozen-EZ Yeast Transformation II kit (Zymo Research). The transformed yeast cells were plated on selective plates (SDCAA –W, –U) and single yeast colonies were selected for amplification.

For the libraries, L2-YIP yeast cells were electroporated with the error-prone PCR library (see library construction below) using a BioRad GenePulser Xcell.

Long-term storage

For individual clones, glycerol stocks were made. Isolated colonies were grown in 5 mL SDCAA overnight at 30°C. 812 μL of saturated overnight yeast culture was combined with 188 μL of 80% glycerol (sterile) and placed into a 2 mL cryo-vial, stored at -80°C.

For libraries and rounds, frozen stocks were made. At least 10^{10} freshly grown cells from each library were inoculated into 1 L low-dextrose SDCAA [as above but with 0.5% dextrose] and grown at 30°C for approximately 3 days. Cells were pelleted and then resuspended in freezing solution (2% glycerol, 0.67% yeast nitrogen base) to a final concentration of ~6 x 10^{10}/mL. Cells were divided into 1 mL aliquots in cryovials. The tubes were then put into a Mr. Frosty (Nalgene) and allowed to slow-freeze overnight -80°C. The next day, cells were moved to liquid nitrogen for long-term storage.

Sorting and Scanning Instrumentation

All fluorescence activated cell sorting was performed on a BD FACSAria (BD Biosciences). For all analysis where sorting was not required, a BD LSRII-HTS (BD Biosciences) was used. Both of these instruments are located in the Flow Cytometry Core Facility at the Koch Institute for Integrative Cancer Research at MIT.

Preparation of Yeast cells constitutively expressing LAP2-Aga1p

The LAP2-Aga1p yeast integrating plasmid was cloned from the YIPlac211-GPD-Avitag-Aga1p plasmid which was a gift from Irwin Chen at Harvard University. First, two restriction sites, *AflII/NheI* were introduced on either side of the Avitag to create YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. The LAP2 tag was then introduced by pre-annealing LAP2 primers with *AflII/NheI* sticky ends and ligating into *AflII/NheI* digested YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. This produced the yeast integrating plasmid YIPlac211-GPD-*AflII*-LAP2-*NheI*-Aga1p.

To integrate the plasmid into the yeast genome, YIPlac211-GPD-*AflII*-LAP2-*NheI*-Aga1p was linearized by digestion with *BsiWI* and transformed into *Saccharomyces cerevisiae* strain BJ5465 using the protocol described above and plated on SDCAA plates lacking uracil. Colonies were picked and grown in SDCAA lacking uracil. A yeast colony with the LAP2-Aga1p construct correctly inserted and with good expression was designated "LAP2-YIP" yeast and displays the LAP2 peptide sequence constitutively on the cell surface as a fusion to the N-terminus of Aga1p.

Creating LplA-NTD libraries

The round zero (R0) LplA-NTD library was made of a mixture of libraries created with varying mutation rates. These libraries were made by error-prone PCR and cloned into LAP2-YIP *S. cerevisiae* using gap repair homologous recombination.

For all libraries, a 1:1:1 mixture of ^{T57I}LplA-NTD:^{F147L}LplA-NTD:^{WT}LplA-NTD was used as a template for the error-prone PCR.

For Library A, the PCR reaction contained 2 μM 8-oxo-2'deoxyguanosine (8-oxo-dGTP), 2 μM 4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP), 200 μM each dNTP, and 0.2 μM each of primers 1F and 1R and reactions were thermocycled 10 times.

For Library B, the PCR reaction was assembled as above but with 2 μM dNTP analogues and was thermocycled 20 times.

For Library C, the PCR reaction was assembled as above but with 20 μM dNTP analogues and was thermocycled 10 times.

For all libraries, mutagenized genes were further amplified in PCR reactions without mutagenic dNTP analogs using primers 1F and 1R. The pCTCON2-Aga2p vector was prepared by digestion with *NheI* and *BamHI*, and gel-purified. The PCR-amplified inserts and digested pCTCON2 vector were mixed in 1:3 ratios and transformed together into LAP2-YIP *S. cerevisiae* by electroporation as described previously (cite Colby et al). Homologous recombination occurred inside the yeast, giving the desired product. Serial dilutions of transformed yeast were plated on SDCAA plates and colonies were counted, to determine transformation efficiencies for each library. A total of ~10^8 cells from each fully grown library culture were mixed together and then the pooled library was pelleted and induced as described above.

11-Br probe labeling and selections on yeast cell surface (Figure 5-7, Figure 5-12)

Yeast cells constitutively displaying the LAP2 peptide and either the induced LplA libraries or individual clones (see General yeast methods for LplA induction details) were pelleted and washed twice with Tris-buffered saline (pH 7.5) with 1 mg/ml bovine serum albumin (TBS-B). Cells were resuspended to a cell density of 2 x 10^8 cells/ml in TBS-B with 500 μM 11-Br-OH, 5 mM $Mg(OAc)_2$ and 3 mM ATP and incubated on a rotator at 30°C for 1–5 hours. Labeling time for Round 1 was 5 hours, for Round 2 was 2.5 hours, for Round 3 was 1.5 hours, and for Round 4 was 1 hour. The reactions were stopped by pelleting the cells and washing with ice-cold TBS-B containing 100 mM EDTA. Cells were washed twice with ice-cold TBS-B.

The ligated 11-Br probe was detected by incubating with 1 μM HaloTag-biotin for 1 hour at 30C. Yeast cells were then pelleted and washed twice with TBS-B. The biotin was detected using phycoerythrin-conjugated streptavidin (SA-PE; Jackson Immunoresearch) in TBS-B at 1:50 dilution. The LplA surface-expression level was detected using mouse monoclonal anti-myc antibody (Roche) in TBS-B at 1:150 dilution. Incubation with SA-PE and anti-myc occurred simultaneously at 4°C for 40 minutes. Cells were pelleted and washed twice with ice-cold TBS-B. Then the yeast cells were incubated with anti-chicken AlexaFluor-488 secondary antibody (AF488; Invitrogen) in TBS-B at 1:100 dilution for 40 minutes at 4°C. Cells were pelleted and washed twice with ice-cold TBS-B then resuspended in TBS-B to a density of ~1 x 10^7 /mL for FACS.

For the libraries, sorted cells were cultured until saturation in SDCAA growth media with 50 μg/mL penicillin, 25 μg/mL kanamycin, and 50 μg/mL streptomycin. After the culture was saturated, ~10^9 yeast cells were pelleted and induced in SGCAA and prepared for the next round of selection.

For testing of the ligation specificity of individual clones, EBY100 yeast were transformed using the Frozen-EZ Yeast Transformation II kit (Zymo Research). The transformed yeast cells were plated on selective plates (SDCAA –W, –U) and single yeast colonies were selected for amplification. Enzyme expression was then induced, and cells were labeled as described above.

Testing cis 11-Br-OH and 10-Br-AMP ligation on mammalian cell surface (Figure 5-9, 5-10, 5-11)

HEK cells were transfected with 600 ng of LAP2-CFP-TM or LAP2(K→A)-CFP-TM and 600 ng of the indicated LplA-NTD-NRX variant using Lipofectamine 2000 (Invitrogen) per manufacturer's protocol. 12 hours post-transfection cells were labeled with 11-Br-OH or 10-Br-AMP. For labeling with 11-Br-OH: cells were incubated with 500 μM 11-Br-OH, 5 mM $Mg(OAc)_2$, and 3 mM ATP for 1 hour at 30°C. For labeling with 10-Br-AMP: cells were incubated with 10 μM 10-Br-AMP and 5 mM $Mg(OAc)_2$ for 20 minutes hour at 30°C. For all samples, cells were washed 3 x 3 minutes with DMEM and then the ligation of the bromo-probe was detected using HaloTag-dye conjugates. Cells were incubated with HaloTag-AlexaFluor647 or HaloTag-AlexaFluor568 for 10 minutes, then washed 3 x 3 min with DMEM and imaged live.

References

1. 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*, 33645-51.

2. Fujiwara, K., *et al.* Global conformational change associated with the two-step reaction catalyzed by Escherichia coli lipoate-protein ligase A. *J Biol Chem* **2010**, *285*, 9971-80.

3. Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **1997**, *15*, 553-7.

4. VanAntwerp, J. J.; Wittrup, K. D. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol Prog* **2000**, *16*, 31-7.

5. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011**, *108*, 11399-404.

6. Zaccolo, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *J Mol Biol* **1996**, *255*, 589-603.

7. Los, G. V., *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008**, *3*, 373-82.

8. Zang, H., *et al.* Efficient and high fidelity incorporation of dCTP opposite 7,8-dihydro-8-oxodeoxyguanosine by Sulfolobus solfataricus DNA polymerase Dpo4. *J Biol Chem* **2006**, *281*, 2358-72.

9. Colby, D. W., *et al.* Engineering antibody affinity by yeast surface display. *Methods Enzymol* **2004**, *388*, 348-58.

Chapter 6. *In vitro* evolution of full-length LplA for improved activity in the endoplasmic reticulum and cell surface

The work described in this chapter is being prepared for publication. Tao Uttamapinant synthesized the pAz probe and the acetoxymethyl ester protected pAz probe (pAz-AM).

Introduction

The goal of this project is to produce a full-length LplA variant that exhibits improved probe ligation activity both in the secretory pathway and when expressed on the mammalian cell surface. We envisioned that this improved ligase could be applied to PRIME labeling of proteins in the endoplasmic reticulum as well as extend ID-PRIME to the labeling of inter-cellular protein-protein interactions.

As previously discussed, using LplA to directly ligate a fluorophore or probe is preferable for intracellular applications over the two-step labeling protocol, which requires two distinct incubation and washout steps. However, the two-step labeling scheme is ideal for an evolution project; it allows us to evolve one LplA variant to ligate the general functional group handle which can then be chemoselectively derivatized with a variety of probes (Figure 6-1). Using this strategy, one enzyme evolution step can enable us to label proteins with virtually any fluorophore or probe.

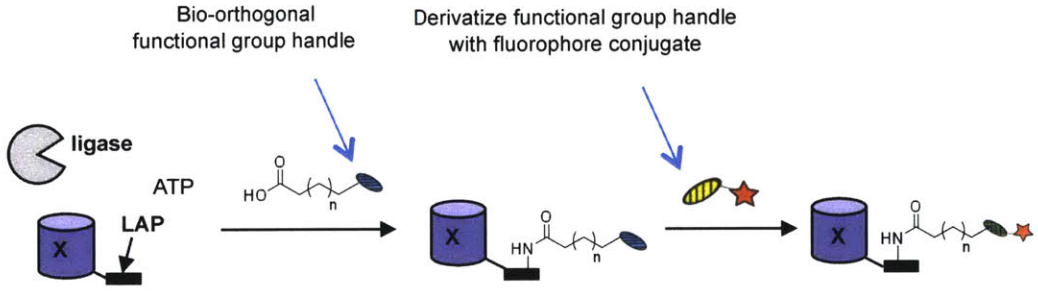


Figure 6-1. Proposed two-step labeling method for enzyme evolution strategy. In the two-step labeling scheme, LplA is first used to enzymatically ligate a bio-orthogonal functional group handle onto the LAP peptide. In the second step, the bio-orthogonal functional group handle is chemoselectively derivatized with any fluorophore- or reporter-conjugate

Unfortunately, our first attempts at using a two-step labeling strategy to evolve LplA-NTD, as described in Chapter 6, produced ligases that were non-specific. This strategy used enzymatic ligation of a bromoalkanoic acid followed by detection using HaloTag conjugates. Not only was the enzymatic ligation of the 11-Br probe non-specific, we also observed significant background sticking of the HaloTag-fluorophore conjugates (Chapter 5). Another drawback of that labeling scheme was its dependence on a HaloTag detection method, which is protein-based and thus not cell-permeable. This

detection method would thus be restricted to cell surface proteins or fixed cells. Because of these issues, we wanted to find a new target probe for the two-step labeling strategy for the evolution of full length LplA

In order to accurately reflect the relative activity of the first enzymatic step in these selections, the second derivatization step needs to be fast and high yielding. LplA has been demonstrated to ligate alkyl azides[1,2], aryl azides[3], and, more recently, picolyl azides[4] with good yield inside cells. Azides can be chemoselectively derivatized via alkynes in a variety of ways, some of which are high yielding and compatible with living cells (Figure 6-2). The traditional [3 + 2] cycloaddition of azides and alkynes was developed by Huisgen; but required high temperatures and pressures in order to produce high yields[5]. Sharpless and colleagues created a Cu(I)-catalyzed version of the reaction that can be performed at room temperature and physiological pH with fast kinetics and high yield[6-8] (Figure 6-2, top). However, these conditions require high concentrations of Cu(I), resulting in cytotoxicity when used on cells[9]. To overcome this limitation, a Cu-free method that uses strain-promoted [3 + 2] cycloaddition with a transcyclooctyne was developed by the Bertozzi group[10,11] (Figure 6-2, middle). This method eliminates the need for copper by relying on the increased reactivity induced by the strain of a transcyclooctyne. The reaction kinetics of this method can be further enhanced by using mono- or di-fluorinated transcyclooctynes[12]. While we have successfully used this method of azide derivatization to label intracellular proteins[2], the kinetics using strain-promoted click (k_{app} = 10^{-3} to 10^{-1} M^{-1} sec^{-1}) are still 10^3- 10^6-fold slower than the copper catalyzed versions (k_{app} = 10^2 to 10^3 M^{-1} sec^{-1}). To address the issue of slow kinetics, Tao Uttamapinant in our lab recently developed a new protocol for cell-compatible copper click chemistry using a copper-chelating picolyl azide probe[4] (Figure 6-2, bottom). This approach allows for the reduction of Cu(I) catalyst but, because the azide probe has a high affinity for copper, there is no concomitant loss in reaction kinetics. Tao's new method reduces the required copper concentration 100 to 1,000-fold compared to the Sharpless protocol, but has similar ligation kinetics[4,7].

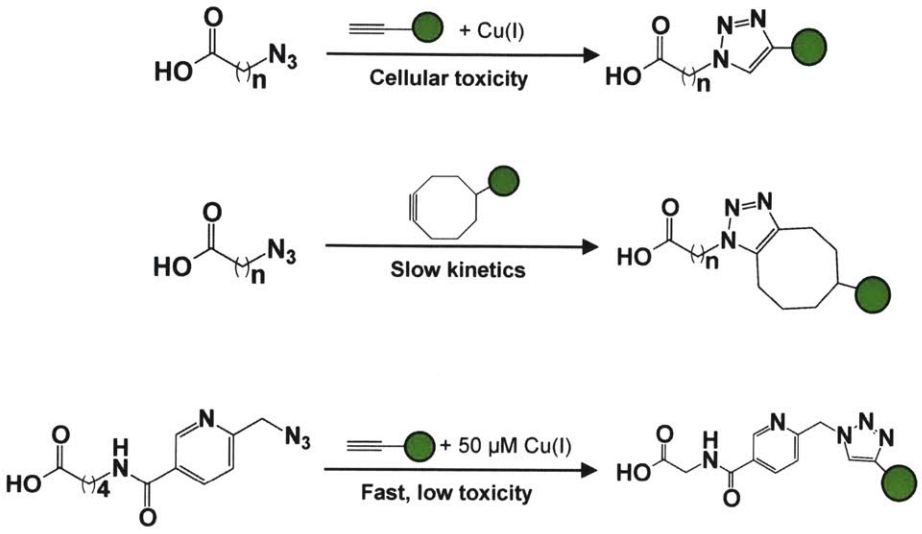


Figure 6-2. Methods for chemoselectively derivatizing azides with alkynes. Copper-catalyzed [3 + 2] cycloaddition developed by Sharpless and Finn is fast at physiological pH, but cytotoxic (top). Copper-free [3 + 2] cycloaddition using strained cyclooctynes developed by Bertozzi lab can be performed in living cells, but the kinetics are slow. Finally, the cell-compatible copper-catalyzed [3 + 2] cycloaddition using copper chelating azides developed by our lab is both fast and nontoxic to cells (bottom).

Tao Uttamapinant demonstrated that when cell-surface proteins were labeled using the picolyl azide probe, the signal obtained following the cell-compatible copper click step was 5- to 20-fold higher than when using alkyl azide probes[4] under identical labeling conditions. We confirmed this result, demonstrating that labeling with the picolyl azide probe (pAz) produced higher labeling signal than with an alkyl azide probe (Az9) when using exogenous ^{W37V}LplA on both the mammalian and yeast cell surface (Figure 6-3). This increase in labeling signal has huge implications for our application. We were previously unable to detect cis ligation on the yeast cell surface when using alkyl-azide probes even with long incubation times (data not shown). After we switched to the picolyl-azide probe, we could detect low cis labeling with an LplA mutant (Results section). Moreover, because both the picolyl azide and the alkyne conjugates are small molecule probes, this detection method is more versatile than the HaloTag detection strategy used for LplA-NTD selections (Chapter 5)

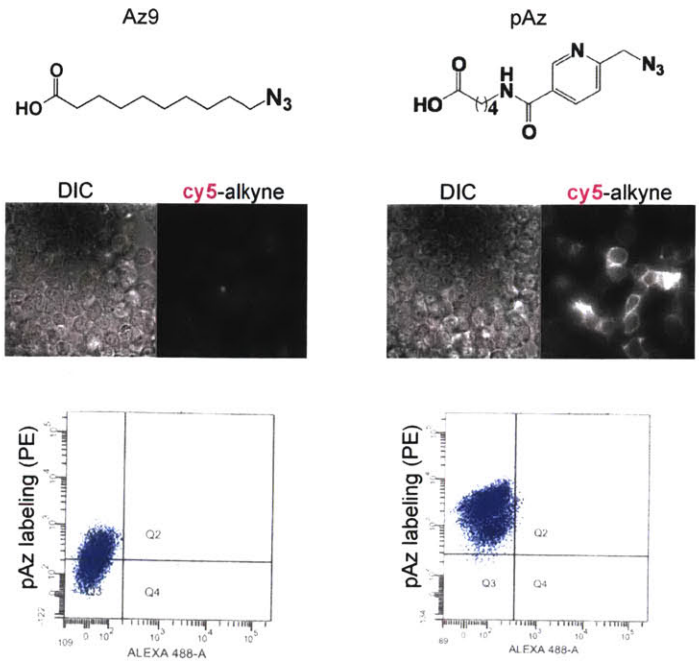


Figure 6-3. Comparison of pAz and Az9 labeling using exogenous ^{W37V}LplA. Top images show labeling of HEK cells expressing E2p on the cell surface using ^{W37V}LplA and either alkyl azide (Az9, left) or picolyl azide (pAz, right). Fluorescent activated cell sorting (FACS) was used to generate the lower plots which show labeling of yeast cells expressing LAP2 on the cell surface using ^{W37V}LplA and either alkyl azide (Az9, left) or picolyl azide (pAz, right).

In this chapter, we describe the *in vitro* directed evolution of a full-length PRIME ligase with improved ligation activity *in vitro*, in the secretory pathway, and on the cell surface. First, we describe how we performed yeast-display of an LplA enzyme library and selected for LplA variants with improved picolyl azide (pAz) probe ligation activity. Second, we present the characterization of the isolated LplA variants *in vitro* and on the surface of yeast and mammalian cells. Finally, we demonstrate how the optimized pAz ligase can be applied to label intercellular neurexin and neuroligin interactions on live cells.

Results

Selection platform based on co-display of ligase and peptide on yeast cell surface

We proposed using a yeast-display system identical to what we previously used for evolution of the truncated LplA (Chapter 5) and pioneered by Chen and colleagues[13]. To briefly reiterate, in this system the enzyme and substrate are co-displayed on the yeast cell surface. The LAP2 peptide (GFEID**K**VWYDLDA) is constitutively expressed as a fusion to the Aga1p mating protein. We can then induce the expression of the LplA enzyme library, which is displayed as a fusion to Aga2p (Figure 6-4A). After applying labeling reagents to the cells, we can measure the amount of cis labeling of LAP2 peptide catalyzed by the tethered enzyme. Although labeling is performed at the cell surface, this strategy allows us to select for LplA variants that have high activity after being processed through the secretory pathway.

The main concern with this evolution platform is the potential for artificially low turnover when using a surface-tethered ligase, which may restrict access to surface-expressed peptide. This is particularly problematic on the yeast cell surface, which is not as fluid as the mammalian cell surface[14]. To first demonstrate that cis labeling is possible with the full-length ligase, yeast displaying ^{WT}LplA and the LAP2 peptide were incubated with lipoic acid and ATP for three hours. After washing, lipoylation of the LAP2 peptide was detected by anti-lipoic acid antibody (phycoerythrin (PE) signal) and enzyme expression level was detected by anti-HA antibody (AlexaFluor488 signal). After antibody staining, the fluorescently labeled yeast pools were analyzed by fluorescence-activated cell sorting (FACS) (Figure 6-4B). When yeast cells displaying ^{WT}LplA were analyzed, LAP2 cis lipoylation was observed, but the signal was not as high as identically prepared yeast cells lipoylated with exogenously applied enzyme (Figure 6-4C). This control is consistent with our previous observations that full-length ^{WT}LplA has decreased activity when expressed on the cell surface compared to when it is expressed and purified from *E. coli*. We also observed lower lipoylation signal when we performed cis lipoylation using yeast cells expressing ^{W37A}LplA, which has about 30% the activity of wild-type *in vitro*, although enzyme expression level was similar (Figure 6-4C, right-

most panel). This result demonstrates that the cis labeling assay is capable of distinguishing between homogeneous populations of a highly active ligase (^{WT}LplA) and a less-active ligase (^{W37A}LplA). This platform faithfully reflects the observed activity disparities between various LplA mutants.

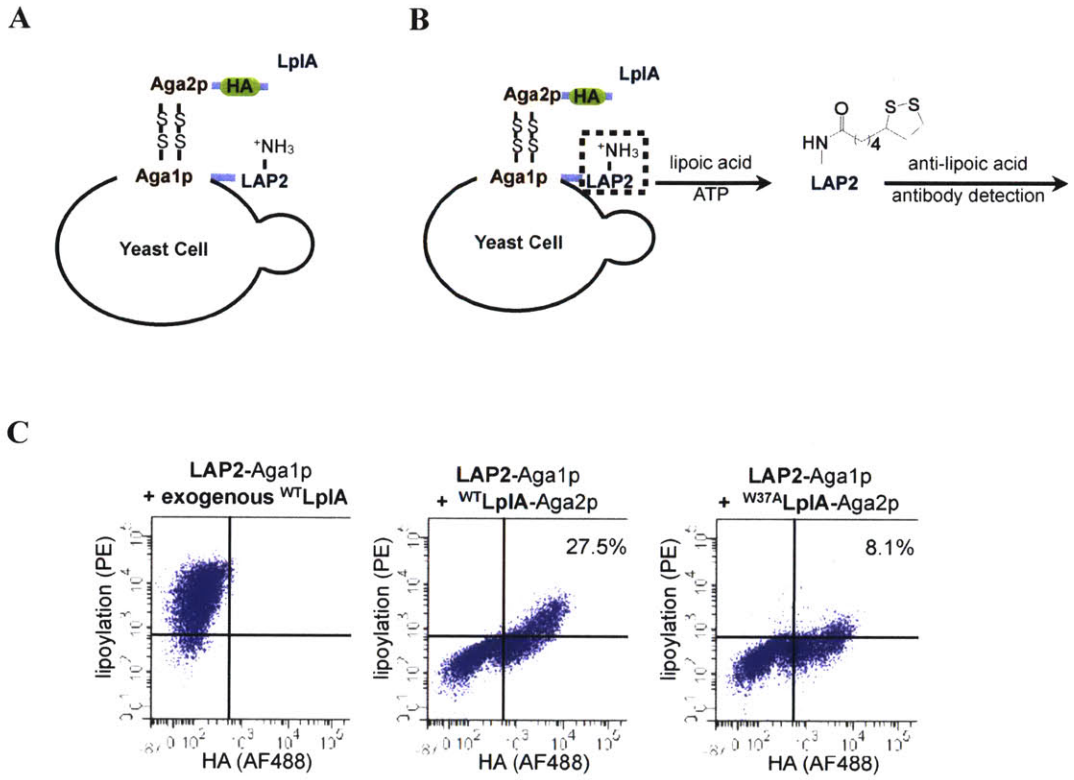


Figure 6-4. Demonstration of cis lipoylation activity using LplA and LAP2 co-displayed on yeast cell surface. (A) Diagram of co-display. Yeast cells constitutively express LAP2 as a fusion to the Aga1p mating protein, and a subset transiently express LplA as a fusion to Aga2p. The HA epitope tag can be used to monitor LplA expression levels. **(B)** Scheme for cis lipoylation protocol. Yeast cells are treated with lipoic acid and ATP and lipoylation is detected by antibody staining and fluorescence activated cell sorting (FACS). **(C)** FACS analysis of yeast cells treated with lipoic acid and ATP for 3 hours. Lipoylated LAP2 was detected by anti-lipoic acid antibody (phycoerythrin (PE), y-axis) and enzyme expression was quantified using anti-HA antibody (AlexaFluor488, x-axis). Each dot in the scatter plot represents a single yeast cell. Yeast cells expressing ^{WT}LplA have higher cis lipoylation signal (27.5%) than yeast cells expressing ^{W37A}LplA (8.1%). Lipoylation of LAP2 yeast with exogenous ^{WT}LplA is shown as a control.

We next wanted to demonstrate that this selection strategy can enrich for high turnover ligases from a heterogeneous population. To do this, we performed one round of model selection on mixtures of ^{WT}LplA and ^{W37A}LplA yeast. For the model selections, yeast cells were mixed in 1:1, 1:10, and 1:100 ratios of ^{WT}LplA: ^{W37A}LplA yeast and then labeled with lipoic acid. After fluorophore staining, cells that exhibited a high ratio of PE to AlexaFluor488 were sorted out via FACS (Figure 6-5). The FACS plots show that

before sorting the population of highly labeled yeast decreased with increasing dilution of ^{WT}LplA but that after sorting, the population of highly labeled yeast was enriched. In this model selection experiment, we observed ~50-fold enrichment of ^{WT}LplA over ^{W37A}LplA (Figure 6-5, chromatograms). These results demonstrate that the detection of LAP2 cis labeling is an effective readout of enzyme activity and that the selection strategy enriches for ligases with high ligation activity.

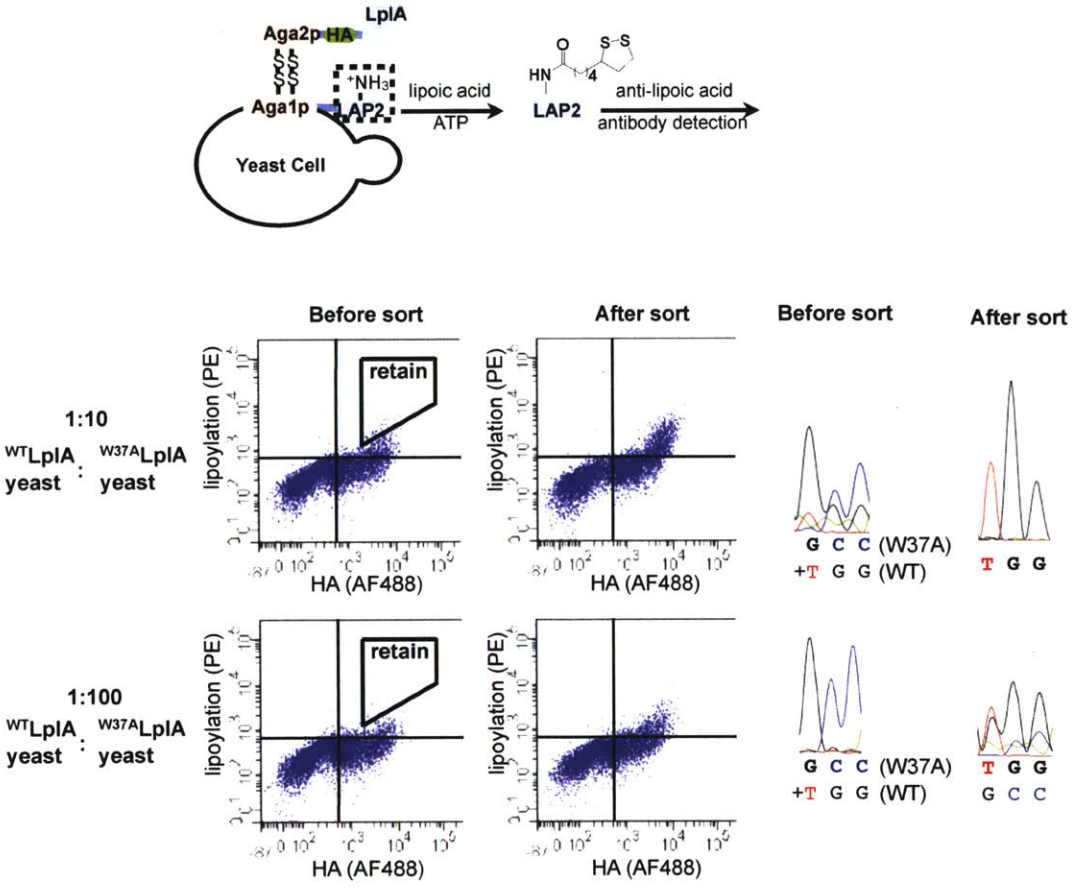


Figure 6-5. Model selection enriches for highly active ligase over less active ligase. Samples were mixed in 1:10 or 1:100 ratios of ^{WT}LplA-expressing yeast to ^{W37A}LplA-expressing yeast. The mixed samples were labeled with lipoic acid and ATP for 3 hours, then stained as previously described and sorted by FACS using the gate shown. Chromatograms show sequencing at residue 37 for the yeast mixtures before and after a single round of selection. Under these conditions, ~50-fold enrichment of ^{WT}LplA is calculated based on sequencing enrichment assay.

Directed evolution of a picolyl azide (pAz) ligase

Rationally designed ligases have no activity on yeast cell surface

After we validated the evolution platform using a model selection for lipoic acid, the next step was to use the platform to select for an LplA variant with improved picolyl azide ligation activity. The general selection scheme is presented in Figure 6-6. In the first step, the yeast are incubated with pAz and ATP allowing for enzymatic ligation of the pAz probe to LAP2 on the cell surface. Then, in order to maximize the fluorescent labeling signal, we perform cell-compatible copper-click with a biotin alkyne followed by streptavidin-PE detection. The enzyme expression level is detected using anti-HA antibody as previously described. This protocol produces an effective fluorescent readout of the enzymatic pAz ligation activity.

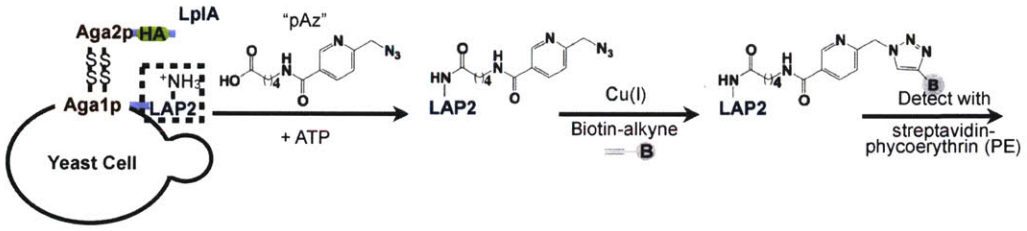


Figure 6-6. Labeling strategy for pAz selection scheme. In the enzymatic ligation step, LplA catalyzes the ATP-dependent attachment of pAz to LAP2. In order to detect pAz ligation, a two-step protocol is used. First, the azide is derivatized with biotin alkyne using cell-compatible copper-click. Then, the biotin is detected using a streptavidin-phycoerythrin (PE) conjugate.

Before performing the selection, we wanted to first test pAz labeling on the yeast cell surface using the best rationally designed ligases. This experiment would provide a benchmark for the selection and further demonstrate the decreased activity of LplA mutants when expressed on the cell surface. We therefore tested the pAz ligation activity in cis using ^{W37V}LplA and ^{W37A}LplA, the two most active pAz ligases *in vitro*. While pAz labeling of the LAP2 yeast using exogenous ^{W37V}LplA produced very high signal in just one hour, no cis labeling was observed with either ^{W37V}LplA or ^{W37A}LplA, even after 12 hours of probe incubation (Figure 6-7A). In fact, over 20 hours of probe incubation was required to observe any pAz cis labeling signal using ^{W37A}LplA on the yeast cell surface (Figure 6-7B). This result was encouraging because it demonstrated that cis pAz ligation

is achievable on the yeast cell surface, although the ligation activities of the rationally designed ligases are simply not high enough to accumulate appreciable labeling signal within a biologically relevant timescale.

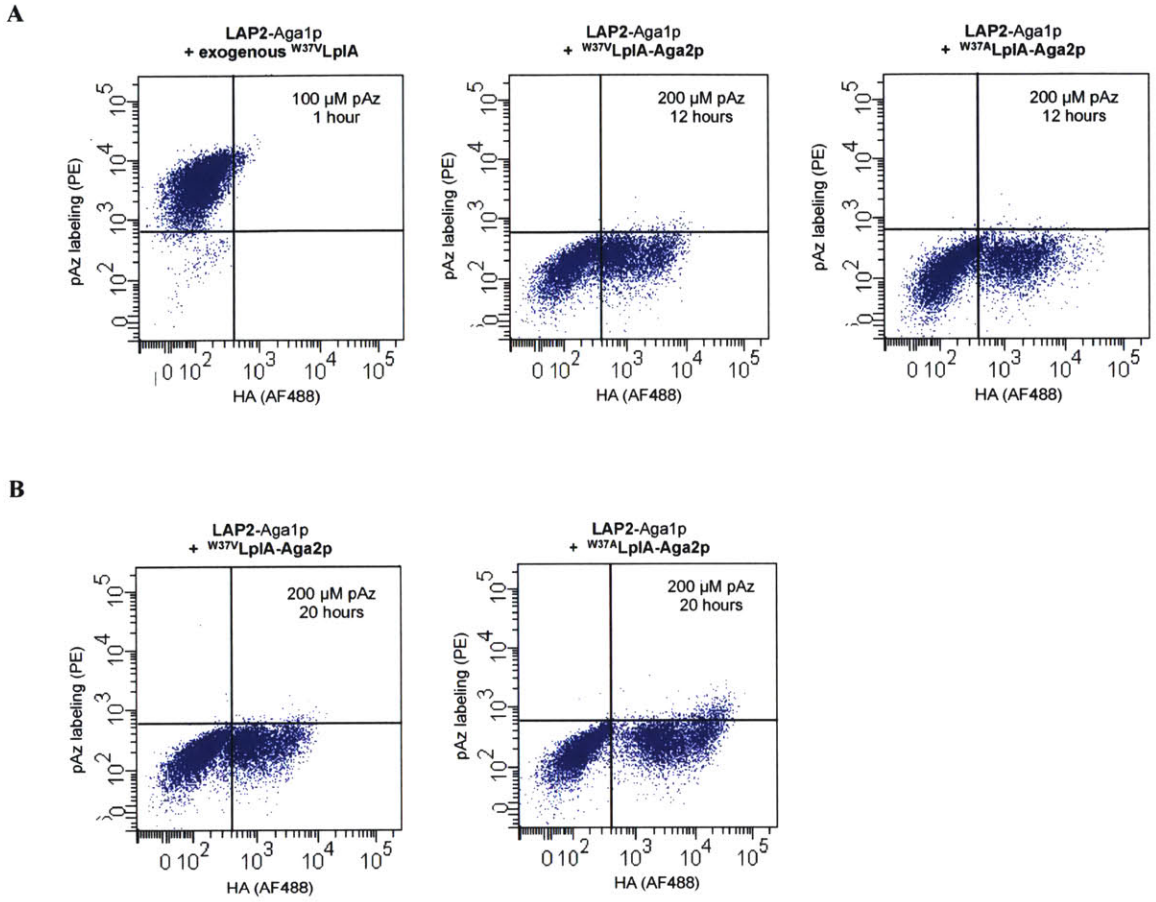


Figure 6-7. Testing cis pAz ligation using rationally designed ligases. (A) Yeast cells expressing either ^{W37V}LplA or ^{W37A}LplA were labeled with ATP and the indicated concentration of pAz probe for the indicated reaction time. Detection of pAz labeling was carried out as previously described. As a control, pAz labeling with exogenous ^{W37V}LplA produces very high labeling signal, even at short incubation times. **(B)** In order to observe cis pAz labeling of LAP2, very long incubation times (20 hours) were required with ^{W37A}LplA-expressing yeast. Even at long incubation times ^{W37V}LplA-expressing yeast do not show appreciable cis pAz labeling.

Selection for an improved pAz ligase

The next step was to create LplA enzyme libraries for the pAz-based activity selections. Although ^{W37V}LplA is the best pAz4 ligase *in vitro*, the libraries were made using ^{W37A}LplA as a template. We chose this template because ^{W37A}LplA has been shown to ligate the pAz probe *in vitro*[4] and has superior expression to ^{W37V}LplA on the yeast cell

surface (Figure 6-7B, x-axes). As previously described for the LplA-NTD libraries (Chapter 5), error-prone PCR was used to generate libraries of LplA gene variants which were then incorporated into the yeast display vector using gap homologous recombination. Since it was not initially clear whether a high mutation rate would be beneficial or detrimental to the activity of LplA, six libraries were produced with varying mutation rates. In the interest of brevity, this thesis will only discuss the two libraries that produced promising LplA variants. The first library, Library A, was made using a low mutation rate, with 1-2 non-silent mutations per gene (8.0×10^7 transformants). Library B was made using a moderate mutation rate of 1-6 non-silent mutations per gene (2.5×10^7 transformants).

Each library was independently subjected to four rounds of selection for pAz4 ligation; yeast pools were amplified between rounds but no further library diversification was performed. Evolutionary pressure for improved catalytic activity was introduced by decreasing the labeling times (from 12 hours to 1 hour) and by collecting a smaller fraction of cells in each successive round. For rounds 3 and 4, we applied additional evolutionary pressure for improved pAz substrate recognition by reducing the concentration of pAz probe (from 200 μM to 100 μM) (Figure 6-8).

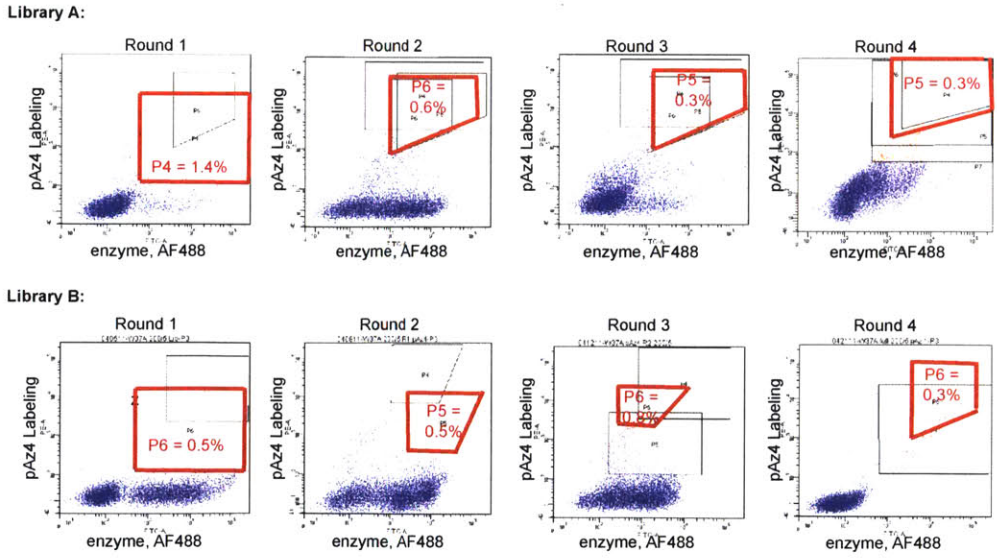


Figure 6-8. Sort windows for each round of pAz selections. Selections were performed using increasingly stringent labeling for each round: Round 1– 200 μM pAz, 12 hours; Round 2– 200 μM pAz, 6 hours; Round 3– 100 μM pAz, 2.5 hours; Round 4– 100 μM pAz, 1 hour. Sort gates are indicated in red with the raw sort percentages reported for each round. Note: recovery percentages reported in Figure 6-9 correct these raw values based on the induction percentage from each round.

After completing the selections, we assayed the yeast from each round under identical conditions to validate our selection strategy (Figure 6-9). While enzyme expression levels (x-axis) remained relatively constant through the selections, we observed a steady increase in pAz labeling intensities (y-axis), with maximal labeling signal observed at round 4. From this FACS data, we concluded that the selection strategy successfully enriched library members with improved pAz ligation activity.

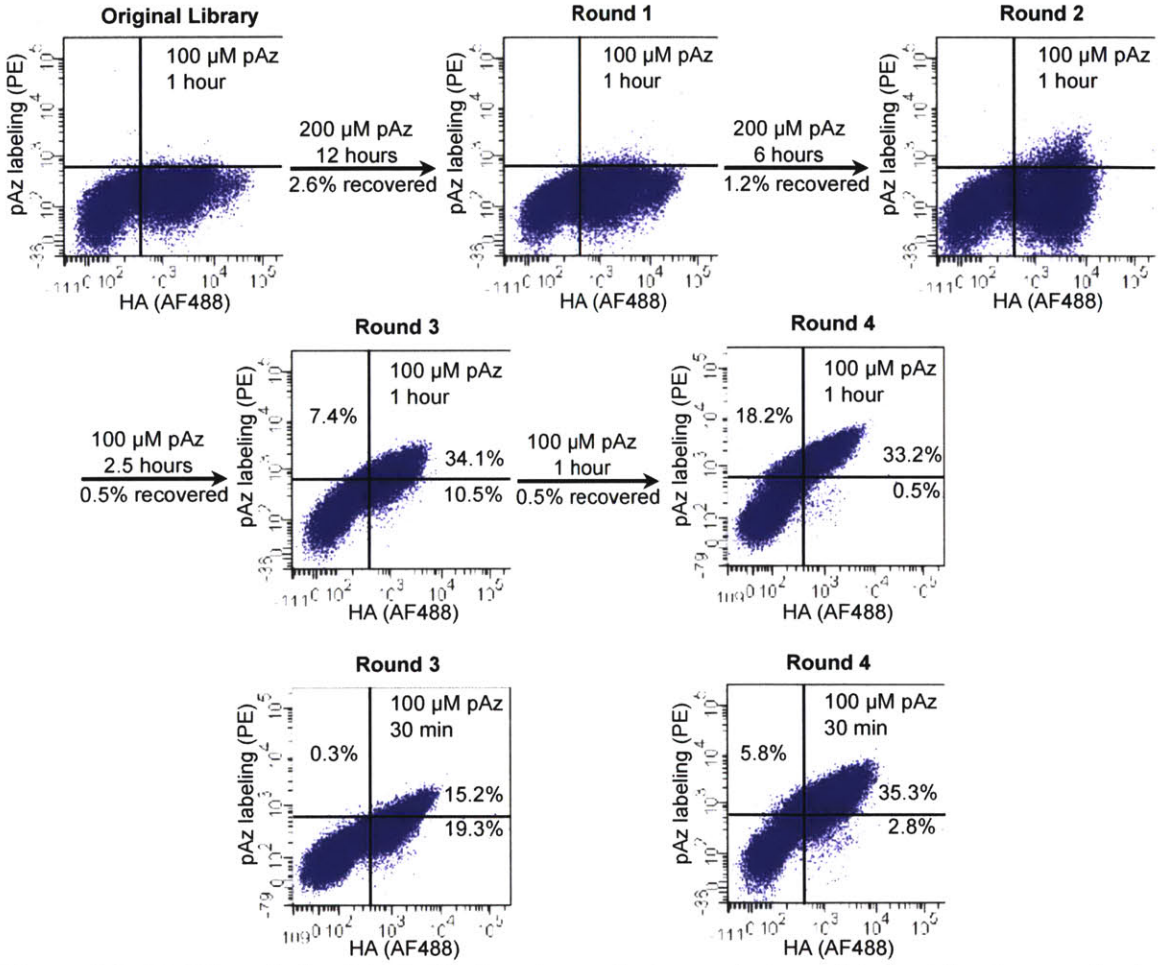


Figure 6-9. FACS analysis of individual sort rounds demonstrate enrichment for improved pAz labeling activity. Re-amplified yeast pools from each round of selection on Library A were labeled under identical conditions (100 μM pAz, 1 hour). Selection conditions used for each round are reported above each arrow along with percentages of cells recovered. To better illustrate the difference between the pAz ligation activities of round 3 and round 4, these rounds were also analyzed using a 30 minute labeling protocol (bottom panels). Library B shows similar enrichment (data not shown).

In order to monitor the library diversity and identify trends in conserved mutations, sequencing analysis was performed on clones isolated from each round for each of the libraries (Table 6-1). For Library A the original library of 8.0×10^7 members was reduced through four rounds of selection to just two consensus clones (A4.6 and A4.7), while Library B was reduced from 2.5×10^7 members to three consensus clones (B4.2, B4.3, and B4.8). While we will discuss the individual mutations in more detail later in this results section, it is interesting to note that the majority of highly conserved mutations were located either in the C-terminal domain of LplA (after residue 256) or in areas far from the small-molecule binding pocket. In fact, only one highly conserved mutation, F147L, is located in a buried region of the protein and near the small molecule binding pocket.

	I23	W37	R38	V44	S100	I103	F147	R181	K224	A248	R258	H267	F281	N286	Q321	Other Mutations
A1.1		A														P287T
A1.2		A														L291V
A1.3		A														W52R, A110T
A1.4		A														D157N
A1.5		A		A												
A1.6	V	A													R	
A1.7		A						S							R	
A1.8		A								T						
A2.1		A														E265K
A2.2		A														F318L
A2.3		A											L			
A2.4		A											S			
A2.5		A								T						
A2.6		A								T						
A2.7		A								T						
A2.8		A										R				
A3.1		A			G											
A3.2		A			G											
A3.3		A					L									
A3.4		A										R				
A3.5		A					L					R				
A3.6		A					L			T						
A3.7		A								T						
A3.8		A								T						
A4.1		A			G											
A4.2		A			G											
A4.3		A			G											
A4.4		A			G											
A4.5		A			G											
A4.6		A			G											
A4.7		A					L					R				
A4.8		A					L					R				
B1.1		A														A150V, F259L, Y302C, M306V, K323R
B1.2		A														V113A, D122G
B1.3		A														F86L, K126E, I177T
B1.4	V	A												G		N16S, L81P, Y139H, H149Y
B1.5		A		A												N166G, D256G
B1.6		A				T										T101I, E324G
B1.7		A							R		G					R183E
B1.8		A							R						R	
B2.1		A														E257G, Q308R, L315P
B2.2		A						G							R	F268L
B2.3		A										R				T188A
B2.4		A	G		G	T										
B2.5		A	G			T										
B2.6		A								T						
B2.7		A											L			H148Y
B2.8		A									G		L			
B3.1		A												D		
B3.2		A	G			T										
B3.3		A	G			T										
B3.4		A											L			
B3.5		A								T			L			
B3.6		A								T						
B3.7		A								T						
B3.8		A								T						
B4.1		A								T						
B4.2		A								T						
B4.3		A								T			L			
B4.4		A								T			L			
B4.5		A								T			L			
B4.6		A								T			L			
B4.7		A								T			L			
B4.8		A											L			

Table 6-1. Sequencing results from each round of selection. After each round, yeast samples were plated and individual clones were randomly selected for sequencing. The LplA clone names sequenced from each round are listed down the rows of the table and the LplA residues that were mutated in the selected clones are listed across the columns. Positions with highly conserved mutations are highlighted in yellow. The W37A mutation was part of the original library design.

Evaluation of evolved ligase mutants on yeast and mammalian cell surface

Evolved ligases are specific on the yeast cell surface

Since we had observed issues with non-specific labeling using consensus clones from previous libraries (Chapter 6), we first wanted to test the activity and specificity of the individual consensus clones on the yeast cell surface. The activities of the five consensus clones were compared by performing pAz cis ligation onto the LAP2 peptide (Figure 6-10 top row). To monitor the extent of nonspecific ligation using these mutants, pAz cis labeling was performed on yeast cells constitutively expressing the acceptor peptide for biotin ligase (AP), which should not be recognized by LplA (Figure 6-10, bottom row). While most of the consensus clones were highly specific, some nonspecific labeling of AP was observed when using B4.3.

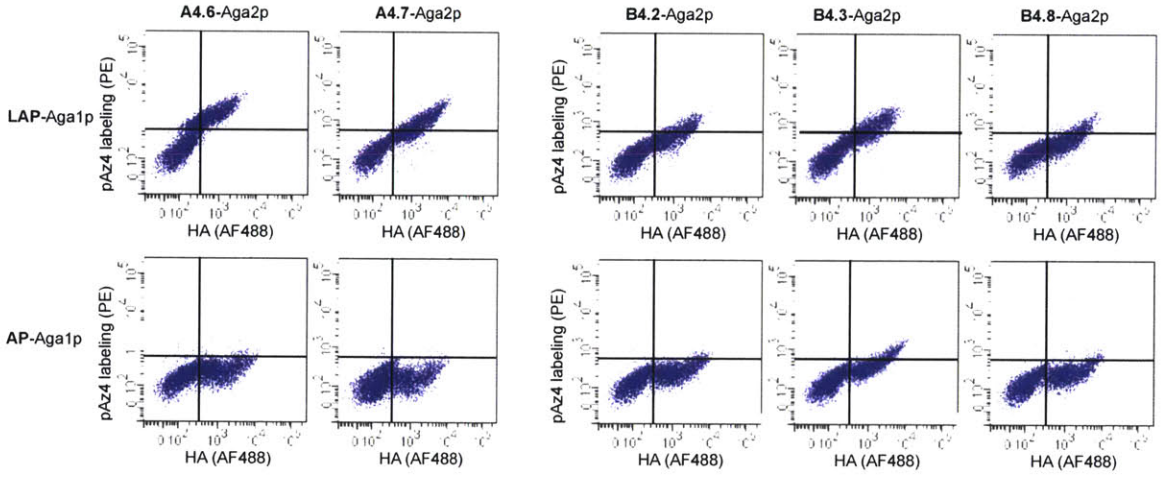


Figure 6-10. Activity and specificity of the five consensus clones. The two consensus clones from Library A and three consensus clones from Library B were amplified after the fourth round of selection. FACS plots show the extent of pAz cis ligation onto LAP2 observed using each LplA variant as a function of its expression level (HA, x-axis) (top row). To analyze the specificity of pAz ligation, we isolated the expression plasmids and introduced them into yeast displaying the acceptor peptide for biotin ligase (AP), which should not be recognized by LplA. FACS plots in the bottom row show an absence of pAz ligation onto AP for all clones except B4.3, which shows significant labeling signal. For all samples, labeling was performed with ATP and 100 μM pAz for 1 hour.

Evolved ligases are active on the mammalian cell surface

We next wanted to demonstrate that the activity we evolved is not context-dependent, and that we are able to achieve high pAz labeling with the enzymes when taken out of the context of the yeast cell surface. To that end, we developed an assay to test pAz labeling in cis on the surface of mammalian cells. We first expressed the ligases

as fusions to neurexin-1β (NRX). Neurexin-1β was chosen because this protein gives good surface expression in mammalian cells and we have previously demonstrated that ^{WT}LplA can catalyze the ligation of lipoic acid in cis when fused to this protein (Chapters 4 and 5). LAP2-CFP-TM was co-expressed and the cells were incubated with ATP and pAz, which was followed with click chemistry using an AlexaFluor647-alkyne (Figure 6-11, scheme). In this assay, we observed labeling of the LAP2 peptide using all of the mutants, which demonstrates that labeling with the evolved ligases is not dependent on a yeast-display context (Figure 7-11A). More importantly, this result was the first demonstration of PRIME labeling using enzymes expressed on the cell surface.

To better assess activity trends, we quantified the extent of labeling signal and compared it to both enzyme and substrate expression levels (Figure 6-11B). The highest ratios were observed with A4.7 (W37A/F147L/H267R) followed by A4.6 (W37A/S100G) and then B4.2 (W37A/A248T) and much lower labeling to expression ratios were obtained with the other consensus clones from Library B. While consistent with the yeast cell surface activity trends, the cis labeling signal observed when using the evolved ligases on the mammalian cell surface was still 3- to 8-fold lower than the exogenous labeling signal.

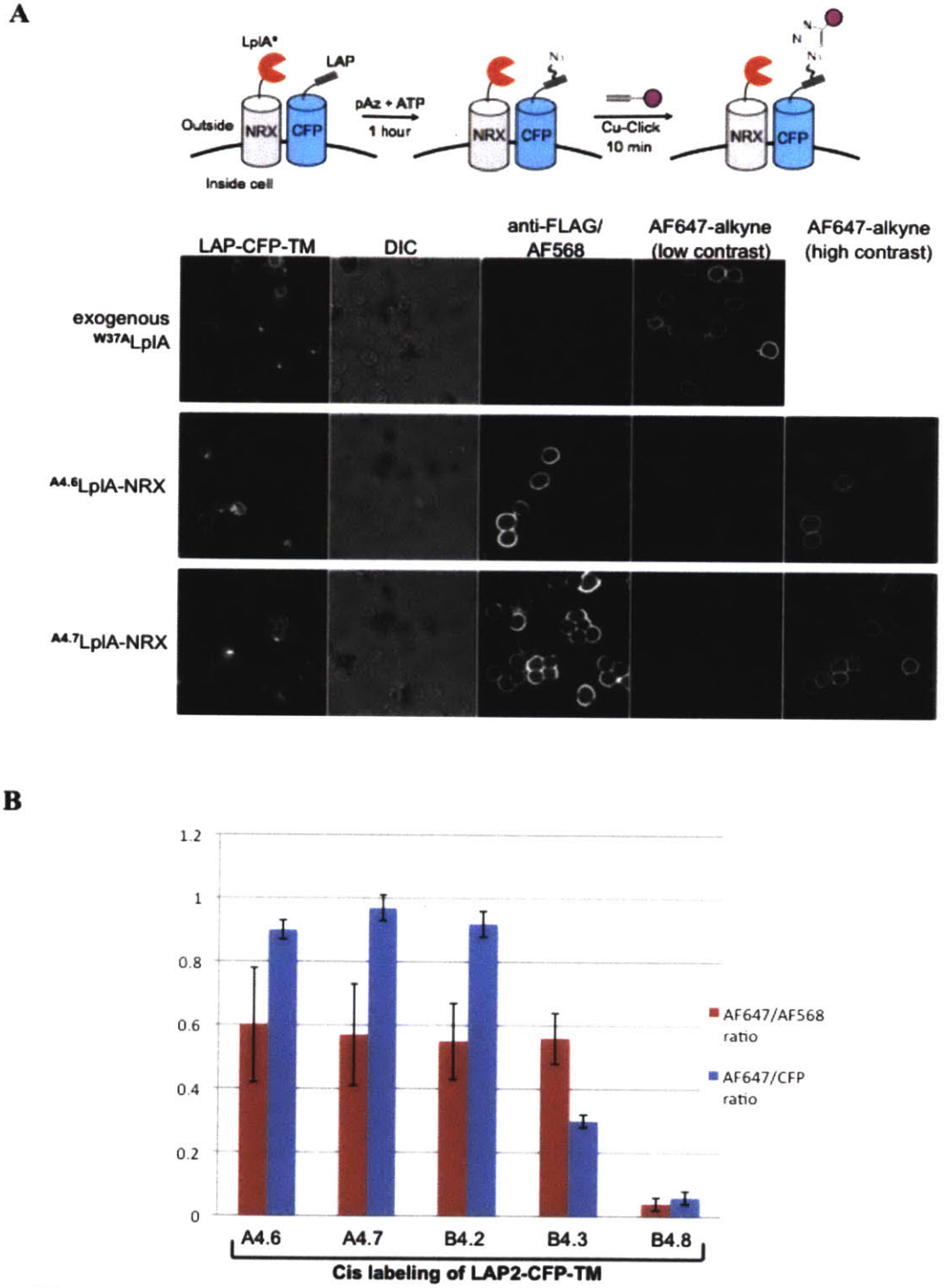


Figure 6-11. Analysis of the cis pAz ligation activity using evolved LplA variants on mammalian cell surface. (A) Each LplA mutant was fused to the N-terminus of neurexin-1β (NRX) and co-expressed with LAP2-CFP-TM on the surface of HEK cells. Cis labeling was performed with addition of 100 μM pAz and ATP for 1 hour. The pAz-LAP2 conjugates were detected by [3 + 2] cycloaddition with AlexaFluor647-alkyne. LplA-NRX expression was detected by live-cell antibody staining with anti-FLAG antibody. All consensus mutants were tested, but representative confocal images for $^{A4.6}$LplA and $^{A4.7}$LplA are shown. **(B)** Quantification of pAz cis labeling extent on the mammalian cell surface for all five LplA mutants. Fluorescence intensities were calculated from confocal images for at least 30 individual cells.

$^{A4.7R}$LplA labels LAP2 fusions with signal equivalent to exogenous enzyme

We next sought to improve the cis labeling activity of the evolved ligases on the mammalian cell surface by incorporating promising mutations from the NTD selections. Through the LplA-NTD selections, we discovered that Thr57 of LplA is a critical position in the enzyme and, when mutated to isoleucine (T57I), improves LplA expression, stability, and activity across a range of probes and applications (see Chapter 6). We hypothesized that this mutation might also improve the activity of the evolved ligases. To test this, we added the T57I mutation to the most active evolved ligase, $^{A4.7}$LplA (W37A/F147L/H267R), to create $^{A4.7R}$LplA (W37A/T57I/F147L/H267R). When we included this new mutant in the cis mammalian cell assay, we found the $^{A4.7R}$LplA was more active than $^{A4.7}$LplA and produced cis pAz labeling signal that was equivalent to exogenous enzyme in all conditions tested (Fig 6-12A). To ensure that the labeling was generalizable, we tested cis labeling of LAP2 fused to neuroligin-1 (LAP2-NLG) and observed a similar increase in activity (Fig 6-12B). When labeling signal was quantified and compared to both enzyme expression level and substrate expression level, $^{A4.7R}$LplA outperformed all of the evolved ligases in the cis ligation assay (Fig. 6-12C). With this result, we determined that $^{A4.7R}$LplA is the best pAz ligase for mammalian cell-surface applications.

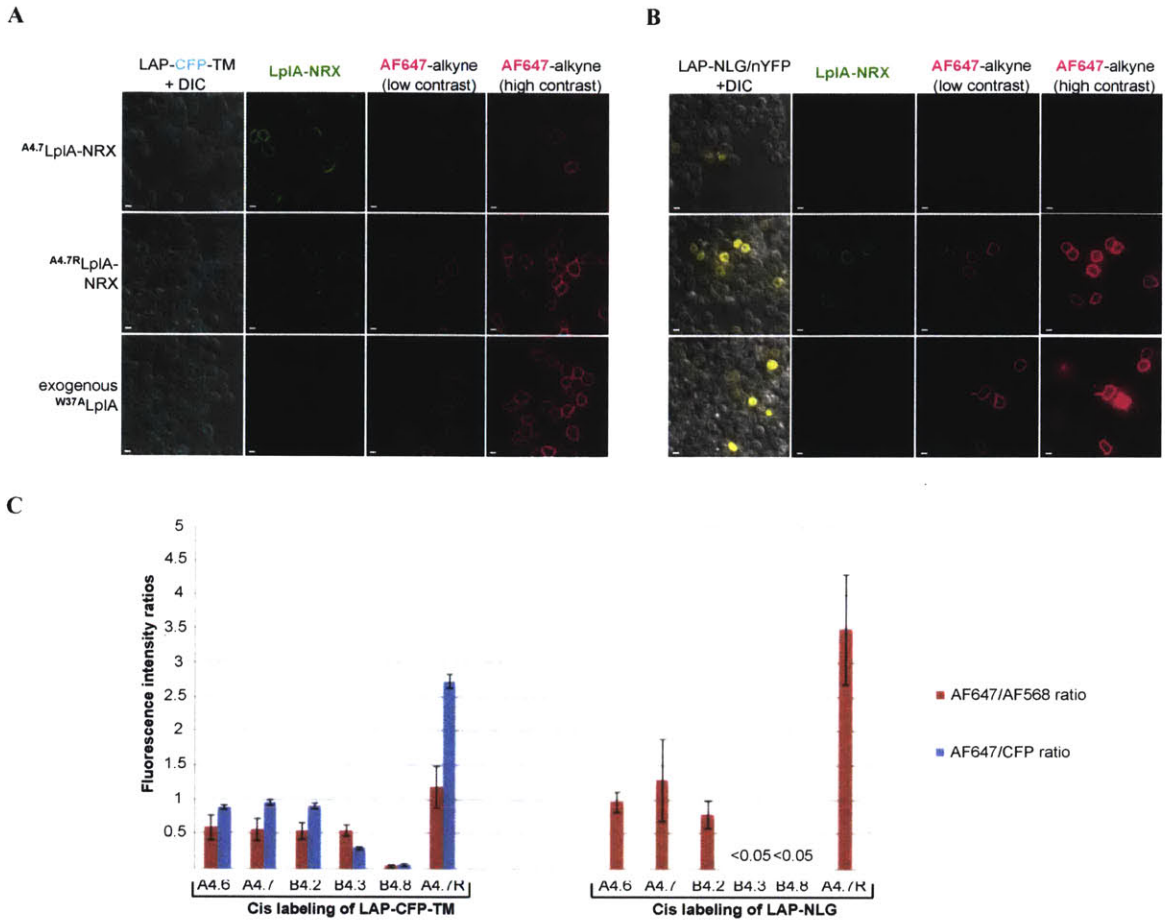


Figure 6-12. $^{A4.7R}$LplA produces cis pAz labeling signal equivalent to exogenous enzyme. (A) $^{A4.7R}$LplA was fused to the N-terminus of neurexin-1β (NRX) and co-expressed with LAP2-CFP-TM on the surface of HEK cells. Cis labeling was performed with addition of 100 μM pAz and ATP for 1 hour. The pAz-LAP2 conjugates were detected by [3 + 2] cycloaddition with AlexaFluor647-alkyne. LplA-NRX expression was detected by live-cell antibody staining with anti-FLAG antibody. Images of exogenous ^{W37A}LplA pAz labeling and cis pAz ligation using $^{A4.7}$LplA are shown for comparison. . **(B)** Cis pAz labeling performed identically to (A) while using LAP2-neuroligin-1 as the target substrate. **(C)** Quantification of pAz cis labeling extent on the mammalian cell surface for all five LplA mutants. Fluorescence intensities were calculated from confocal images for at least 30 individual cells for each condition. The labeling signal obtained with $^{A4.7R}$LplA greatly exceeds any of the ligases obtained from selections.

$^{A4.7R}$LplA can be used for PRIME labeling in the endoplasmic reticulum

Finally, we wanted to test the ability of the optimized pAz ligase ($^{A4.7R}$LplA) to label proteins when expressed in the endoplasmic reticulum (ER). To increase the intracellular probe concentration and improve labeling signal, a cell-permeable form of the pAz probe with an acetoxymethyl protecting group was used (pAz-AM). In HEK cells, the ligase was targeted to the ER using a KDEL-motif and co-expressed with

LAP2-neuroligin-1, which can be labeled by the ER enzyme as it is trafficked through the secretory pathway. After labeling in the ER, the labeled pAz-LAP2-neuroligin was detected by performing click at the cell surface (Figure 6-13, scheme). When using ER-targeted $^{A4.7R}$LplA, LAP2-neuroligin-1 was highly labeled (Figure 6-13). However, when the identical experiment was performed using $^{A4.7}$LplA, similar pAz labeling signal was observed. This result suggests that the addition of T57I may improve surface activity to a greater extent than ER activity, which is a hypothesis that we are continuing to explore. We also note that pAz labeling using exogenously applied enzyme is still 3- to 4-fold higher than when using $^{A4.7R}$LplA-ER or $^{A4.7}$LplA-ER, indicating that enzyme activity in the ER can still be improved further. Nevertheless, this was the first demonstration of PRIME labeling using an ER-targeted ligase.

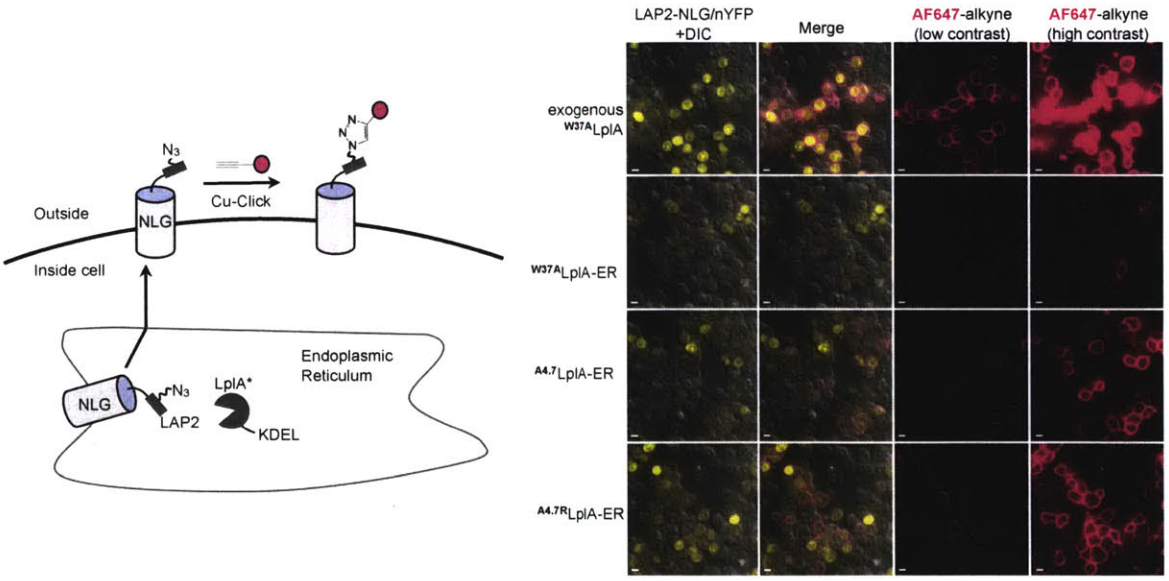


Figure 6-13. $^{A4.7R}$LplA has improved activity in the endoplasmic reticulum. HEK cells were co-transfected with the indicated ligase (retained in the ER via a KDEL tag) and LAP-neuroligin-1 (NLG). Membrane permeable pAz-AM was added to cells for 4 hours, and labeling of LAP-NLG was detected by Click at the cell surface with AF647-alkyne. For comparison (top row), labeling of LAP-NLG was performed with exogenous ^{W37A}LplA. Scale bars 10 μM.

In vitro characterization of evolved pAz ligases

Evolved ligases have improved activity in vitro

Although we have demonstrated that $^{A4.7R}$LplA gives the highest pAz labeling signal on the mammalian cell surface and in the endoplasmic reticulum, we wanted to

more thoroughly compare the activities of all evolved ligases *in vitro*. We first used high-performance liquid chromatography (HPLC) assays to monitor the extent of pAz ligation onto LAP2 peptide in a ten minute labeling protocol with either 200 μM or 50 μM pAz probe using the LplA variants. Table 6-2 summarizes the results of this assay using the five evolved ligases, $^{A4.7R}$LplA, and the best rationally designed pAz ligases: ^{W37V}LplA and ^{W37A}LplA. We also included the individual mutations of $^{A4.7}$LplA with the hope of deconvoluting the contributions of each mutation on activity. To investigate whether increases in enzyme activity were probe-specific, we also performed these assays using both lipoic acid and 7-hydroxycoumarin probe (Table 6-2).

We note that the trends for *in vitro* pAz ligation activity follow closely with the trends we observe on both the mammalian and yeast cell surface. All of the evolved ligases except for B4.3 give improved labeling over ^{W37A}LplA, the library starting-point. However, only A4.7 and A4.7R give improved pAz *in vitro* labeling compared to the best rationally designed ligase, ^{W37V}LplA. We also observed that $^{A4.7R}$LplA did not show improvement over $^{A4.7}$LplA at high pAz probe concentrations, and only minimal improvement at the lower pAz probe concentration (Table 6-2). This result contrasts with the mammalian cell surface data, where the addition of T57I drastically improved pAz labeling activity. Given these results, we hypothesize that the T57I mutation may be serving to stabilize or solubilize the enzyme when it is expressed on the cell surface, which is neither advantageous nor deleterious when working with purified enzymes. This conclusion is supported by the previous analysis of how T57I may contribute to improved activity (see Chapter 5). We also found that, for pAz ligation, it seems that the effects of F147L and H267R in combination are largely additive (Table 6-2).

	200 μM pAz4	50 μM pAz4	500 μM lipoic acid	500 μM coumarin
WT	N/A	N/A	65.7 ± 0.5	N/A
W37A	28.7 ± 1.7	9.6 ± 1.5	38.5 ± 1.2	3.9 ± 0.2
W37V	58.8 ± 2.7	20.9 ± 0.3	31.8 ± 2.8	18.3 ± 2.0
A4.6	44.9 ± 3.5	21.9 ± 2.1	46.3 ± 2.1	14.2 ± 0.7
A4.7	95.0 ± 3.7	26.8 ± 2.1	86.4 ± 3.2	10.5 ± 0.6
B4.2	55.75 ± 0.85	13.2 ± 0.6	65.7 ± 1.4	3.4 ± 0.6
B4.3	6.8 ± 0.50	1.2 ± 0.08	4.1 ± 0.2	0.2 ± 0.01
B4.8	37.3 ± 1.9	10.6 ± 2.0	15.2 ± 1.8	1.6 ± 0.1
A4.7R	93.14 ± 3.1	33.6 ± 0.5	88.1 ± 2.7	10.2 ± 0.3
W37A F147L	61.2 ± 8.6	16.7 ± 1.1	77.3 ± 4.1	1.9 ± 0.3
W37A H267R	68.9 ± 2.8	15.6 ± 0.9	88.1 ± 2.5	2.7 ± 0.2

Table 6-2. In vitro ligation activities of LplA variants with pAz, lipoic acid, and 7-hydroxycoumarin probes. Ligases were used at 2 μM concentration and with the indicated probe concentrations. All reaction times were 10 minutes. Extent of probe ligation onto 150 μM LAP2 peptide was monitored by HPLC. All measurements were performed in triplicate. Cell shading indicates relative activities, darker colors indicate higher relative activities. Errors, +/- 1 s.d.

$^{A4.7R}$LplA mutations may contribute in various ways to increased pAz ligation activity

We next sought to rationalize how the two evolved mutations (F147L and H267R) might be contributing to overall activity improvements of the enzyme. The F147L mutation has frequently appeared in previous LplA library selections, and always occurred when performing selections against alkyl azide probe ligation (see Chapter 5). We had previously hypothesized that this mutation may be involved in accommodating the bound azide structure or otherwise contributing to catalysis in a pAz probe-specific manner. This hypothesis is supported by the result that when F147L is introduced into wild-type LplA, it does not improve the k_{cat} for lipoylation (Figure 6-14). However, two results indicate that this hypothesis might be incorrect. First, when we docked the pAz-AMP structure into the lipoyl-AMP binding pocket of LplA[15] the F147L mutation is not in close proximity to the azide moiety (Figure 6-15), which places in doubt the direct involvement of F147L in pAz4 binding or recognition. Second, in contrast to the result obtained when F147L was added to ^{WT}LplA, we observed significantly improved lipoylation activity when F147L was combined with the W37A mutation (Table 6-2).

This result indicates that F147L may be a potentiating mutation that serves to improve enzyme activity only when Trp37 is mutated. Future work exploring this issue will involve analysis of the alkyl azide probe ligation activity of F147L mutants with and without mutation at the Trp37 residue.

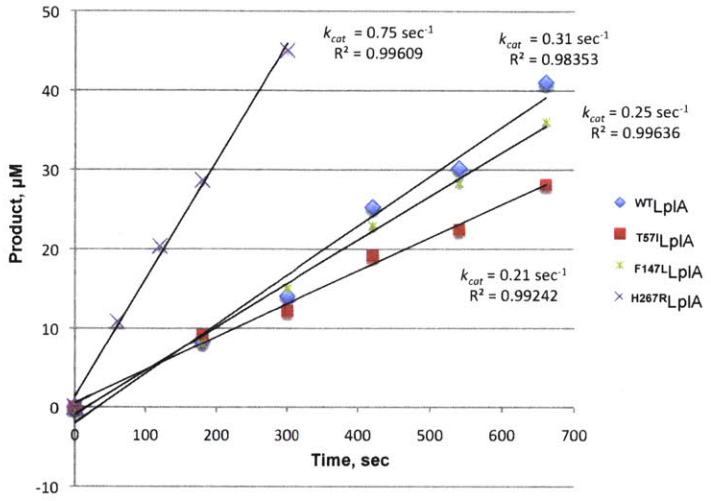


Figure 6-14. Lipoylation kinetics when mutations of $^{A4.7R}$LplA are added individually to ^{WT}LplA. Ligation reactions were assembled with 400 µM lipoic acid, 400 µM LAP2 peptide, and 200 nM of the indicated enzyme. Product amounts were quantified by HPLC. We estimate that these reaction conditions are saturating for all three LplA substrates, such that the slopes of these lines represent V_{max}, and from these values k_{cat} can be determined. Reactions were performed in triplicate; errors, +/- 1 s.d.

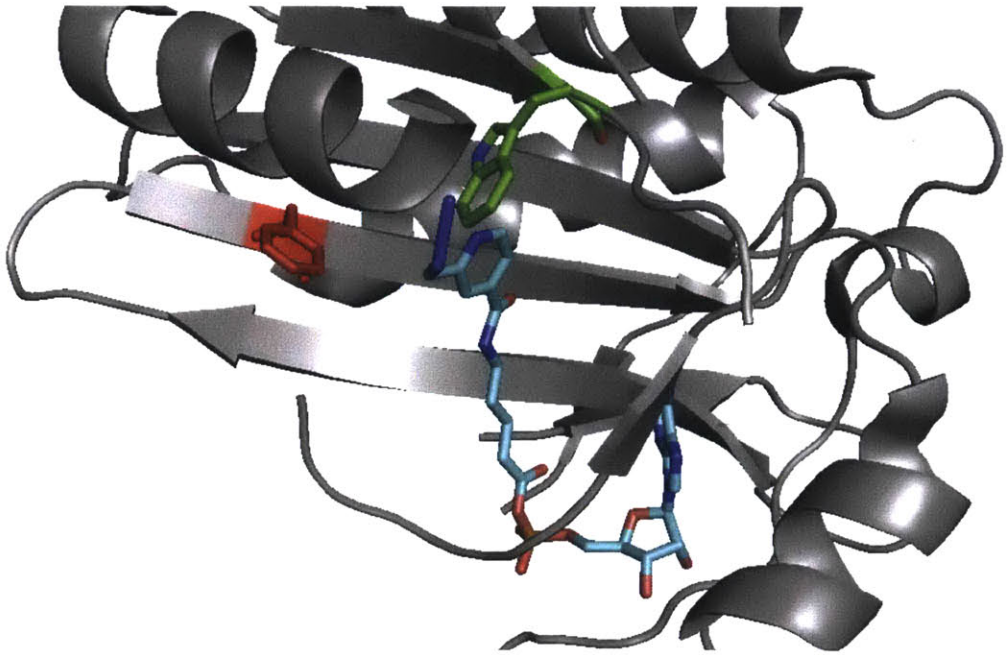


Figure 6-15. F147 is probably not directly involved in pAz binding. Docked structure of pAz-AMP in lipoyl-AMP binding pocket of LplA (PDB 3A7R). W37 position is shown in green stick and F147 position is shown in red stick.

Unlike F147L, the H267R mutation seems to generally increase enzyme activity independent of the small molecule probe being used (Table 6-2). His267 is located at the interface between the N- and C-terminal domains of the apo LplA structure[16] (Figure 6-16A). As discussed in Chapter 5, LplA undergoes a global conformational change after formation of the lipoyl-AMP and before transfer to the protein substrate[15] (Figure 6-16B). Upon formation of the adenylate ester, the C-terminal domain of LplA flips approximately 180 degrees, adopting a stretched conformation and exposing the site of protein substrate binding. We propose that the mutation of His267 to a bulky arginine residue is helping to promote this global conformational change by pressing the C-terminal domain away from the N-terminal domain and pre-ordering the open conformation of the enzyme[15] (Figure 6-16C). This hypothesis is supported by the fact that when the H267R mutation is introduced into the wild-type LplA, we notice a 2-fold improvement in the k_{cat} of the lipoylation reaction (Figure 6-14). We note that the other individual mutations of $^{A4.7R}$LplA did not produce this increase in lipoylation activity when introduced into the wild-type ligase (Figure 6-14). This suggests that the contributions of these mutations to improved pAz ligation activity may be the result of probe-specific effects or require the cooperation of the W37A mutation (as we suspect with F147L). It is also possible that the rate enhancement effects are only are evident in the context of cell surface expression and stability and are thus less apparent when performing *in vitro* reactions (as we suspect with T57I).

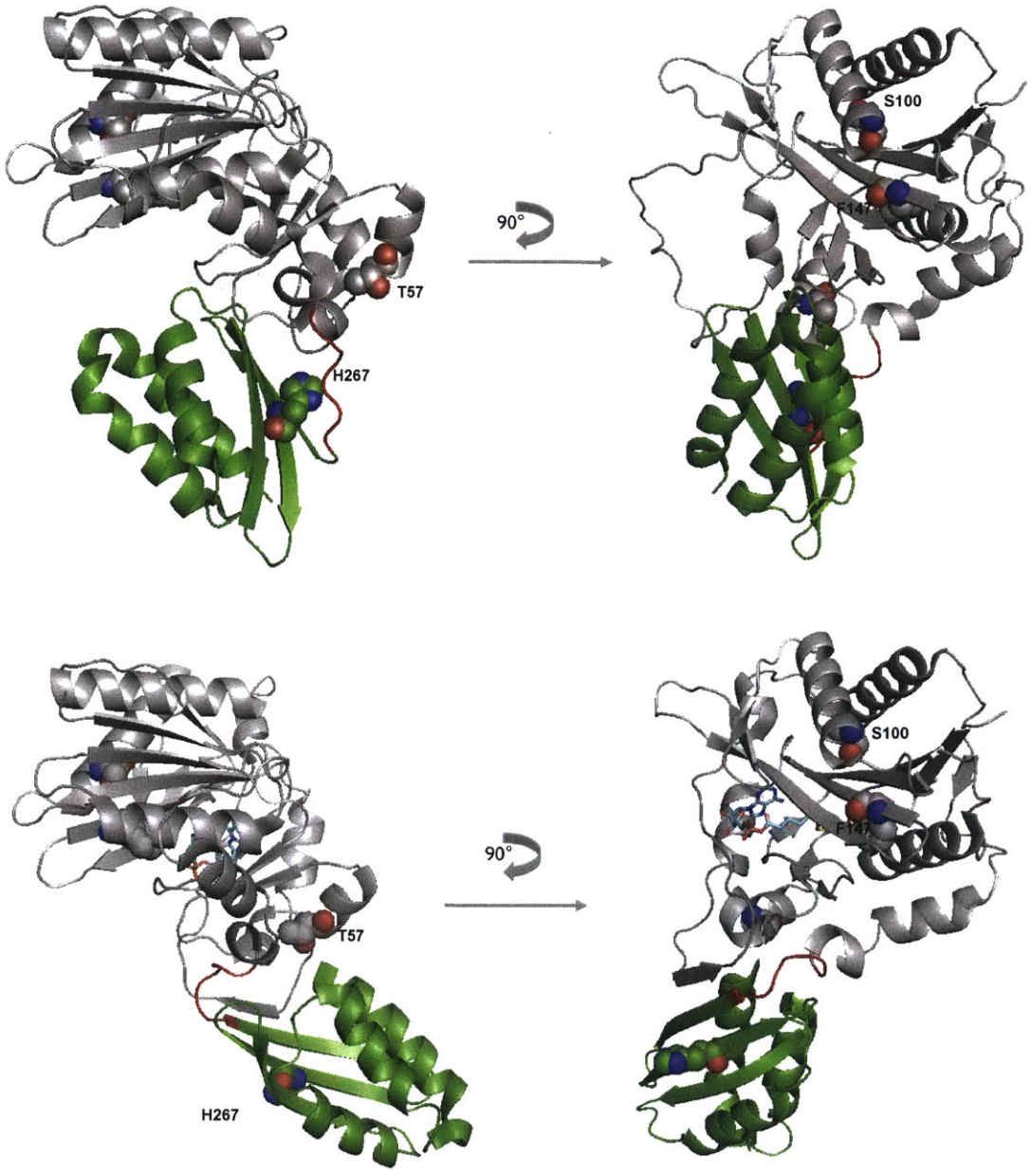


Figure 6-16. H267R may promote conformational change upon lipoyl-AMP formation and pre-order the activated complex. (A) Apo *E. coli* LplA (PDB 1X2G) **(B)** Lipoyl-AMP bound *E. coli* LplA (PDB 3A7R). N-terminal domain and C-terminal domain are in grey and green, respectively. LplA undergoes a global conformational change upon lipoyl-AMP formation (cyan stick). In both proteins the amino acids of $^{A4.7R}$LplA are shown in spheres. All structure views created using PyMol.

In vitro characterization of optimized pAz ligase

$^{A4.7R}$LplA covalently ligates pAz onto LAP2 site-specifically

We further characterized the optimized pAz ligase ($^{A4.7R}$LplA) *in vitro* with specificity controls. Figure 6-17A shows HPLC traces from 10 minute *in vitro* LAP2 labeling reactions using $^{A4.7R}$LplA, ^{W37V}LplA, and ^{W37A}LplA. This result clearly shows that $^{A4.7R}$LplA is more active than the best rationally designed ligases under identical labeling conditions. A control experiment where ATP is omitted demonstrates that pAz ligation using $^{A4.7R}$LplA is ATP-dependent. Finally, the identity of the pAz-LAP2 conjugate was confirmed by mass spectrometry (Figure 6-17B).

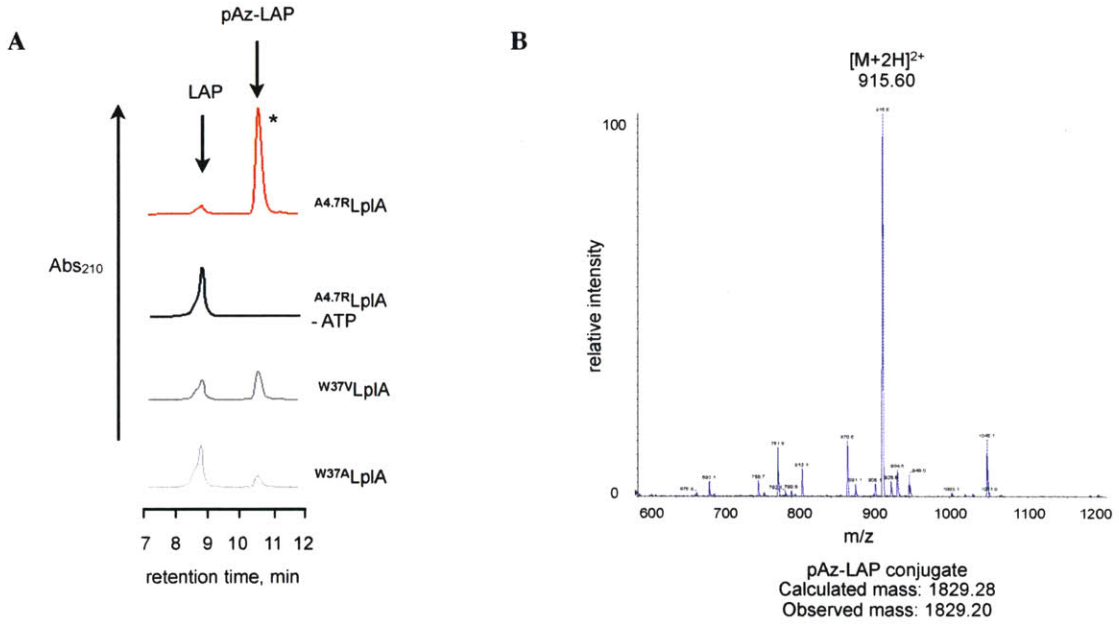


Figure 6-17. $^{A4.7R}$LplA *in vitro* characterization and specificity controls. (A) HPLC traces show conversion of LAP2 peptide to picolyl azide-LAP2 conjugate by $^{A4.7R}$LplA, ^{W37A}LplA and ^{W37V}LplA. A negative control, with ATP omitted, is shown in black. **(B)** Starred peak in (A) was analyzed by mass spectrometry and confirmed to be the covalent pAz-LAP2 conjugate.

Next, the sequence specificity of $^{A4.7R}$LplA was confirmed by performing pAz ligation on mammalian cell lysate. To demonstrate that the ligase obtained from yeast-display selections is also sequence-specific, we performed this experiment in parallel using $^{A4.7}$LplA. We prepared lysate from HEK cells expressing LAP fused to yellow fluorescent protein (LAP2-YFP) and performed pAz labeling using either $^{A4.7R}$LplA or

$^{A4.7}$LplA followed by detection using biotin-alkyne and western blot with streptavidin-HRP (Figure 6-18). While lanes 1 and 4 demonstrate specific labeling of LAP2-YFP, control lanes using untransfected lysate or lysate expressing a LAP2 mutant where the target lysine is mutated to an alanine do not show labeling. This result demonstrates that the optimized pAz ligase and the best pAz ligase obtained by yeast-display evolution both retain the high sequence-specificity of LplA.

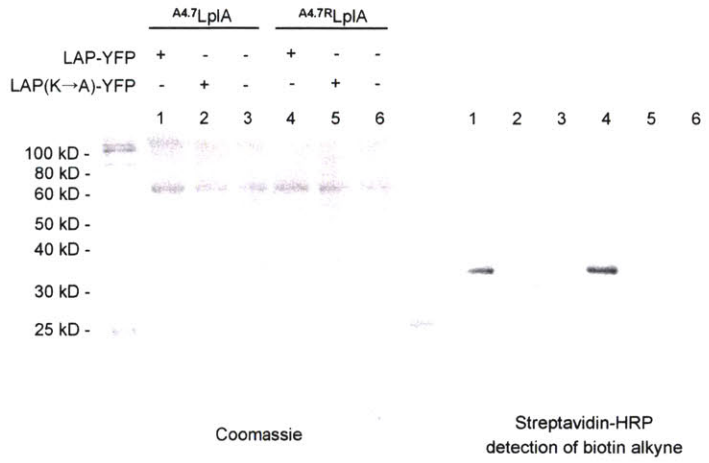


Figure 6-18. $^{A4.7R}$LplA and $^{A4.7}$LplA retain the high sequence specificity of the original ligase. HEK cells expressing LAP2-YFP were lysed and labeled with either $^{A4.7}$LplA or $^{A4.7R}$LplA. Ligated pAz was detected using biotin-alkyne followed by western blotting with streptavidin-HRP. Controls are shown using untransfected lysate or lysate prepared from HEK cells expressing LAP2(K→A)-YFP.

$^{A4.7R}$LplA has a faster k_{cat} compared to rationally designed pAz ligases

We next used HPLC to measure the kinetics of pAz ligation onto LAP2. For $^{A4.7R}$LplA we determined the k_{cat} to be 0.30 sec^{-1}, while for $^{A4.7}$LplA we calculated a k_{cat} of 0.41 sec^{-1} (Figure 6-19). These values can be compared to those calculated for ^{W37A}LplA and ^{W37V}LplA (k_{cat} 0.09 and 0.21 sec^{-1}, respectively). The k_{cat} values for pAz ligation using the evolved and optimized ligases are ~2 to 4-fold higher than for the best rationally designed ligases. This result confirms that the evolutionary pressure we applied during our selections was successful in enriching for ligases with improved catalytic activities.

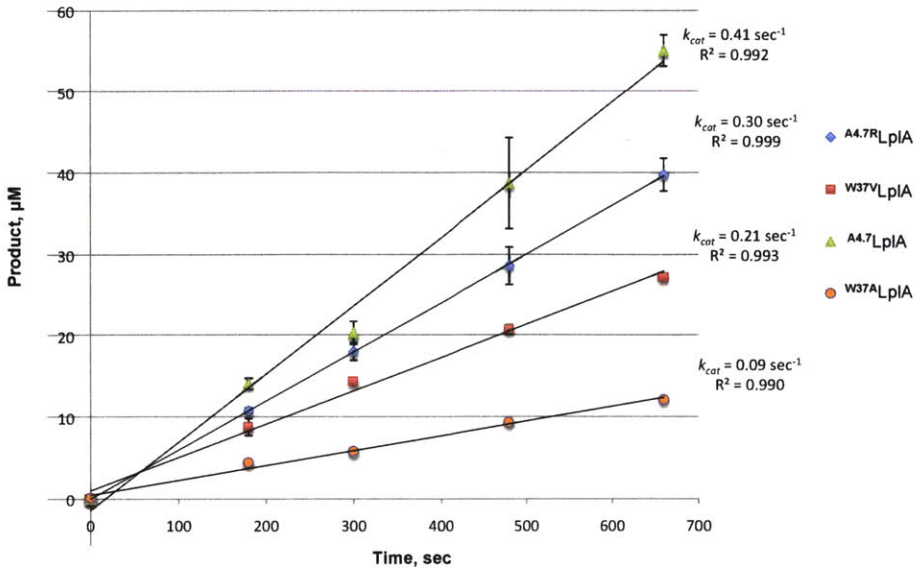


Figure 6-19. $^{A4.7R}$LplA has improved pAz ligation kinetics compared to rationally designed ligases. Ligation reactions were assembled with 400 µM LAP2 peptide, 400 µM pAz probe, 1 mM ATP, and 200 nM of the indicated enzyme. Product amounts were quantified by HPLC. Reactions were performed in triplicate; errors, +/- 1 s.d. Assuming reaction conditions are saturating for all LplA substrates, the slopes of the lines give V_{max} and k_{cat} was calculated for each enzyme.

Application to labeling protein-protein interactions

$^{A4.7R}$LplA can be used to visualize intercellular protein-protein interactions in HEK cells

We have demonstrated in this chapter that the optimized pAz ligase produces better pAz labeling *in vitro*, in the endoplasmic reticulum, and in cis on the mammalian cell surface. We next wanted to apply $^{A4.7R}$LplA to the labeling of inter-cellular contacts and explore the potential for using the evolved ligase to extend ID-PRIME and allow for the labeling of intercellular protein-protein interactions.

First, we developed a trans pAz labeling assay to test the ability of $^{A4.7}$LplA to label intercellular contacts between HEK cells. In this assay, we transfect one pool of HEK cells with $^{A4.7R}$LplA-neurexin-1β and a separate pool of HEK cells with LAP2-neruoligin-1. After transfection, we lift the two populations and then mix and co-plate them on the same coverslip. If we can successfully label inter-cellular protein-protein contacts, when the two populations of HEK cells come into contact should result in labeling of the LAP2-neuroligin by the $^{A4.7R}$LplA-neurexin (Figure 6-20, scheme). While the goal is to ultimately label neurexin and neuroligin interactions at a synapse, where

this interaction is biologically relevant, intercellular labeling in HEK cells serves as an important proof-of-principle.

Figure 6-20 shows that we are able to label the inter-cellular contact of these two proteins with just a 5-minute pAz labeling protocol in HEK cells. Controls where the experiment was performed identically but using either ^{W37A}LplA or the alanine mutant of LAP2 demonstrate the specificity of the inter-cellular labeling. Interestingly, when performing trans pAz labeling using $^{A4.7R}$LplA, the labeling signal observed after the 5-minute labeling protocol is brighter than that observed when using exogenously applied enzyme.

Encouraged by this result, we next wanted to determine whether the labeling signal we observe in the trans assay is dependent on the interaction between the neurexin and neuroligin. To test this, we performed the trans labeling experiment using a fusion of $^{A4.7R}$LplA to an interaction-deficient mutant of neurexin-1β (^{D137A}NRX) (Figure 6-20). Although some intercellular contacts were labeled using the interaction deficient version of neurexin, the labeling signal never exceeded 25% of what we observed in the presence of protein-protein interaction. This control demonstrated that labeling is largely interaction dependent. We propose that with some further optimization this method may be applicable to the imaging of biologically relevant protein-protein interactions on the surface of living mammalian cells or even perhaps at the neuronal synapse.

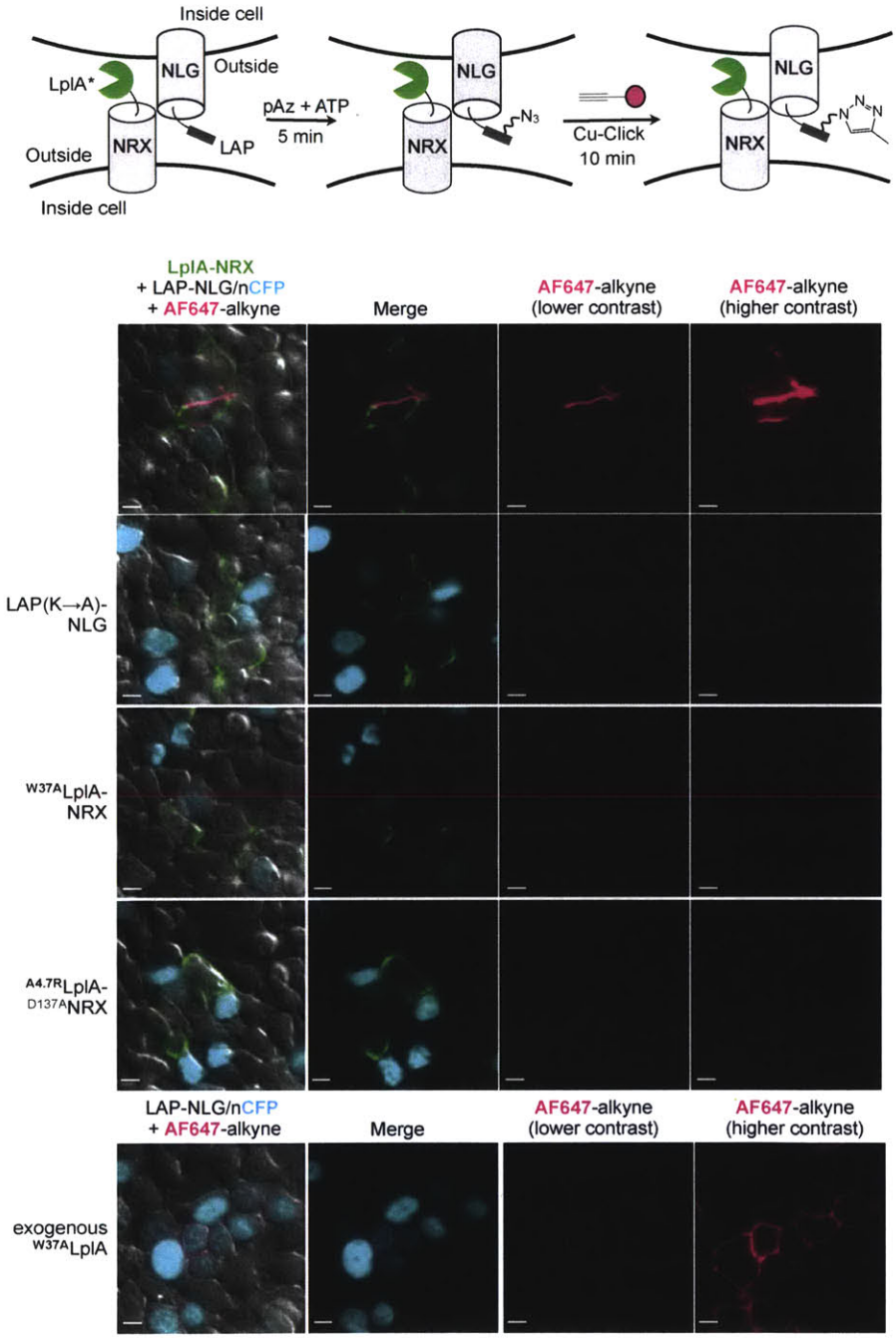


Figure 6-20. Imaging neurexin and neuroligin interactions using trans pAz labeling. HEK cells expressing $^{A4.7R}$LplA as a fusion to neurexin-1β (NRX) were co-plated with HEK cells expressing LAP2 as a fusion to neuroligin-1 (NLG, with a nuclear CFP co-transfection marker). When the two populations of cells come into contact, $^{A4.7R}$LplA can ligate pAz onto LAP2-NLG in just 5 minutes. As controls, trans labeling fails when using LAP2(K→A)-NLG or ^{W37A}LplA-NRX (second and third rows). When $^{A4.7R}$LplA is fused to an interaction-deficient mutant of neurexin-1b, trans labeling is inhibited (fourth row). Labeling signal can be compared to that observed using exogenous ^{W37A}LplA for 5 minutes (bottom row).

Comparison of GRASP methodology and trans pAz ligation using $^{A4.7R}$LplA

We were eager to compare our method to GRASP methodology (GFP Reconstitution Across Synaptic Partners), another method that can image protein-protein interactions in live cells. This method uses a split super-folder GFP (sfGFP) to image protein-protein contacts across cells or synapses and has been demonstrated in tissue culture and even transgenic mice[17,18]. A large portion of sfGFP is fused to one interacting partner and the remaining smaller portion of sfGFP is fused to the other interacting partner. When expressed separately, the two proteins are non-fluorescent. Upon co-localization of the protein interaction partners, the sfGFP is reconstituted and becomes fluorescent after chromophore maturation.

We first tried the trans experiment using the exact GRASP constructs described in the work of Kim and colleagues, where the larger piece of sfGFP is fused to CD4 and the smaller piece of sfGFP is fused to neuroligin-1[18]. These constructs were designed as a pair that would report on synapse formation as determined by spatial proximity. Unfortunately, we weren't able to observe GFP complementation in a trans assay in HEK cells, despite observing GFP complementation when the two pieces were expressed in the same cell (data not shown). We hypothesized that when these constructs are expressed in neurons, scaffolding proteins and endogenous protein assemblies force the inter-cellular interaction of CD4 and neuroligin, which in turn promotes sfGFP complementation. In support of this hypothesis, when the sfGFP pieces were fused to neurexin-1β and neuroligin-1, which we have already demonstrated interact well across HEK cells, we observed good GFP complementation in the trans assay (Figure 6-21). Moreover, when we used an interaction-deficient form of neurexin-1β we observed a significant decrease in number of GFP complementation events and a decrease in the GFP fluorescence intensity at those events. These two results, taken together, suggest that the GRASP method may be interaction-dependent in some contexts.

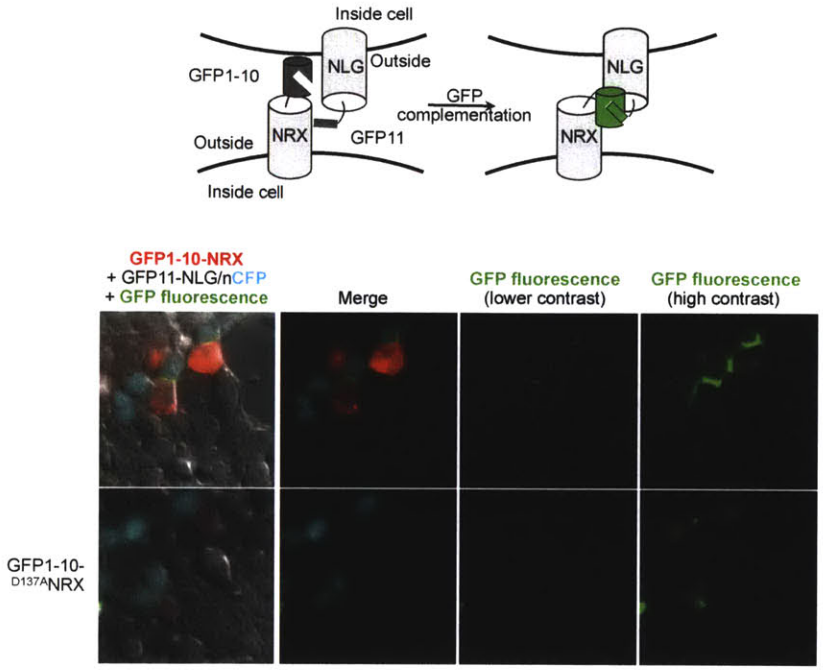


Figure 6-21. Imaging of neurexin and neuroligin interactions using GRASP. HEK cells expressing sfGFP1-10 fused to neurexin-1β (NRX, with mCherry co-transfection marker) were co-plated with HEK cells expressing sfGFP11 fused to neuroligin-1 (NLG, with CFP co-transfection marker). Where the two populations of cells come into contact, GFP can reconstitute between the cells (top row). When a mutant of neurexin that is incapable of interacting with neuroligin is used (^{D137A}NRX), GRASP signal decreases dramatically (bottom row).

To get a direct comparison of the GRASP method and our pAz labeling method, the labeling signal to noise ratio was determined for each protocol (additional fields of view for both experiments can be found in Figure 6-22). We found that the trans pAz ligation method gives a signal to noise ratio of 26, while the GRASP method produces a signal to noise ratio of 11. This indicates that the trans pAz labeling technique is more sensitive, producing higher labeling signal on average. An additional advantage of our method is that while GFP complementation is trapping, our label is not and persists after interaction. We envision that, using our method, a protein-protein interaction could be detected and then time-lapse imaging can be used to monitor subsequent trafficking of the labeled interacting partner.

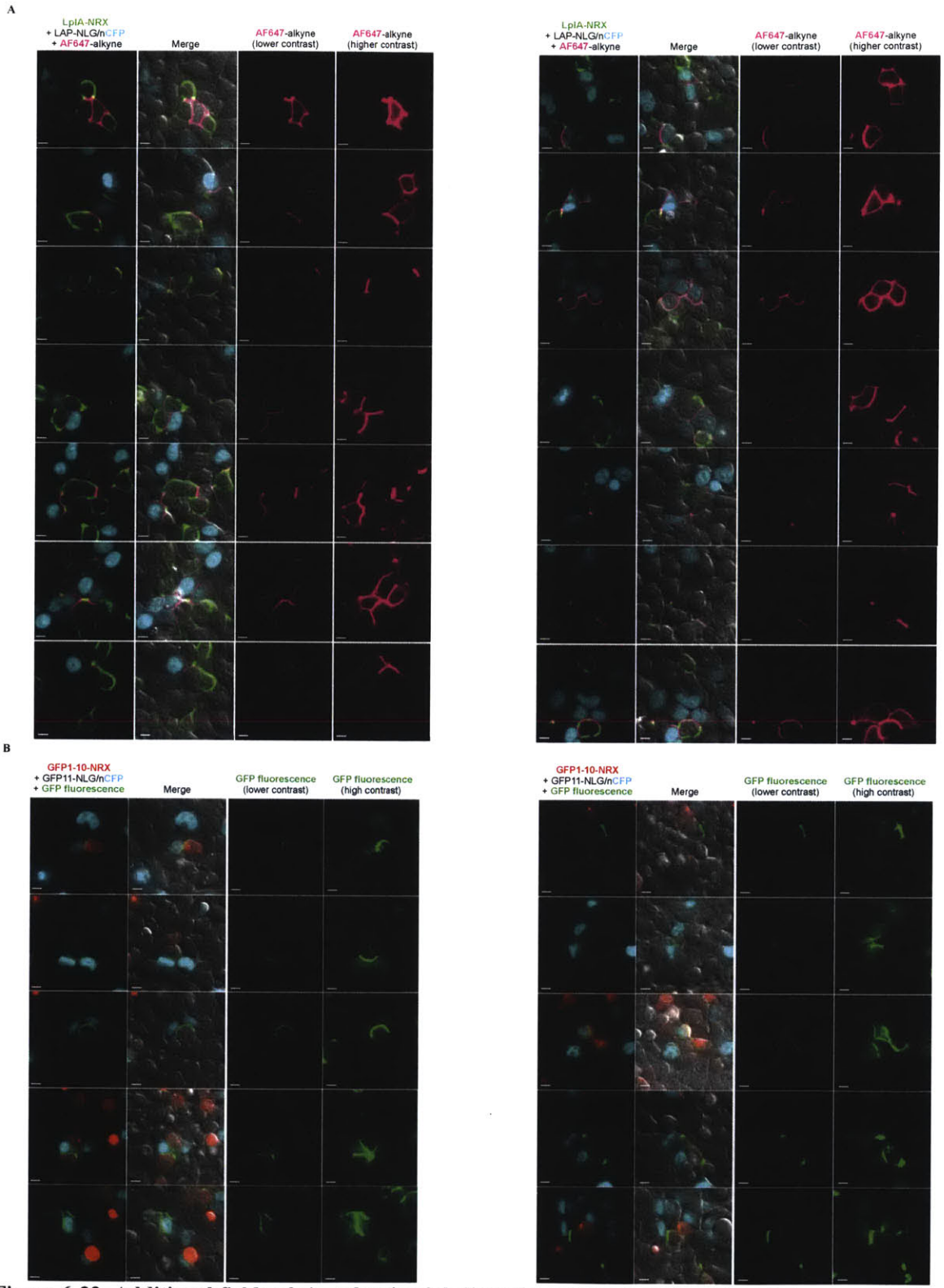


Figure 6-22. Additional fields of view showing labeling of intercellular interactions between neurexin and neuroligin. (A) Trans pAz labeling of LAP2-neuroligin using $^{A4.7R}$LplA-neurexin-1β. From the same experiment as Figure 6-20. Scale bars, 10 μM. **(B)** Trans GRASP labeling using GFP11-neuroligin and GFP1-10-neurexin-1β. From same experiment as figure 6-21. Scale bars, 10 μM.

Conclusion

Using yeast-display evolution of *E. coli* lipoic acid ligase, we have engineered a PRIME ligase ($^{A4.7R}$LplA) with improved picolyl azide (pAz) labeling properties. We have demonstrated the improved pAz ligation activity of this ligase *in vitro*, and also report the first demonstrations of PRIME labeling using enzyme expressed on the surface and in the endoplasmic reticulum of living cells. While there have been other recent examples using yeast-display to evolve improved enzyme function[13,19], to our knowledge this is the first example of the evolution of activity for an application where there was no observable activity using the original enzyme.

The expansion of the PRIME labeling methodologies to the endoplasmic reticulum designates PRIME as one of the most versatile labeling methods currently available. While other protein-based methods such as SNAP/CLIP[20] and HaloTag[21] are applicable to oxidizing compartments of the cell, the large size of these tags can interfere with target protein function. With the developments described in this chapter, the PRIME method now uniquely combines a peptide tag, high labeling specificity, and the ability to label target proteins in both reducing and oxidizing compartments of the cell. Moreover, the wide variety of fluorophores and probes that can be incorporated using PRIME make it an attractive method for a wide range of biological applications.

We have further demonstrated the application of this ligase to the labeling of intercellular protein-protein interactions. We were able to label the interaction of neurexin-1β and neuroligin-1 in HEK cells in just 5 minutes. This result is the first demonstration of interaction-dependent PRIME (ID-PRIME) labeling of cell-surface protein-protein interactions. Compared to the GRASP method for labeling synaptic partners, our method gives better signal and produces a label that is non-trapping. We hope to soon extend this trans pAz4 labeling method to the imaging of synaptic partners in live neurons.

Experimental

General yeast display methods

Growth and induction. Unless otherwise noted, cells were grown in SDCAA growth media [2% dextrose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.86% Monosodium phosphate] at 30°C with shaking for 20 hours. To induce protein expression, the cultures were diluted to a cell density of ~10^6 cells/ml in SGCAA media (identical to SDCAA above but with 2% galactose as the carbon source). Induction in SGCAA was performed at 30°C for 20 hours.

Transformation. For single clones, LAP2-YIP (see below) yeast cells were transformed with DNA using the Frozen-EZ Yeast Transformation II kit (Zymo Research). The transformed yeast cells were plated on selective plates (SDCAA –W, –U) and single yeast colonies were selected for amplification. For the libraries, LAP2-YIP yeast cells were electroporated with the error-prone PCR library (see library construction below) using a BioRad GenePulser Xcell.

Long-term storage. For individual clones, glycerol stocks were made. Isolated colonies were grown in 5 mL SDCAA overnight at 30°C. 812 μL of saturated overnight yeast culture was combined with 188 μL of 80% glycerol (sterile) and placed into a 2 mL cryo-vial, stored at -80°C. For libraries and rounds, frozen stocks were made. At least 10^{10} freshly grown cells from each library were inoculated into 1 L low-dextrose SDCAA [as above but with 0.5% dextrose] and grown at 30°C for approximately 3 days. Cells were pelleted and then resuspended in freezing solution (2% glycerol, 0.67% yeast nitrogen base) to a final concentration of ~6 x 10^{10}/mL. Cells were divided into 1 mL aliquots in cryovials. The tubes were then put into a Mr. Frosty (Nalgene) and allowed to slow-freeze overnight -80°C. The next day, cells were moved to liquid nitrogen for long-term storage.

FACS instrumentation. All fluorescence activated cell sorting was performed on a BD

FACSAria (BD Biosciences). For all analysis where sorting was not required, a BD LSRII-HTS (BD Biosciences) was used. Both of these instruments are located in the Flow Cytometry Core Facility at the Koch Institute for Integrative Cancer Research at MIT.

Preparation of yeast cells constitutively expressing LAP2-Aga1p

The LAP2-Aga1p yeast integrating plasmid was cloned from the YIPlac211-GPD-Avitag-Aga1p plasmid, which was a gift from Irwin Chen at Harvard University. First, two restriction sites, *AflII/NheI* were introduced on either side of the Avitag to create YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. The LAP2 peptide was then introduced by pre-annealing LAP2 primers with *AflII/NheI* sticky ends and ligating into *AflII/NheI* digested YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. This produced the yeast integrating plasmid YIPlac211-GPD-*AflII*-LAP2-*NheI*-Aga1p.

To integrate the plasmid into the yeast genome, YIPlac211-GPD-*AflII*-LAP2-*NheI*-Aga1p was linearized by digestion with *BsiWI* and transformed into *Saccharomyces cerevisiae* strain *TM134* using the transformation protocol described above and plated on SDCAA plates lacking uracil. Colonies were picked and grown in SDCAA lacking uracil. A yeast colony with the LAP2-Aga1p construct correctly inserted and with good expression was designated "LAP2-YIP" yeast and displays the LAP2 peptide sequence constitutively on the cell surface as a fusion to the N-terminus of Aga1p.

Lipoylation on Yeast cell-surface (Figure 6-4**)**

Yeast cells constitutively displaying the LAP2 peptide and inducibly displaying *E. Coli* LplA or ^{W37A}LplA (see General yeast methods for LplA induction details) were pelleted and washed twice with phosphate-buffered saline (pH 7.5) with 1mg/ml bovine serum albumin (PBS-B). Cells were resuspended to a cell density of 2 x 10^8 cells/ml in PBS-B with 500 μM lipoic acid, 5 mM $Mg(OAc)_2$ and 3 mM ATP and incubated on a rotator at 30°C for 1-3 hours. The reactions were stopped by pelleting the cells and washing with ice-cold PBS-B containing 100 mM EDTA. Cells were washed an additional time with ice-cold PBS-B. The ligated lipoic acid was detected using rabbit

polyclonal anti-lipoic acid antibody (Calbiochem) and LplA surface-expression level was detected using mouse monoclonal anti-HA antibody (Roche) in PBS-B at 1:200 dilutions for 40 minutes at 4°C. Cells were pelleted and washed twice with ice-cold PBS-B. Then the yeast cells were incubated with anti-rabbit R-phycoerythrin (Invitrogen) and anti-mouse AlexaFluor-488 secondary antibodies (AF488; Invitrogen) in PBS-B at 1:100 dilutions for 40 minutes on a rotator at 4°C. Cells were pelleted and washed twice with ice-cold PBS-B then resuspended in PBS-B to a density of ~1 x 10^7 for FACS.

Model Selection (Figure 6-5**)**

LAP2-YIP yeast displaying wild-type LplA or the less-active ^{W37A}LplA mutant were mixed in ratios of 1:100 and 1:10 wt:W37A and labeled for 3 hours as described above in **Lipoylation on Yeast cell surface**. After incubation with fluorophore-conjugated antibodies, 10^7 cells from each mixture were sorted. The top 0.1% and top 1.0% of the PE/AlexaFluor488 double positive population for the 1:100 and 1:10 experiments, respectively, were collected. Collected cells were cultured until saturation in SDCAA growth media with 50 μg/mL penicillin, 25 μg/mL kanamycin, and 50 μg/mL streptomycin. Plasmid DNA was harvested using the Zymoprep Yeast Plasmid Miniprep kit (Zymo Research), and the recovered LplA genes were amplified using the primers PctAmp-F 5'-CGTTCCAGACTACGCTCTGCAG and PctAmp R 5'-CGAGATCTGATAACAACAGTGTAG and sent for sequencing analysis. Pre-and post-sort yeast were re-propagated and labeled again as described above and analyzed by FACS for qualitative comparison. Quantitative analysis of enrichment from pre- to post-sort was determined by analyzing nucleotide intensities on the sequencing chromatogram.

Creating LplA libraries A and B

The round zero (R0) LplA libraries were cloned into LAP2-YIP *S. cerevisiae* using gap repair homologous recombination.

For Library A, the PCR reaction contained 2 μM 8-oxo-2'deoxyguanosine (8-oxo-dGTP), 2 μM 4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP), 200 μM each dNTP, and 0.2 μM each of primers 1F and 1R and reactions were thermocycled 10 times.

For Library B, the PCR reaction was assembled as above but with 200 μM dNTP analogues and was thermocycled 5 times.

For Library C, the PCR reaction was assembled as above but with 20 μM dNTP analogues and was thermocycled 10 times.

For Library D, the PCR reaction was assembled as above but with 20 μM dNTP analogues and was thermocycled 20 times.

For both Library A and B, mutagenized genes were further amplified in PCR reactions without mutagenic dNTP analogs using primers: 1F (3'- caaggtctgcaggctagtggtggaggaggctctggtggaggcggtagcggaggc-5') 1R (3'-ctacactgttgttatcagatctcgagctattacaagtcctcttcagaaataagcttttgttc-5').

The pCTCON2-Aga2p vector was prepared by digestion with *NheI* and *BamHI*, and gel-purified. The PCR-amplified inserts and digested pCTCON2 vector were mixed in 1:3 ratios and transformed together into LAP2-YIP *S. cerevisiae* by electroporation as described previously[22]. Homologous recombination occurred inside the yeast, giving the desired product. Serial dilutions of transformed yeast were plated on SDCAA plates and colonies were counted, to determine transformation efficiencies for each library. A total of ~10^9 cells from the fully grown library cultures were pelleted and induced as described above for the first round of selection.

pAz labeling and selections on yeast cell surface (Figure 6-7, 6-8, 6-9, 6-10**)**

Yeast cells constitutively displaying the LAP2 peptide and either the induced LplA libraries or individual clones (see General yeast methods for LplA induction details) were pelleted and washed twice with phosphate-buffered saline (pH 7.5) with 1mg/ml bovine serum albumin (PBS-B). Cells were resuspended to a cell density of 2 x 10^8 cells/ml in PBS-B with 100-200 μM picolyl azide (pAz), 5 mM $Mg(OAc)_2$ and 3 mM ATP and incubated on a rotator at 30°C for 1–12 hours (see Figure 2A and Supporting Figure x for exact library selection conditions). The reactions were stopped by pelleting the cells and washing with ice-cold PBS-B containing 100 mM EDTA. Cells were washed twice with ice-cold PBS to remove residual BSA, which inhibits the subsequent click reaction.

The ligated pAz was detected by performing Cu-catalyzed click chemistry with biotin-alkyne. The click reagents were prepared as follows: 500 μM THPTA, 100 μM CuSO4, 2.5 mM sodium ascorbate were incubated at room temperature for 10 minutes then diluted with PBS to appropriate volume for the given final concentrations. Biotin-alkyne was then added to a final concentration of 4 μM. The labeled yeast cells were incubated with the click reagents at room temperature on a rotator for 10 minutes, then pelleted and washed twice with PBS-B.

The biotin was detected using phycoerythrin-conjugated streptavidin (SA-PE; Jackson Immunoresearch) in PBS-B at 1:50 dilution. The LplA surface-expression level was detected using mouse monoclonal anti-HA antibody (Roche) in PBS-B at 1:200 dilution. Incubation with SA-PE and anti-HA occurred simultaneously at 4°C for 40 minutes. Cells were pelleted and washed twice with ice-cold PBS-B. Then the yeast cells were incubated with anti-mouse AlexaFluor-488 secondary antibody (AF488; Invitrogen) in PBS-B at 1:100 dilution for 40 minutes at 4°C. Cells were pelleted and washed twice with ice-cold PBS-B then resuspended in PBS-B to a density of ~1 x 10^7 /mL for FACS.

For the libraries, sorted cells were cultured until saturation in SDCAA growth media with 50 μg/mL penicillin, 25 μg/mL kanamycin, and 50 μg/mL streptomycin. After the culture was saturated, ~10^9 yeast cells were pelleted and induced in SGCAA and prepared for the next round of selection.

Analysis of Yeast Pools from Individual selection rounds (Figure 6-9)

Yeast harvested from each round of selection were amplified and induced as described above. All pools were then treated identically with 100 μM pAz, 5 mM magnesium acetate, and 3 mM ATP for 1 hour at 30°C. Rounds 3 and 4 were also labeled under identical concentrations for only 30 minutes at 30°C to distinguish between the activities of these rounds.

To sequence individual clones, yeast were plated on SDCAA plates, single colonies were amplified in SDCAA media, and plasmid was isolated using the Zymoprep Yeast Plasmid Miniprep kit (Zymo Research). To increase DNA concentration, LplA genes were PCR-amplified from plasmid using the primers PctAmp-F and PctAmp-R

(sequences under Model Selections) and sent for sequencing analysis.

General mammalian cell culture methods

Human embryonic kidney (HEK) cells were cultured in Minimum Essential Medium (MEM; Cellgro) supplemented with 10% v/v fetal bovine serum (FBS; PAA Laboratories). All cells were maintained at 37°C and under 5% CO_2. For imaging, cells were plated as a monolayer on glass cover slips. Adherence of HEK cells was promoted by pre-coating the coverslip with 50 μg/ml fibronectin (Millipore).

Fluorescence imaging

Cells were imaged in Dulbecco's Phosphate Buffered Saline (DPBS) in confocal mode. All images shown are using a Zeiss Axiovert 200M inverted microscope with a 40x oil-immersion objective. This instrument is equipped 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 (DPSS) lasers (all 50 mW). CFP (405 laser excitation, 445/50 emission), YFP and AlexaFluor-488 (491 laser excitation, 528/38 mission), AlexaFluor-568 (561 laser excitation, 617/73 emission), AlexaFluor-647 (630/20 excitation, 680/30 emission).

Cis pAz ligation on mammalian cell surface (Figure 6-11, 6-12)

HEK cells were transfected at ~70% confluency with expression plasmids for FLAG-^{W37A}LplA-neurexin1β (or variant) (600 ng for a 0.95 cm2 dish) and the LAP fusion protein of interest (600 ng) using Lipofectamine 2000 (Invitrogen). Where indicated, a nuclear co-transfection marker (H2B-YFP) was used (50-100 ng). 12–24 hours after transfection, cells were incubated with 100 μM pAz, 5 mM $Mg(OAc)_2$, and 3 mM ATP in serum-free MEM for 1 hour at 37°C. The cells were then rinsed with MEM twice and Dulbecco's Phosphate Buffered Saline (DPBS) once.

The ligated pAz was detected by performing Cu-catalyzed click chemistry with biotin-alkyne. The click reagents were prepared as follows: 500 μM THPTA, 100 μM $CuSO_4$, 2.5 mM sodium ascorbate were incubated at room temperature for 10 minutes then diluted to volume with DPBS. AlexaFluor-647-alkyne was then added to a final

concentration of 4 μM. The labeled HEK cells were incubated with the click reagents at room temperature for 10 minutes, washed twice with 1 mg/mL BSA in DPBS (DPBS-B).

The LplA surface-expression level on the HEK cells was detected using mouse monoclonal anti-FLAG antibody (Stratagene) in DPBS-B at 1:300 dilution with incubation at room temperature for 10 minutes. After washing, cells were incubated with either anti-mouse AlexaFluor-568 or anti-mouse AlexaFluor-488 secondary antibody (Invitrogen) in DPBS-B at 1:300 dilution for 10 minutes at room temperature. Cells were washed twice with DPBS-B and then imaged.

Quantification of cis labeling signal (Figure 6-11, 6-12)

For quantification of imaging data, cells with surface signal:background >2:1 in the AlexaFluor647 channel were selected for analysis. The background corrected mean intensity data for each ROI was calculated using Slidebook 5.0 software. The bar graphs show average fluorescence intensity ratios for ~25 cells for each condition (With B4.3 and B4.6 cis labeling onto LAP-NLG, we were unable to find 25 cells with signal:background >2:1, these values reflect averages for just 8-10 cells).

ER labeling (Figure 6-13)

HEK cells were transfected at ~70% confluency with expression plasmids for FLAG-^{W37A}LplA-ER (or variant) (600 ng for a 0.95 cm2 dish) and LAP-neuroligin (600 ng) using Lipofectamine 2000 (Invitrogen). A nuclear co-transfection marker (H2B-YFP) was used (50 ng). 24 hours after transfection, cells were incubated with 25 μM pAz-AM in serum-free MEM for 4 hours at 37°C. The media was then replaced 3 times over 30 minutes at 37°C. The cells were then rinsed with DPBS once and the ligated pAz was detected by Cu-catalyzed click as described in **Cis pAz ligation on mammalian cell surface.**

Trans pAz ligation on mammalian cell surface (Figure 6-20, 6-22)

One 0.95 cm^2 dish of HEK cells was transfected with FLAG-^{W37A}LplA-neurexin1b (or variant). Another 0.95 cm^2 dish of HEK cells was transfected with LAP-neuroligin-1 (600 ng) and a nuclear CFP co-transfection marker (H2B-CFP, 50 ng). Transfection was

performed at ~70% confluency using Lipofectamine 2000 (Invitrogen). 4 hours after transfection cells were washed twice with MEM and then placed back in the incubator with complete media. After 6 hours of incubation, cells from each dish were rinsed quickly once with DPBS and then incubated in each dish with trypsin-EDTA (_____) for 5 minutes at 37°C. The two pools of cells were then mixed well and co-plated on fresh cover slips coated in with 50 μg/ml fibronectin (Millipore).

After co-plating, cells were incubated an additional 12 hours and then labeled. For labeling, cells were incubated 100 μM pAz, 5 mM $Mg(OAc)_2$, and 3 mM ATP in serum-free MEM for 5 minutes at room temperature. The cells were then rinsed with MEM twice and Dulbecco's Phosphate Buffered Saline (DPBS) once.

Cu-catalyzed click chemistry and live-cell antibody staining were then performed as described in **Cis pAz ligation on mammalian cell surface.**

Grasp comparison (Figure 6-21, 6-22)

One 0.95 cm^2 dish of HEK cells was transfected with sfGFP1-10 fused to neurexin-1b (or variant, 400 ng) and a whole-cell mCherry co-transfection marker (50 ng). Another 0.95 cm^2 dish of HEK cells was transfected with sfGFP11-neuroligin-1 (400 ng) and a nuclear CFP co-transfection marker (H2B-CFP, 50 ng). Transfection was performed at ~70% confluency using Lipofectamine 2000 (Invitrogen). Co-plated cultures were prepared as described in **Trans pAz ligation on mammalian cell surface**. After co-plating cells were incubated for 20 hours and then imaged.

Protein expression and purification

LplA mutants demonstrating activity on the mammalian cell surface were cloned into bacterial expression vector (pYFJ16) and enzymes were expressed in BL21-DE3 bacteria with IPTG induction as previously described (cite).

After Ni-NTA purification, LplA enzymes were stored in a stabilizing buffer: 20 mM Tris pH 7.5, 1 mM DTT and 10% glycerol in aliquots at −80 °C.

Western blotting of mammalian cell lysate (Figure 6-18)

HEK cells expressing LAP2-YFP or LAP2(K→A) were lysed under hypotonic conditions in 1 mM HEPES pH 7.5 with 5 mM $MgCl_2$, 1 mM phenylmethylsulphonyl fluoride, and protease inhibitor cocktail (Calbiochem). After three cycles of freeze-thaw, cells were mixed by vortexing for 2 min. Lysate was cleared by centrifugation and stored in aliquots at −80 °C. To label with azide, lysates were incubated with 200 μM pAz, 3 mM ATP, 5 mM $Mg(OAc)_2$ and 200 nM of the indicated enzyme for 10 hours at 30 °C.

Azide derivatization was performed by adding 500 μM THPTA, 100 μM $CuSO_4$, 2.5 mM sodium ascorbate and 40 μM biotin-alkyne and incubating for 20 minutes at room temperature. Each reaction sample was boiled in protein loading buffer containing 2-mercaptoethanol for 10 minutes then divided in half and run on duplicate 12% SDS-PAGE gels. One gel was subjected to Coomassie staining, the other was transferred to a nitrocellulose membrane for western blot analysis.

The membrane was blocked with 3% BSA in TBS + 0.5% Tween-20 (TBS-T) for 1 hour at room temperature. The membrane was then incubated with streptavidin-horseradish peroxidase conjugate (Pierce, 1:3000 dilution) in 3% BSA in TBS-T for 40 minutes at room temperature. The membrane was washed four times with TBS-T then the blot was developed with Supersignal West Femto substrate (Pierce) and visualized on an Alpha Innotech ChemiImager 5500 instrument.

***In vitro* ligation reactions** (Figure 6-17, Table 6-2)

LplA mutants were purified as described above. For pAz ligation reactions: reactions were assembled with 2 μM enzyme, 150 μM LAP2 (GFEIDKVWYDLDA), 5 mM $Mg(OAc)_2$, 3 mM ATP and either 200 μM pAz4 or 50 μM pAz4 in DPBS + 10% glycerol. Reactions were incubated for 10 minutes then quenched with EDTA (C_f = 150 mM) and analyzed via HPLC. For lipoic acid and coumarin ligations: reactions were assembled with 2 μM enzyme, 150 μM LAP2, 5 mM $Mg(OAc)_2$, 3 mM ATP and 500 μM probe in DPBS + 10% glycerol. Reactions were incubated for 10 minutes then quenched with EDTA (C_f = 150 mM) and analyzed on a Varian Prostar HPLC using a reverse-phase *C18 Microsorb-MV 100 column* (250 × 4.6 mm). Chromatograms were recorded at 210 nm. For analysis of reactions using the LAP2 synthetic peptide

(GFEIDKVWYDLDA), we used a 14-minute gradient of 25–60% acetonitrile in water with 0.1% trifluoroacetic acid at 1 ml/minute flow rate. Percent conversions were calculated by dividing the product peak area by the sum of (product + starting material) peak areas.

Mass spectrometric analysis (Figure 6-17)

Starred peaks from Fig. 3A were manually collected and injected onto an Applied Biosystems 200 QTRAP mass spectrometer. The flow rate was 10 µl/min and mass spectra were recorded under the positive-enhanced multicharge mode.

Kinetic measurements (Figure 6-14, 6-19)

Reactions were assembled with 400 µM pAz4, 400 µM LAP2 (GFEIDKVWYDLDA), 1 mM ATP, 5 mM Mg(OAc)2, and 200 nM of enzyme and kept at 30C. After 3, 5, 8, and 11 minutes, an aliquot was drawn from the reaction vial, quenched with EDTA (Cf = 150 mM) and the product was quantified via HPLC as described previously in the section on ***in vitro* ligation reactions**. A plot of product concentration vs. time was made and the data was fit to a simple linear regression wherein the slope corresponds to the initial rate of the reaction (V_0). The value of kcat was calculated from the Michaelis-Menten equation $V_{max} = k_{cat}$*[Enzyme] at substrate-saturating conditions. Measurements were performed in triplicate.

References

1. Fernandez-Suarez, M., *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat Biotechnol* **2007**, *25*, 1483-7.

2. Yao, J. Z., *et al.* Fluorophore targeting to cellular proteins via enzyme-mediated azide ligation and strain-promoted cycloaddition. *J Am Chem Soc*, *134*, 3720-8.

3. Baruah, H.; Puthenveetil, S.; Choi, Y. A.; Shah, S.; Ting, A. Y. An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* **2008**, *47*, 7018-21.

4. Uttamapinant, C., *et al.* Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angew Chem Int Ed Engl.*

5. Huisgen, R.; Mloston, G.; Polborn, K. 1,3-Dipolar Activity in Cycloadditions of an Aliphatic Sulfine(,)(1). *J Org Chem* **1996**, *61*, 6570-6574.

6. Demko, Z. P.; Sharpless, K. B. An intramolecular [2 + 3] cycloaddition route to fused 5-heterosubstituted tetrazoles. *Org Lett* **2001**, *3*, 4091-4.

7. Wang, Q., *et al.* Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am Chem Soc* **2003**, *125*, 3192-3.

8. Himo, F., *et al.* Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J Am Chem Soc* **2005**, *127*, 210-6.

9. Link, A. J.; Vink, M. K.; Tirrell, D. A. Presentation and detection of azide functionality in bacterial cell surface proteins. *J Am Chem Soc* **2004**, *126*, 10598-602.

10. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **2004**, *126*, 15046-7.

11. Agard, N. J.; Baskin, J. M.; Prescher, J. A.; Lo, A.; Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol* **2006**, *1*, 644-8.

12. Baskin, J. M., *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc Natl Acad Sci U S A* **2007**, *104*, 16793-7.

13. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011**, *108*, 11399-404.

14. Jones, R. P.; Greenfield, P. F. Ethanol and the fluidity of the yeast plasma membrane. *Yeast* **1987**, *3*, 223-32.

15. Fujiwara, K., *et al.* Global conformational change associated with the two-step reaction catalyzed by Escherichia coli lipoate-protein ligase A. *J Biol Chem* **2010**, *285*, 9971-80.

16. 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*, 33645-51.

17. Yamagata, M.; Sanes, J. R. Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Front Mol Neurosci*, *5*, 18.

18. Kim, J., *et al.* mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nat Methods*, *9*, 96-102.

19. Agresti, J. J., *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* **2010**, *107*, 4004-9.

20. Gautier, A., *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* **2008**, *15*, 128-36.

21. Los, G. V., *et al.* HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008**, *3*, 373-82.

22. Colby, D. W., *et al.* Engineering antibody affinity by yeast surface display. *Methods Enzymol* **2004**, *388*, 348-58.

Chapter 7. Combining yeast-display evolution and *in vitro* compartmentalization for the directed evolution of LplA

The results of this chapter are unpublished. This project was performed in collaboration with Dr. Pascaline Mary and Dr. Ilke Akartuna, members of the Weitz Lab at Harvard University. Pascaline and Ilke fabricated all microfluidic devices and picoinjection devices used in this project. We thank Dr. Irwin Chen (Harvard University) for providing plasmids.

Introduction

Yeast-display can be a powerful tool for the evolution of enzymatic activity. As we showed in Chapter 6, just four rounds of selection on a yeast-display library produced LplA variants with ~5-fold increase in catalytic activity over the parent enzyme and drastically improved activity in the secretory pathway. Further demonstrating the power of this technique, the two other examples of evolving improved catalysis using yeast display have produced variants with 2- to 150-fold increases in catalytic activity[1,2]. However, the application of yeast-display to enzyme evolution does present some significant problems. First, the method requires that the enzymatic product be linked to the yeast cell surface. Second, the yeast-display evolution platform may produce artificially low turnover due to the requirement for surface-tethered ligase and the steric constraints introduced by the protein scaffolds. For example, in our case, very active LplA variants may have their product conversion limited by steric access to LAP2 peptide instead of by intrinsic activity. Finally, while evolutionary pressure can be easily controlled on the yeast surface, the co-display method forces both the enzyme and peptide substrate to be expressed with very high effective molarities. This limits the ability to both evolve for improved catalytic efficiency and select for lower peptide substrate K_M.

The necessity of linking enzymatic product to the yeast cell surface has recently been circumvented in a collaborative effort between the Weitz lab at Harvard University and the Klibanov lab at MIT[3]. In this work, the authors used a drop-based microfluidic device to encapsulate yeast cells that were displaying a library of HRP mutants. The drops essentially served as reaction vessels, with each drop containing at most one yeast cell and a fluorogenic HRP substrate. The emulsion compartmentalized the individual yeast cells and effectively linked the diffusible HRP product (phenotype) to the active HRP mutant expressed on the yeast cell (genotype). Although similar *in vitro* compartmentalization (IVC) techniques have been performed using protein-coated beads[4,5] instead of a yeast cell, this new method presents a few significant advantages. First, traditional methods for IVC use homogenization or "mixing" to create the oil-in-water emulsion[6,7], which produces polydisperse droplets with non-uniform volumes. The

authors use microfluidic devices to create monodisperse droplets with identical and tunable volumes[8]. While this is not the first example of such a device being used for IVC[9,10], achieving consistent volumes and thus uniform concentrations of reaction components is critical when evolving enzyme activity. Second, the authors were able to directly sort the fluorescent droplets within the microfluidic device, eliminating the need to perform bulk FACS after encapsulation and reaction incubation. This shortens the overall protocol time and eliminates potential droplet or library loss in the transfer to FACS sorting. Third, because the emulsion droplets can be sorted directly, there is no need to perform downstream coupling of the diffusible product to the gene, as required by all previous bead-based methods[5]. Finally, the authors had previously introduced a method for using voltage-controlled microfluidic channels to coalesce droplets and to add very precise volumes to existing droplets[11]. This ability to control timing of reagent additions to the encapsulated yeast cells is exciting and permits extremely creative and tailored selection strategies.

We immediately saw the advantages of combining *in vitro* compartmentalization and yeast-display for the evolution of LplA. We recognized that labeling LAP2 on the yeast cell surface with LplA library members that were freely diffusible in solution would solve the problem of artificially low turnover due to steric inaccessibility of the enzyme. This would also eliminate any labeling artifacts due to high local concentrations of the tethered enzyme. As outlined in Figure 7-1, we envisioned using droplets to encapsulate individual yeast cells at which point we could cleave LplA from the yeast cell surface. We could then rely on the droplet to restrict soluble LplA to labeling only the LAP2 on the yeast cell it came from, thus preserving the genotype-phenotype linkage. We hoped this evolution strategy would evade the limitations of co-display using a surface-tethered enzyme and produce LplA variants with still greater catalytic activity.

In this chapter, we will describe the development and optimization of a directed evolution technique that combines yeast-display and *in vitro* compartmentalization for the evolution of multiple-turnover ligases. We perform initial model selections and then apply the method to evolve ligases with improved lipoylation activity over wild-type LplA. Although these initial experiments demonstrate the power of the method in

evolving ligases with improved activity, there are technical hurdles that will need to be addressed before this platform can be used to select based on unnatural probe ligation.

Results

Combined yeast-display and IVC selection scheme

We began a collaboration with the Weitz lab to perform the microfluidic encapsulation of yeast cells. The general scheme for the evolution platform is depicted in Figure 7-1. We first altered the traditional co-display of LplA and LAP2 by incorporating a tobacco etch virus (TEV) protease cleavage site between the LplA and the HA tag (Figure 7-1A). This configuration ensures that the expression level of the enzyme can still be monitored after cleavage by antibody staining for the HA tag, which remains on the cell surface after enzyme cleavage. We then use a microfluidic drop-making device to co-flow the yeast cells and a solution of TEV protease (Figure 7-1B). The aqueous phase is then flowed into fluorocarbon oil with a biologically inert surfactant that creates the individual droplets. The cells are introduced at a concentration such that each 6 pL droplet contains at most 1 yeast cell. The droplets are then collected into a delay line and the TEV protease cleaves the enzyme from the surface of the yeast (Figure 7-1C).

After performing the TEV cleavage inside the droplet, we then need to introduce the labeling reagents into each droplet so the cleaved LplA can label LAP2 on the yeast cell surface. The Weitz lab has developed a voltage-controlled channel which can precisely inject a small volume of aqueous solution into individual droplets[11] (Figure 7-1D). We propose using this "picoinjection" device to deliver labeling reagents to each droplet. The droplets can then be incubated for any desired length of time, as droplets made using microfluidic devices are stable up to two days[3,11] (Figure 7-1E). The droplets encapsulating very active ligases will also contain highly labeled yeast. Finally, we can break the emulsions, collect the yeast cells, detect the extent of labeling, and sort via FACS (Figure 7-1F).

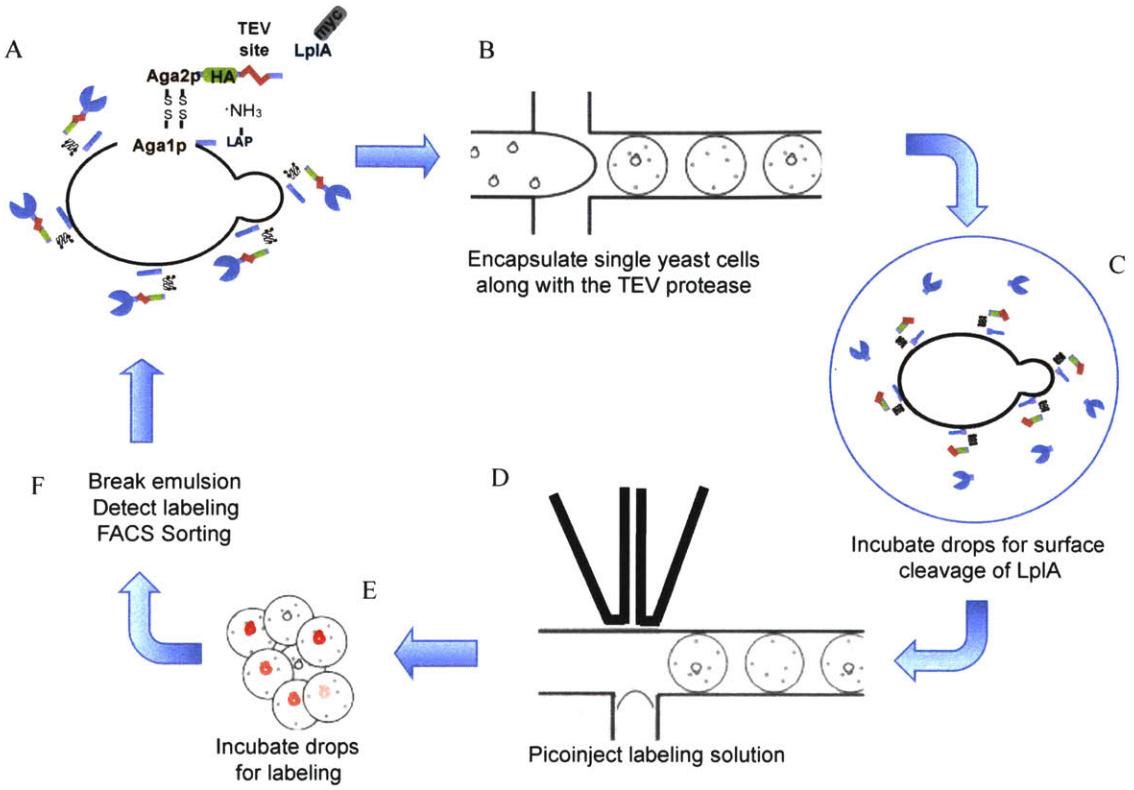


Figure 7-1. Selection scheme for combined yeast-display and IVC selections (A) Yeast cells are prepared as previously, the enzyme has an HA tag and a myc tag. A TEV cleavage site was introduced between the LplA and the Aga2p protein. The HA tag persists after LplA cleavage; the myc tag does not. **(B)** Yeast and TEV protease are co-flowed into an oil stream. The resulting droplets are 6 pL (23 microns). The yeast are at a concentration such that at there will be at most 1 yeast cell per droplet. **(C)** After forming, the droplets are incubated at room temperature for 40 minutes so the LplA can be cleaved from the yeast cell surface. **(D)** Pre-formed droplets containing cleaved LplA are flowed through a picoinjection device. A voltage-controlled channel (electrode in black) allows for the picoinjection of 0.6 pL (1:10 drop volume). Each drop receives a precise and consistent amount of labeling reagent. **(E)** Drops are incubated for any length of time to allow labeling to occur **(F)** After labeling, emulsions can be destroyed, yeast recovered, labeling detected, and FACS sorting performed. Figure modified from Agresti, *et. al.*[3]

^{WT}LplA can be fully cleaved from the yeast cell surface and lipoylates LAP2 at low concentrations

This is a very complicated experiment compared to the traditional tethered-enzyme yeast display. Therefore, before embarking on model selections, we wanted to ensure we could meet three essential criteria. First, we wanted to demonstrate that we observe labeled yeast when LplA is provided at concentration ranges achievable within the droplet. Second, we need to show that we can achieve complete cleavage of LplA from the yeast cell surface using TEV protease and on a reasonable timescale. Finally, we wanted to demonstrate that the combination of yeast-display and in vitro

compartmentalization actually produces improved labeling signal over the traditional tethered yeast-display method. These three achievements would demonstrate this method is promising enough to continue on to model selections.

To meet the first criterion, we determined the range of concentrations that we could practically produce inside each droplet. There are approximately 10^4 enzyme molecules displayed per yeast cell[1]. If each 6 pL droplet contains a single yeast cell with 70-100% of the enzyme cleaved and in solution, then enzyme concentrations will range from 25-40 nM. We performed lipoylation of LAP2 yeast using ^{WT}LplA at high concentrations and at concentrations within that range and detected lipoylation using antibody staining. Figure 7-2 shows 1D FACS lipoylation histograms of those labeling experiments. When a high concentration of ^{WT}LplA is used, the yeast cells are highly labeled even at the shortest labeling time tested (20 min). However when ^{WT}LplA is provided at low concentrations, LAP2 lipoylation level is very low. It requires very long labeling times (20 hours) to start seeing a significant population of labeled yeast using just 25 nM of enzyme. However, as previously mentioned, the drops are extremely stable and long probe incubation times are feasible.

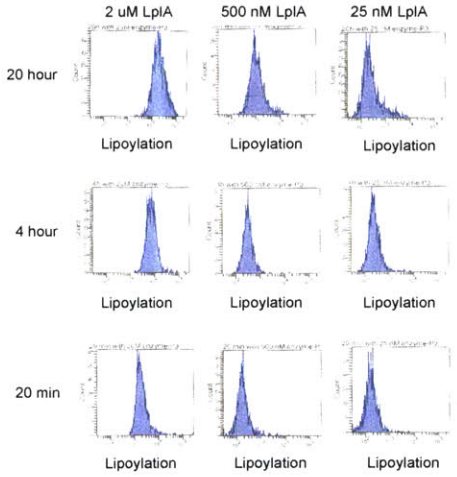


Figure 7-2. Analysis of lipoylation activity using low concentrations of ^{WT}LplA Yeast expressing LAP2 were labeled with lipoic acid and ATP using the indicated ^{WT}LplA concentration for the indicated amount of time. Lipoylation was detected by antibody staining and cells were analyzed by FACS.

The next goal was to demonstrate that efficient and high-yielding TEV cleavage can be achieved within the droplets. We first expressed and purified the S219V mutant of TEV protease, which has been shown to have superior cleavage properties over the wild-

type protease[12]. We applied the purified protease to the surface of yeast cells expressing ^{WT}LplA with the TEV protease site added to the linker. In this assay, the extent of ^{WT}LplA cleavage can be monitored by loss of myc signal. Unfortunately, we observed very little cleavage using a standard TBS-B buffer (Figure 7-3A). This is probably not surprising given the report that TEV protease activity is improved in reducing buffers, and 1 mM DTT is recommended[13]. However, since Aga1p and Aga2p are connected to each other via disulfides[14], we worried that DTT might be too reducing and disrupt the disulfides, causing loss of the whole Aga2p scaffolding fusion. If that is the case, then incubation with TBS-B containing 1 mM DTT will show loss of myc signal in the absence of TEV protease, indicating that the entire Aga2p-HA-TEV-LplA-myc fusion is being removed from the cell surface. In practice, that is exactly what we observe when using DTT (Figure 7-3B). We therefore investigated whether a milder reducing buffer system might provide an environment that is reducing enough for high TEV protease activity but not so reducing as to disrupt the Aga1p-Aga2p disulfide. After considerable optimization, we determined that a mixture of 3 mM reduced glutathione and 0.3 mM oxidized glutathione (GSH/GSSG) produced minimal disulfide cleavage in the absence of TEV protease. When this buffer system was used with TEV protease, we noticed a huge increase in TEV cleavage activity, with over 90% of the LplA cleaved from the yeast cell surface after just 40 minutes of incubation with the protease (Figure 7-3C).

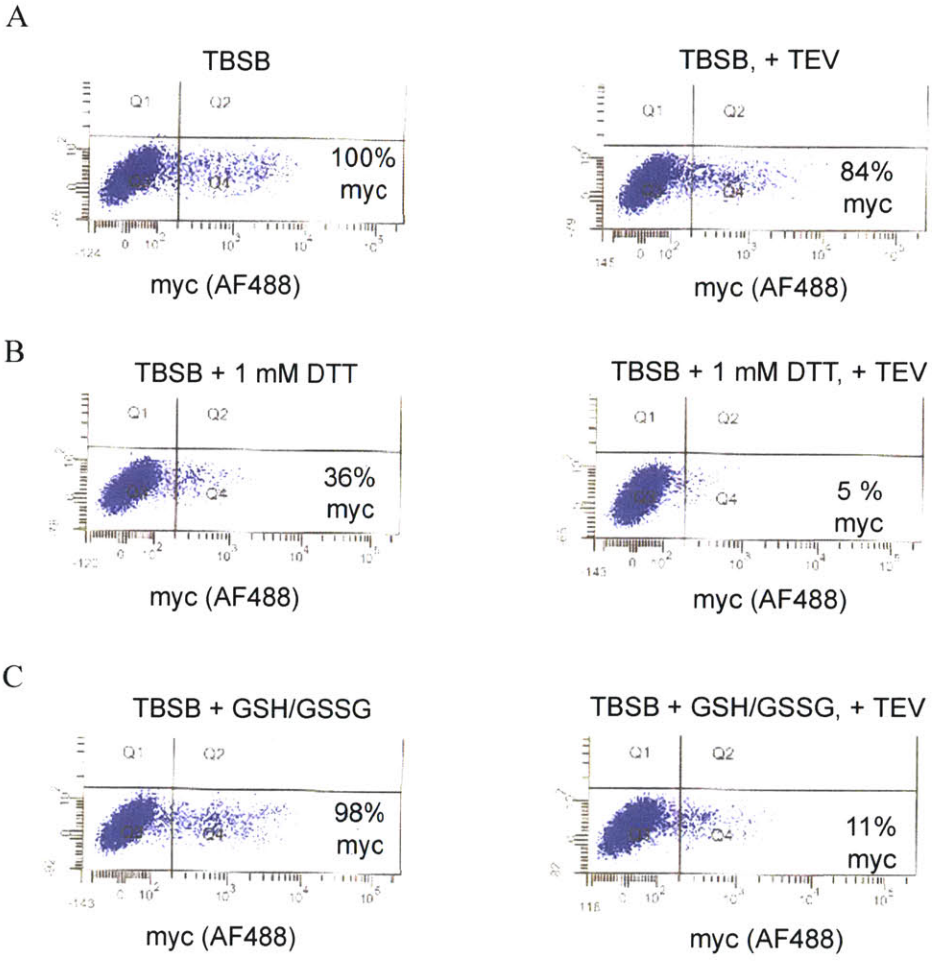


Figure 7-3. TEV cleavage in optimized buffer is complete within 40 minutes (A) Yeast expressing ^{WT}LplA with a linker containing a TEV protease site are incubated with and without TEV in TBS-B. **(B)** The experiment is performed as in (A) but with 1 mM DTT. Under these conditions, a significant portion of the ^{WT}LplA is cleaved in the absence of protease, corresponding to DTT-mediated cleavage of the disulfide between Aga1p and Aga2p. **(C)** When a buffer system of reduced and oxidized glutathione is used, TEV cleavage activity improves and there is no disruption of the Aga1p/Aga2p disulfide.

Finally, we wanted to determine whether the IVC combined with enzyme cleavage actually produced noticeable changes in yeast cell-surface labeling signal when using ^{WT}LplA. To assay for this, we performed labeling using the scheme outlined in Figure 7-1 using lipoic acid. This experiment used TEV protease cleavage of the enzyme along with in vitro compartmentalization in droplets (+TEV, +IVC) (Figure 7-4D). We compared the lipoylation signal observed under these conditions with the lipoylation

signal observed using traditional tethered-enzyme yeast display (-TEV, no IVC) (Figure -7A). As controls, when we perform either IVC or TEV cleavage alone we do not observe an increase in LAP2 lipoylation (B and C). We concluded from these results that eliminating steric restrictions for LAP2 labeling by ^{WT}LplA led to better enzyme turnover and improved the lipoylation signal when compared to the tethered-enzyme system.

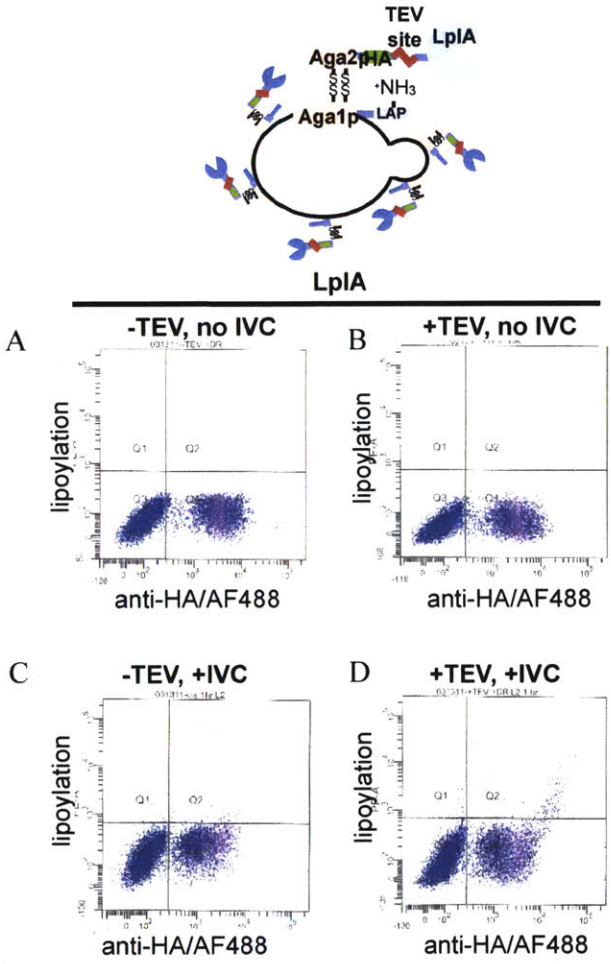


Figure 7-4. Demonstration of consistent picoinjection and improved labeling using combined yeast-display and IVC. (A) Yeast cells co-expressing ^{WT}LplA ligase and the LAP2 are lipoylated using cis labeling. **(B)** and **(C)** are controls where the yeast cells are treated with either IVC or with TEV, but not both. **(D)** Yeast cells co-expressing the ^{WT}LplA ligase and the LAP2 are lipoylated using *in vitro* compartmentalization and incubation with TEV protease. (D) is the only condition that gives high signal in this 1 hour labeling protocol. This indicates that IVC combined with yeast-display may be more sensitive than traditional tethered yeast-display.

Yeast-display and IVC combined selection gives enrichment equivalent to cis selections

The next step was to perform a model selection to demonstrate that the new evolution platform can distinguish between active and inactive, or less active, ligases. To do this, we prepared a mixture of 1:1, 1:10, or 1:100 ratios of ^{WT}LplA:^{W37V}LplA yeast. We then performed one round of selection (see Experimental) for lipoylation activity using IVC and TEV cleavage. After just one round of selection we observed ~50-fold enrichment of ^{WT}LplA over ^{W37V}LplA (Figure 7-5). This enrichment factor matches the enrichment we observed in model selections using tethered enzymes. We had hoped that the new method would improve our ability to discriminate between active and less active ligases. However, this does not appear to be the case based on the enrichment in the model selection, at least for lipoylation.

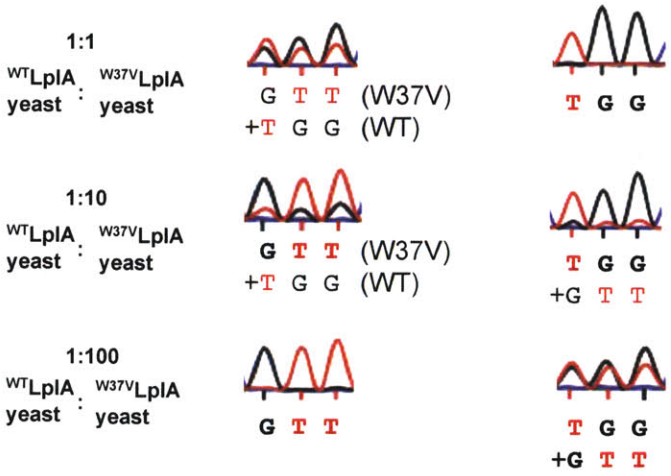


Figure 7-5. Sequencing results of model selection using yeast-display and *in vitro* compartmentalization. Model selections were performed as described in experimental, when using 1:1 and 1:10 ratios of ^{WT}LplA to ^{W37V}LplA enrichment was complete. With 1:100 ratios, enrichment of ^{WT}LplA was only half complete. Therefore, enrichment using this assay is calculated at ~50-fold.

Library generation and selections for increased lipoylation

We next created a series of error-prone PCR libraries based on ^{W37A}LplA. These libraries had mutation rates that ranged from an average of just one amino acid per gene to up to 12 amino acid mutations per gene. Unlike with the tethered yeast-display selections, for these selections we pooled the yeast libraries, giving a final library

diversity of 3.0 x10^8 library members. We performed four rounds of selection for lipoic acid. We initially tried several unnatural probes such as alkyl azide and bromo-undecanoic acid. However, we found that picoinjection using these more hydrophobic probes presented a significant problem with drop stability. The drop coalescence and inconsistent injection volumes that we observed were both the result of the highly concentrated dyes disrupting the oil-water interface; only lipoic acid was capable of being picoinjected reproducibly (data not shown). We therefore proceeded with selections based on lipoylation activity. We performed four rounds of selection on the pooled library. Evolutionary pressure for improved catalytic activity was introduced by decreasing the labeling times (from 1.5 hours to 20 minutes) and by collecting a smaller fraction of cells in each successive round (Figure 7-6). We performed sequencing of randomly selected clones at each round. By round 4, the original library of ~10^8 mutants had been reduced to just two consensus clones. One of the consensus clones had just one mutation (F281S) in addition to W37A, but the other was highly mutated with 14 amino acid mutations in addition to W37A.

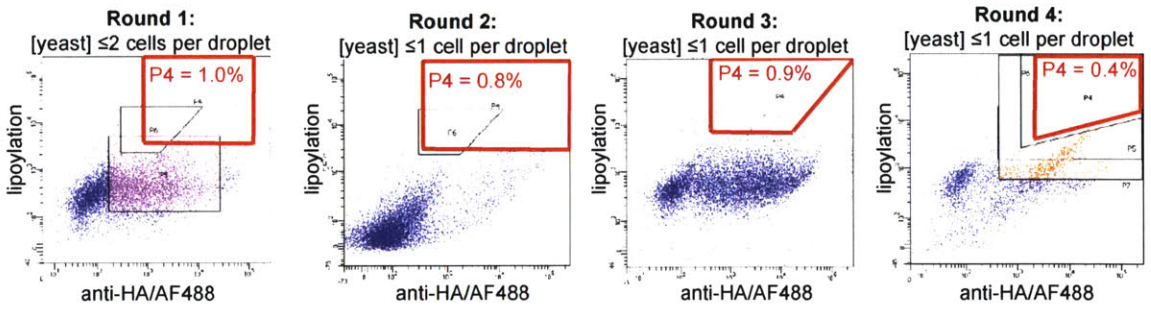


Figure 7-6. Sort windows for each round of lipoylation selection using yeast display and IVC. Selections were performed using increasingly stringent labeling for each round, all rounds used 500 μM lipoic acid: Round 1– 1.5 hour incubation; Round 2– 1 hour incubation; Round 3–40 minute incubation; Round 4– 20 minute incubation. Sort gates are indicated in red with the raw sort percentages reported for each round. Note: for Round 1 a higher concentration of yeast cells was used to reduce the amount of time spent performing picoinjection and sorting.

Characterization of consensus clones

We next wanted to characterize the lipoylation activity of these clones both on the mammalian cell surface and *in vitro*. We first tested $^{W37A+F281S}$LplA and ^{15Mut}LplA in a cis lipoylation assay. The enzymes were co-expressed as fusions to neurexin-1b along with LAP2-CFP-TM on the HEK cell surface. The cells were incubated with lipoic acid and

ATP for 1 hour and then cis lipoylation of LAP2 was detected by live-cell antibody staining (Figure 7-7). We observed that neither consensus clone performed as well as ^{WT}LplA and, more importantly, neither performed as well as ^{W37A}LplA, the original ligase. This was disconcerting, particularly when we also observed no increase in cis lipoylation on the yeast cell surface using these clones (data not shown). We developed two hypotheses for the lack of improved cis lipoylation activity. The first hypothesis was that perhaps selection where enzyme is freely diffusing inside the droplet is biasing against enzymes that are active as fusions. This would be unfortunate, but the enzymes developed using this strategy might still be applicable to intracellular or *in vitro* protein labeling. The second hypothesis was that the selections did not actually produce enzymes with increased activity and the apparent increase in labeling signal is due to probe or antibody sticking. To get more information, we decided to test the LplA variants in a trans lipoylation assay.

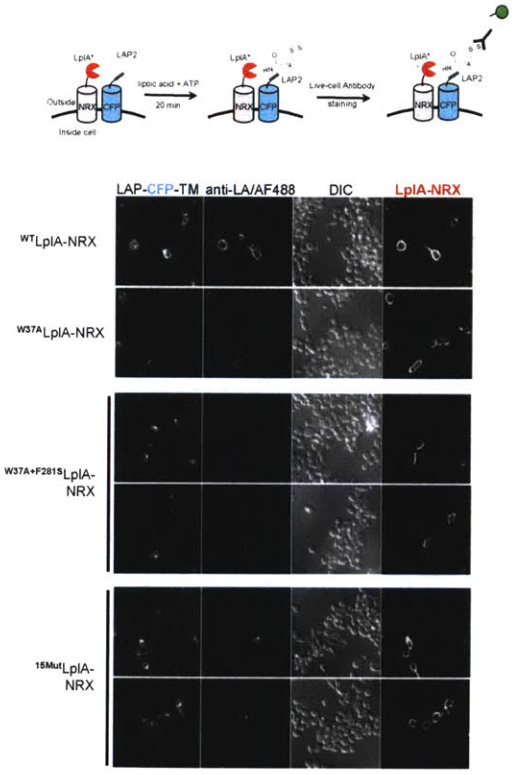


Figure 7-7. Testing cis lipoylation activity of consensus clones on the mammalian cell surface. Each LplA mutant was fused to the N-terminus of neurexin-1β (NRX) and co-expressed with LAP2-CFP-TM on the surface of HEK cells. Cis labeling was performed with addition of 500 μM lipoic acid and ATP for 20 minutes. After labeling the lipoyl-LAP2 conjugates were detected by live-cell antibody staining, which was also performed with anti-FLAG antibody.

In a trans ligation assay, we transfect one pool of HEK cells with LplA-neurexin-1β and a separate pool of HEK cells with LAP2-neruoligin-1. After transfection, we lift the two populations and then mix and co-plate them on the same coverslip. If the consensus clones have lipoylation activity when fused to a protein, then a contact point between the two populations of HEK should result in labeling of the LAP2-neuroligin (Figure 7-8, scheme). We performed the trans assay with ^{WT}LplA and with the two consensus clones (Figure 7-8). While ^{15Mut}LplA was inactive in this assay, the $^{W37A+F281S}$LplA variant gave better trans lipoylation activity than ^{WT}LplA. This demonstrates that, at least for some applications, the mutants evolved using yeast-display combined with IVC exhibit higher lipoylation activity than even the wild-type ligase. More importantly, this result demonstrates that a ligase that was selected for activity when freely diffusing can be expressed as a fusion and retain high lipoylation activity.

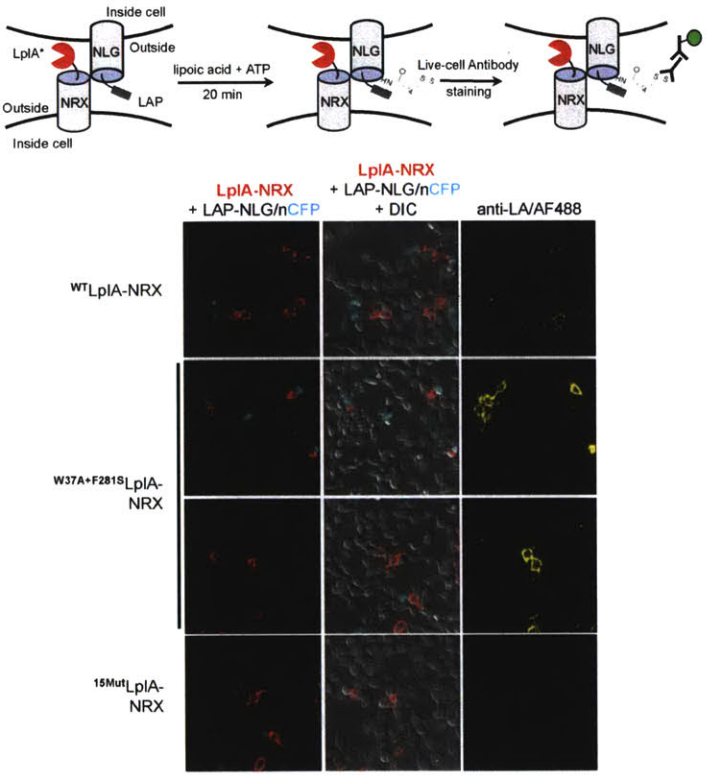


Figure 7-8 Imaging neurexin and neuroligin contacts using lipoylation. HEK cells expressing $^{W37A+F281}$LplA as a fusion to neurexin-1β (NRX) were co-plated with HEK cells expressing LAP2 as a fusion to neuroligin-1 (NLG, with a nuclear CFP co-transfection marker). When the two populations of cells come into contact, $^{W37A+F281S}$LplA can ligate lipoic acid onto LAP2-NLG in just 20 minutes. As controls, trans lipoylation is very low when using ^{WT}LplA-NRX and fails with ^{15Mut}LplA-NRX.

Finally, we sought to compare the *in vitro* lipoylation activity of the $^{W37A+F281S}$LplA consensus clone. We performed a 5-minute lipoylation reaction using purified enzyme and monitored conversion of LAP2 to lipoyl-LAP2 using HPLC (Figure 7-9). In a 5-minute lipoylation reaction, $^{W37A+F281S}$LplA produced twice as much product than ^{WT}LplA. This was a significant result because we had not previously obtained, through rational design or enzyme evolution, an LplA mutant that has better lipoylation activity than ^{WT}LplA in vitro. We determined that the lipoylation k_{cat} for the selected mutant $^{W37A+F281S}$LplA is 0.83 sec^{-1}, which is almost three times higher than the lipoylation k_{cat} for ^{WT}LplA (0.28 sec^{-1}). Because of the extremely poor activity we observed with ^{15Mut}LplA on the mammalian cell surface, we did not characterize this variant *in vitro*.

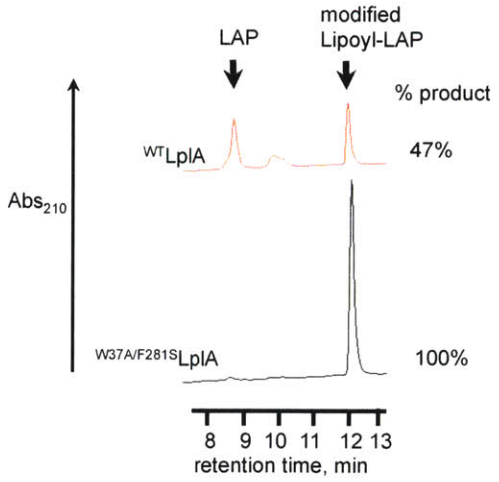


Figure 7-9. *In vitro* lipoylation of LAP2 using ^{WT}LplA and $^{W37A+F281S}$LplA. Reactions were assembled with 2 μM enzyme, 150 μM LAP2 (GFEIDKVWYDLDA), 5 mM $Mg(OAc)_2$, 3 mM ATP and 500 μM lipoic acid. Reactions were incubated for 5 minutes then quenched with EDTA (C_f = 150 mM) and analyzed via HPLC.

Conclusion

In this chapter we present initial progress towards the evolution of multiple-turnover ligases using a technique that combines yeast display evolution and *in vitro* compartmentalization. We demonstrate that this powerful method can be applied to select for an LplA variant with improved lipoylation kinetics over even wild-type enzyme. This evolution method allows large libraries to be processed in a short time (2,000 droplets per second)[3] and eliminates the need for surface-linked product.

However, there are some significant challenges to this method. First, picoinjecting hydrophobic substrates can be difficult due to miscibility issues with the already-formed yeast droplets. Moreover, the requirement for high concentrations of stock labeling solutions means that target probes must be able to be synthesized in very high yields and soluble in aqueous buffer at high concentrations. Finally, the real power of this technique is the ability to directly sort droplets via FACS. As described herein, this method requires a purification step to isolate the yeast, label with antibodies, and then sort by FACS. Future applications of this method by our lab should focus on using fluorescent (or fluorogenic) substrates so that FACS sorting can be performed on droplets within the device, which simplifies the technology a good deal.

Experimental

General Yeast methods

Growth and induction

Unless otherwise noted, cells were grown in SDCAA growth media [2% dextrose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.86% Monosodium phosphate] at 30°C with shaking for 20 hours. To induce protein expression, the cultures were diluted to a cell density of ~10^6 cells/ml in SGCAA media (identical to SDCAA above but with 2% galactose as the carbon source). Induction in SGCAA was performed at 30°C for 20 hours.

Transformation

For single clones, L2-YIP yeast cells were transformed with DNA using the Frozen-EZ Yeast Transformation II kit (Zymo Research). The transformed yeast cells were plated on selective plates (SDCAA –W, –U) and single yeast colonies were selected for amplification.

For the libraries, L2-YIP yeast cells were electroporated with the error-prone PCR library (see library construction below) using a BioRad GenePulser Xcell.

Long-term storage

For individual clones, glycerol stocks were made. Isolated colonies were grown in 5 mL SDCAA overnight at 30°C. 812 μL of saturated overnight yeast culture was combined with 188 μL of 80% glycerol (sterile) and placed into a 2 mL cryo-vial, stored at -80°C.

For libraries and rounds, frozen stocks were made. At least 10^{10} freshly grown cells from each library were inoculated into 1 L low-dextrose SDCAA [as above but with 0.5% dextrose] and grown at 30°C for approximately 3 days. Cells were pelleted and then resuspended in freezing solution (2% glycerol, 0.67% yeast nitrogen base) to a final concentration of ~6 x 10^{10}/mL. Cells were divided into 1 mL aliquots in cryovials. The tubes were then put into a Mr. Frosty (Nalgene) and allowed to slow-freeze overnight -80°C. The next day, cells were moved to liquid nitrogen for long-term storage.

Sorting and Scanning Instrumentation

All fluorescence activated cell sorting was performed on a BD FACSAria (BD Biosciences). For all analysis where sorting was not required, a BD LSRII-HTS (BD Biosciences) was used. Both of these instruments are located in the Flow Cytometry Core Facility at the Koch Institute for Integrative Cancer Research at MIT.

Preparation of Yeast cells constitutively expressing LAP2-Aga1p

The E2p and LAP2-Aga1p yeast integrating plasmid was cloned from the YIPlac211-GPD-Avitag-Aga1p plasmid which was a gift from Irwin Chen at Harvard

University. First, two restriction sites, *AflII/NheI* were introduced on either side of the Avitag to create YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. The E2p tag was introduced by amplifying with primers containing *AflII/NheI* sticky ends and ligating into *AflII/NheI* digested YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. The LAP2 tag was introduced by pre-annealing LAP2 primers with *AflII/NheI* sticky ends and ligating into *AflII/NheI* digested YIPlac211-GPD-*AflII*-Avitag-*NheI*-Aga1p. This produced the yeast integrating plasmids YIPlac211-GPD-*AflII*-E2p-*NheI*-Aga1p and YIPlac211-GPD-*AflII*-LAP2-*NheI*-Aga1p.

To integrate the plasmid into the yeast genome, the yeast-integrating plasmids were linearized by digestion with *BsiWI* and transformed into *Saccharomyces cerevisiae* strain *BJ5465* using the protocol described above and plated on SDCAA plates lacking uracil. Colonies were picked and grown in SDCAA lacking uracil. A yeast colony with the LAP2-Aga1p construct correctly inserted and with good expression was designated "E2p-YIP" yeast and displays E2p constitutively on the cell surface as a fusion to the N-terminus of Aga1p. A yeast colony with the LAP2-Aga1p construct correctly inserted and with good expression was designated "LAP2-YIP" yeast and displays the LAP2 peptide sequence constitutively on the cell surface as a fusion to the N-terminus of Aga1p.

IVC optimizations of labeling and TEV cleavage (Figure 7-2, 7-3)

Testing lipoylation sensitivity

LAP2-YIP yeast cells were incubated with 500 μM lipoic acid, 3 mM ATP, 5 mM Mg(OAc) and the indicated concentration of WTLplA at room temperature for the indicated reaction times. Lipoylation was detected on the yeast cell surface by 1:200 dilution of anti-lipoic acid (rabbit) and incubating with yeast cells for 1 hour at 4 C. After 2 washes with PBSB, the yeast cells were incubated with 1:100 dilution of anti-Rabbit-phycoerythrin for 1 hour at 4C. After 2 washes with PBSB, cells were analyzed via FACS.

Optimization of TEV cleavage buffer

Yeast cells expressing BirA-TEV (Gift of Irwin Chen, Harvard University) were

incubated with 8 μM TEV for 30 minutes at room temperature in either standard buffer (TBSB) or two more reducing buffers: TBSB + 1 mM DTT and TBSB + 3 mM GSH/0.3 mM GSSG. Controls with no TEV added are included. The TBSB myc signal is set to 100% and all other conditions normalized to that value.

IVC method testing

LAP2-YIP yeast displaying WTLplA-TEV was prepared as previously described. The yeast were pelleted and washed twice with TBS-B + 3 mM GSH/0.3 mM GSSG and resuspended to a final concentration of 2.5 x 10^8 cells/mL. Yeast were prepared with 30% OPTIPREP (vendor) and 0.001% Pluronic F147. Yeast and 20 μM TEV (C_f droplet = 10 μM) were co-flowed into the dropmaking device at 50 μL/hour (for –TEV case, buffer was used in place of TEV solution). Droplets were formed within 0.1 sec of aqueous phase mixing by using a 75 μL/hour flow of HFE7500 + 1.8% RAS surfactant (Weitz lab, proprietary). The droplets were collected in a 40 minute incubation line for TEV cleavage and then flowed into the picoinjection device. 10x labeling solution was prepared: 5 mM lipoic acid, 30 mM ATP, 50 mM $Mg(OAc)_2$. Droplets were collected in 6 fractions on ice for 1 hour. Then fractions were incubated at RT for 1 hour for labeling to occur. After 1 hour labeling, 200 μL of 20% perfluorooctanol in HFE 3500 was added to each fraction and vortexed. Samples were spun down and the lower oil layer was removed. Samples were washed again with 200 μL of 20% perfluorooctanol in HFE3500, vortexed, spun, and oil removed. A final wash/vortex/spin step was performed with 200 μL of FL-40. Yeast cells were then pooled and washed 1x with PBS-B containing 100 mM EDTA followed by 2x washes with PBS-B, removing any supernatant remaining oil. Labeling for FACS involved adding 1:200 dilutions of anti-lipoic acid (rabbit) and anti-HA (mouse) and incubating with yeast cells for 1 hour at 4 C. After 2 washes with PBSB, the yeast cells were incubated with 1:100 dilutions of anti-Rabbit-phycoerythrin and anti-mouse AF488 for 1 hour at 4C. After 2 washes with PBSB, cells were analyzed via FACS.

IVC model selections (Figure 7-5)

E2p-YIP yeast displaying wild-type LplA-TEV or the less-active ^{W37V}LplA-TEV

mutant were mixed in ratios of 1:1, 1:10, and 1:100 wt:W37V. These samples were subjected to in vitro compartmentalization in microfluidic devices and picoinjected with the 10x lipoylation solution described above (C_f 500 μM lipoic acid, 3 mM ATP, 5 mM $Mg(OAc)_2$). The emulsion was broken and yeast cells were isolated for antibody staining for lipoylation and enzyme expression as described in **IVC method testing**. Cells were sorted and the top 9.0%, 0.89%, and 0.1% of the PE/AlexaFluor488 double positive population for the 1:1, 1:10, and 1:100 experiments, respectively, were collected. Collected cells were cultured until saturation in SDCAA growth media with 50 μg/mL penicillin, 25 μg/mL kanamycin, and 50 μg/mL streptomycin. Plasmid DNA was harvested using the Zymoprep Yeast Plasmid Miniprep kit (Zymo Research), and the recovered LplA genes were amplified using the primers PctAmp-F 5'-CGTTCCAGACTACGCTCTGCAG and PctAmp R 5'-CGAGATCTGATAACAACAGTGTAG and sent for sequencing analysis. Quantitative analysis of enrichment from pre- to post-sort was determined by analyzing nucleotide intensities on the sequencing chromatogram. Note: for the model selection, the enzymes used have E. Coli optimized codons. It wasn't until the library design that humanized codon versions of the LplA-TEV constructs were made.

Creating LplA libraries

The round zero (R0) LplA library was made of a mixture of libraries created with varying mutation rates. These libraries were made by error-prone PCR and cloned into LAP2-YIP *S. cerevisiae* using gap repair homologous recombination.

For all libraries, a ^{W37A}LplA was used as the template for the error-prone PCR.

For Library A, the PCR reaction contained 2 μM 8-oxo-2'deoxyguanosine (8-oxo-dGTP), 2 μM 4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one (dPTP), 200 μM each dNTP, and 0.2 μM each of primers 1F and 1R and reactions were thermocycled 10 times.

For Library B, the PCR reaction was assembled as above but with 200 μM dNTP analogues and was thermocycled 5 times.

For Library C, the PCR reaction was assembled as above but with 20 μM dNTP

analogues and was thermocycled 10 times.

For Library D, the PCR reaction was assembled as above but with 20 μM dNTP analogues and was thermocycled 20 times.

For all libraries, mutagenized genes were further amplified in PCR reactions without mutagenic dNTP analogs using primers 1F and 1R. The pCTCON2-Aga2p vector was prepared by digestion with *NheI* and *BamHI*, and gel-purified. The PCR-amplified inserts and digested pCTCON2 vector were mixed in 1:3 ratios and transformed together into LAP2-YIP *S. cerevisiae* by electroporation as described previously(Chapter 5 and 6). Homologous recombination occurred inside the yeast, giving the desired product. Serial dilutions of transformed yeast were plated on SDCAA plates and colonies were counted, to determine transformation efficiencies for each library. A total of ~10^8 cells from each fully grown individual library culture were mixed together and then the pooled library was pelleted and induced for labeling and sorting as described above.

IVC Library selections (Figure 7-7)

The pooled library described above was induced for 20 hours in SGCAA and washed twice with TBSB + 3 mM GSH/ 0.3 mM GSSG. Cells were resuspended to 5 x 10^8 for the first round of selection and to 2.5 x 10^8 for all other rounds. Cells were prepared for IVC as described above. For all selection rounds, the cells were picoinjected with the 10x lipoylation solution described above (C_f droplets = 500 μM lipoic acid, 3 mM ATP, 5 mM $Mg(OAc)_2$). Droplets were collected on ice for 15 hours in 10 minute fractions. For the first round, labeling was performed at RT for 90 minutes. For the second round, labeling was performed for 1 hour, then 40 minutes and 20 minutes for rounds 3 and 4, respectively.

Emulsions were broken and yeast cells were labeled as described previously. The cells were sorted cultured until saturation in SDCAA growth media with 50 μg/mL penicillin, 25 μg/mL kanamycin, and 50 μg/mL streptomycin. After the culture was saturated, ~10^9 yeast cells were pelleted and induced in SGCAA and prepared for the next round of selection.

Cis lipoylation on mammalian cell surface (Figure 7-8)

HEK cells were transfected with 600 ng of LAP2-CFP-TM and 600 ng of either ^{WT}LplA-NRX, or indicated consensus mutants using Lipofectamine 2000 (Invitrogen) per manufacturer's protocol. 12 hours post-transfection cells were labeled. For labeling with lipoic acid: cells were incubated with 500 μM lipoic acid, 5 mM $Mg(OAc)_2$, and 1 mM ATP for 20 minutes at 30C. For all samples, cells were washed 2 x 5 minutes with MEM and then lipoylation was detected via live-cell antibody staining.

Live-cell immunofluorescence staining was performed with 4 μg/ml anti-lipoic acid antibody (CalBiochem) and 4 μg/ml anti-FLAG antibody (Stratagene) in 1% BSA in DPBS for 15 minutes. After three quick washes, cells were incubated with 4 μg/ml goat anti-rabbit antibody conjugated to AlexaFluor 488 and 4 μg/ml goat anti-mouse AlexaFluor 568 (both Invitrogen) in 1% BSA in DPBS for 15 minutes, then washed and imaged.

Trans lipoylation on the mammalian cell surface (Figure 7-9)

One 0.95 cm^2 dish of HEK cells was transfected with FLAG- ^{WT}LplA-NRX (or variant). Another 0.95 cm^2 dish of HEK cells was transfected with LAP-neuroligin-1 (600 ng) and a nuclear CFP co-transfection marker (H2B-CFP, 50 ng). Transfection was performed at ~70% confluency using Lipofectamine 2000 (Invitrogen). 4 hours after transfection cells were washed twice with MEM and then placed back in the incubator with complete media. After 6 hours of incubation, cells from each dish were rinsed quickly once with DPBS and then incubated with trypsin-EDTA for 5 minutes at 37°C. The two pools of cells were then mixed well and co-plated on fresh cover slips coated in with 50 μg/ml fibronectin (Millipore).

After co-plating, cells were incubated an additional 12 hours and then labeled with lipoic acid as described above in **cis lipoylation on the mammalian cell surface**.

Protein expression and purification

LplA mutants demonstrating activity on the mammalian cell surface were cloned into bacterial expression vector (pYFJ16) and enzymes were expressed in BL21-DE3

bacteria with IPTG induction as previously described.

After Ni-NTA purification, LplA enzymes were stored in a stabilizing buffer: 20 mM Tris pH 7.5, 1 mM DTT and 10% glycerol in aliquots at −80 °C.

***In vitro* ligation reactions** (Figure 7-10)

LplA mutants were purified as described above. For lipoylation reactions: reactions were assembled with 2 μM enzyme, 150 μM LAP2 (GFEIDKVWYDLDA), 5 mM $Mg(OAc)_2$, 3 mM ATP and 500 μM lipoic acid. Reactions were incubated for 5 minutes then quenched with EDTA (C_f = 150 mM) and analyzed via HPLC. We used a 14-minute gradient of 25–60% acetonitrile in water with 0.1% trifluoroacetic acid at 1 ml/minute flow rate. Percent conversions were calculated by dividing the product peak area by the sum of (product + starting material) peak areas.

Kinetic measurements

Reactions were assembled with 500 μM lipoic acid, 400 μM LAP2 (GFEIDKVWYDLDA), 1 mM ATP, 5 mM $Mg(OAc)_2$, and 100 nM of enzyme and kept at 30°C. After 3, 5, 7, 9, and 11 minutes, an aliquot was drawn from the reaction vial, quenched with EDTA (C_f = 150 mM) and the product was quantified via HPLC as described previously in the section on ***in vitro* ligation reactions**. A plot of product concentration vs. time was made and the data was fit to a simple linear regression wherein the slope corresponds to the initial rate of the reaction (V_0). The value of kcat was calculated from the Michaelis-Menten equation $V_{max} = k_{cat}$*[Enzyme] at substrate-saturating conditions. Measurements were performed in triplicate.

References

1. Chen, I.; Dorr, B. M.; Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc Natl Acad Sci U S A* **2011**, *108*, 11399-404.

2. Lin, Y., *et al.* Display of a functional hetero-oligomeric catalytic antibody on the yeast cell surface. *Appl Microbiol Biotechnol* **2003**, *62*, 226-32.

3. Agresti, J. J., *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* **2010**, *107*, 4004-9.

4. Griffiths, A. D.; Tawfik, D. S. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J* **2003**, *22*, 24-35.

5. Stapleton, J. A.; Swartz, J. R. Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PLoS One*, *5*, e15275.

6. Griffiths, A. D.; Tawfik, D. S. Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol* **2006**, *24*, 395-402.

7. Lu, W. C.; Ellington, A. D. In vitro selection of proteins via emulsion compartments. *Methods*.

8. Paegel, B. M. Microfluidic landscapes for evolution. *Curr Opin Chem Biol* **2010**, *14*, 568-73.

9. Zhu, Y.; Power, B. E. Lab-on-a-chip in vitro compartmentalization technologies for protein studies. *Adv Biochem Eng Biotechnol* **2008**, *110*, 81-114.

10. Wu, N., *et al.* Management of the diffusion of 4-methylumbelliferone across phases in microdroplet-based systems for in vitro protein evolution. *Electrophoresis*, *31*, 3121-8.

11. Abate, A. R.; Hung, T.; Mary, P.; Agresti, J. J.; Weitz, D. A. High-throughput injection with microfluidics using picoinjectors. *Proc Natl Acad Sci U S A* **2010**, *107*, 19163-6.

12. Kapust, R. B., *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **2001**, *14*, 993-1000.

13. Tropea, J. E.; Cherry, S.; Waugh, D. S. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol Biol* **2009**, *498*, 297-307.

14. Boder, E. T.; Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat Biotechnol* **1997**, *15*, 553-7.