A fluorophore ligase for site-specific protein labeling inside living cells

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
A fluorophore ligase for site-specific protein labeling inside living cells

Chayasith Uttamapinant1, Katharine A. White1, Hemanta Baruah1,3, Samuel Thompson, Marta Fernández-Suárez4, Sujiet Puthenveetil, and Alice Y. Ting1

Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Biological microscopy would benefit from smaller alternatives to green fluorescent protein for imaging specific proteins in living cells. Here we introduce PRIME (PRobe Incorporation Mediated by Enzymes), a method for fluorescent labeling of peptide-fused recombinant proteins in living cells with high specificity. PRIME uses an engineered fluorophore ligase, which is derived from the natural Escherichia coli enzyme lipoic acid ligase (LplA). Through structure-guided mutagenesis, we created a mutant ligase capable of recognizing a 7-hydroxycoumarin substrate and catalyzing its covalent conjugation to a transposable 13-amino acid peptide called LAP (LplA Acceptor Peptide). We showed that this fluorophore ligation occurs in cells in 10 min and that it is highly specific for LAP fusion proteins over all endogenous mammalian proteins. By genetically targeting the PRIME ligase to specific subcellular compartments, we were able to selectively label spatially distinct subsets of proteins, such as the surface pool of neutrophin and the nuclear pool of actin.

Results

Structure-Guided Mutagenesis and Screening for a Coumarin Fluorophore Ligase. We selected 7-hydroxycoumarin as our fluorophore for this study because its excitation and emission wavelengths are suitable for live-cell imaging (excitation 387–405 nm; emission 448 nm (13)), it is brighter than Enhanced Blue Fluorescent Protein (EBFP) [ε = 36,700 M⁻¹ cm⁻¹ and QY 0.7 for the anionic form (13), compared to ε = 31,500 M⁻¹ cm⁻¹ and QY 0.2 for EBFP (14)], and it is small and hydrophobic, which increases the likelihood that it will be accepted by LplA. We synthesized four 7-hydroxycoumarin structures (Fig. 1B) with varying linkers separating the fluorophore and the carboxylic acid that is activated by LplA to the adenylate ester.

Using a high-performance liquid chromatography (HPLC) readout of ligation onto E2p, a domain from LplA’s natural protein substrate pyruvate dehydrogenase (15), we found that wild-type LplA, not surprisingly, does not use any of the coumarin substrates. We then examined the crystal structure of E. coli LplA bound to lipoic acid (16), and the related structure of Thermoplasma acidophilum LplA bound to lipoyl-AMP (17), to identify candidate positions for mutagenesis. Thirteen residues lie within 7.5 Å of the dithiolane moiety of lipoic acid in either LplA structure: N16, L17, V19, E20, F35, W37, S71, S72, H79, T87, R140, F147, H149 (E. coli numbering; Fig. S1A and B). We mutated each of these positions to alanine one by one and tested for uptake of the two smaller coumarin probes (n = 3.4). Only W37A LplA gave detectable ligation product.

Fig. S1B shows that W37 is 8.5 Å from the dithiolane ring in the E. coli structure, whereas the analogous residue in T. acidophilum LplA, Y39, is much closer, at 3.9 Å (Fig. S1A). A new structure of E. coli LplA with lipoyl-AMP bound (18) instead of lipoic acid suggests that the enzyme undergoes a conformational change upon formation of the adenylate ester. Interestingly, in this new structure, the W37 position is much closer to


The authors declare no conflict of interest.

This is a PNAS Direct Submission.

1C.U., K.A.W., and H.B. contributed equally to this work.
2To whom correspondence should be addressed. E-mail: ating@mit.edu
3Present address: Adimab, Inc., 16 Cavendish Court, Lebanon, NH 03766.
4Present address: Center for Engineering in Medicine, Harvard Medical School, Massachusetts General Hospital, 114 16th Street, Charlestown, MA 02129.
5This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.0914067107/-/DCSupplemental.
Cyclooctyne

**LplA mutant**

**Conversion**

**CELL BIOLOGY**

[Image]

ing LplA substrate specificity (19).

position most important (out of nine positions tested) for expand-

work to engineer an aryl azide ligase also identified W37 as the

ligation reactions catalyzed by lipoic acid ligase (LplA). The middle row shows

prepared a panel of all possible W37 LplA mutants and screened

Uttamapinant et al. PNAS

![Fig. 2.](image)

**Fig. 1.** Engineering a coumarin ligase. (A) Natural and engineered ligation reactions catalyzed by lipoic acid ligase (LplA). The middle row shows two-step probe targeting via alkyl azide 7 ligation followed by [3 + 2] cycload-

dition (12), and the bottom row shows direct fluorophore ligation by an LplA mutant. LAP = LplA Acceptor Peptide (20). The red circle represents any probe. (B) Relative activities of W37 LplA mutants with 7-hydroxycoumarin derivatives. Assays were performed under two conditions: 12 hours with LplA’s natural protein substrate E2p (15) or 40 min with LAP4.3D (20) (see SI Methods). No product was detected for the black-shaded quadrants. (C) Percent of LAP4.3D peptide converted to the coumarin (n = 4) conjugate for the eight most active W37 LplA mutants. Enzymes were used at 1 μM and reaction times were 40 min. Values are normalized to the percent conversion obtained with the best mutant, W37V, which is set to 100%. Measurements were performed in triplicate. Errors, ±1 s.d.

they against all four coumarin probes (Fig. 1B). We found that 11 mutations were inactive, although most of these were still able to ligate lipoic acid. Of the eight active mutants, little product was detected with the longer coumarin probes (n = 5.6), whereas the most product was seen with coumarin 4 (n = 4). We therefore reassayed these eight mutants under milder conditions with coumarin 4 to differentiate among them. Instead of Et2p, the HPLC assay was performed with an “LplA acceptor peptide” (LAP).

**Fig. 2.** In vitro characterization of coumarin ligase. (A) Natural and engineered ligation to LAP2 peptide (20) to the coumarin 4 conjugate by W37V LplA or W37I LplA (red traces). Negative controls with wild-type LplA, or ATP omitted, are shown in black. Standard peaks were collected and analyzed by mass spectrometry (Fig. 52). (B) Coumarin labeling of mammalian cell lysate to determine sequence-specificity. Lanes 1 and 4 show coumarin 4 ligation to LAP2-YFP catalyzed by W37V LplA and W37I LplA, respectively. Negative controls are shown with wild-type LplA (lane 7), an alanine mutation in LAP2 (lanes 2 and 5), and untransfected lysate (lanes 3 and 6). The same samples were also analyzed by blotting with anti-lipoic acid antibody. Lysate in lane 8 was subjected to in vitro lipoylation with exogenous wild-type LplA.

![Image](image)

the dithiolane ring (3.6 Å, Fig. S1C). We note that our previous work to engineer an ary1 azide ligase also identified W37 as the position most important (out of nine positions tested) for expanding LplA substrate specificity (19).

Encouraged by the coumarin ligation activity of W37V LplA, we prepared a panel of all possible W37 LplA mutants and screened

![Image](image)

**Coumarin Ligation Kinetics and Sequence Specificity.** We proceeded to characterize the two most active coumarin ligases identified in our screen: W37V LplA and W37I LplA. First, HPLC assays were performed, this time using our most kinetically efficient LAP called “LAP2” (20). Controls in Fig. 2A show that coumarin ligation to LAP2 requires ATP and is not catalyzed by wild-type LplA. The HPLC product peaks from the reactions with both ligases were collected and analyzed by mass spectrometry, confirming the identity of the covalent coumarin 4-LAP2 adducts (Fig. S2).

Second, we used HPLC to measure the kinetics of coumarin ligation onto LAP2 (Fig. S3A). For W37V LplA, we obtained a kcat of 0.019 ± 0.004 sec⁻¹ and a Km of 56 ± 20 μM. W37I LplA had a similar kcat (0.016 ± 0.002 sec⁻¹) but a much higher Km of 261 ± 69 μM. These values can be compared to previously determined kinetic constants for natural and unnatural LplA-cata-

lyzed ligations (Fig. S3B). Whereas the kcat of coumarin ligation is 10-30-fold poorer than previous kcat values determined for lipic acid (11, 15, 16), azidoalkanoic acid (11), and aryl azide (19), the coumarin Km for W37V LplA is better than previous Km values for unnatural substrates. In contrast, the coumarin Km for W37I LplA is much poorer. For live-cell labeling, we were uncertain whether low coumarin Km would confer an advantage, because intracel-
cular coumarin concentrations following probe loading may exceed 300 μM (vide infra). We therefore tested both LplA mutants for coumarin ligation activity in living cells.

Third, we characterized the sequence specificity of both coumarin ligases. It was unclear if remodeling of the lipoic acid binding pocket would affect LplA’s recognition of peptide and protein substrates. In addition, mammalian cells express protein substrates for their own mitochondrial lipoic acid ligase that might cross-react with E. coli LplA (15). To test for these possibilities, we performed coumarin labeling of LAP2-YFP (yellow fluorescent protein) in mammalian cell lysate and analyzed the samples by in-gel coumarin fluorescence imaging and anti-lipoic acid Western blotting. Fig. 2B shows that the endogenous protein substrates of mammalian lipoic acid ligase are present in the lysate (see anti-lipoic acid blot) but are not labeled by coumarin. In addition, no other endogenous proteins are labeled by coumarin, even though many are expressed at much higher levels than LAP2-YFP according to the Coomassie stain. It is possible that mitochondrial proteins are not labeled by coumarin because they are already saturated with lipoic acid. This idea is supported by the observation that in vitro lipoylation of lysate does not increase the lipoylation level of these endogenous proteins (lane 8 of Fig. 2B).

We also found that LAP2-YFP proteins were not detectably lipoylated, even though free lipoic acid may be present in the cytosol (lanes 1, 4, and 7). To confirm site-specificity, we mutated the central lysine in LAP2 to alanine and found that this abolished labeling by both W37VLplA and W37LplA (lanes 2 and 5). Thus, both coumarin ligases retain the high sequence-specificity of wild-type LplA.

Characterization of PRIME Labeling Inside Living Mammalian Cells. To test coumarin ligation inside cells, we first addressed the issue of specificity of wild-type LplA. Thus, both coumarin ligases retain the high sequence-specificity of wild-type LplA.

**Characterization of PRIME Labeling Inside Living Mammalian Cells.** To test coumarin ligation inside cells, we first addressed the issue of probe delivery. Coumarin 4 has 1–2 negative charges at physiological pH, making it membrane-impermeant. We protected either the carboxylic acid alone or both the acid and the 7-hydroxyl group with acetoxymethyl (AM) groups (coumarin-AM2; structure shown in Fig. 3A), which are known to be cleaved inside mammalian cells by endogenous esterases (21). Over time, excess anionic fluorophore is removed from cells via the action of organic anionic transporters (22). We found that one AM protecting group was not sufficient for membrane permeability; however, coumarin-AM2 entered cells readily (Fig. S4A).

We used the “wedge method” (see SI Methods) to measure the intracellular concentration of coumarin probe immediately after loading. With 20 μM coumarin-AM2, the intracellular coumarin concentration was ~469 μM after 10 min of loading (Fig. S4B). This concentration exceeds the coumarin Km for both W37VLplA and W37LplA. Empirical optimization showed that coumarin probe could be thoroughly “washed out” of cells by replacing the media several times over 30 min (Fig. S4C).

Next, we evaluated LplA expression in mammalian cells. The wild-type E. coli LplA gene was found to express poorly in Human Embryonic Kidney (HEK) and HeLa cells. Fusions to fluorescent proteins helped, but we obtained the best results by resynthesizing the LplA gene with human-preferred codons. Humanized genes for both W37VLplA and W37LplA gave robust and reproducible expression in a variety of cell lines, and their levels could be titrated by modulating the plasmid concentration used for liposome-mediated transfection (Fig. S5A).

Combining the optimized coumarin loading protocol, humanized LplA gene, and a LAP2-YFP fusion construct, we performed coumarin labeling with both W37VLplA and W37LplA in live HEK cells (Fig. S5B). We observed specific labeling with both coumarin ligases, evident from comparing the coumarin signal for LAP2-YFP and LAP2(K→A)-YFP. W37LplA gave higher coumarin signal/background ratios on average, consistent with its superior
Comparison of PRIME and FIAsh Labeling. We quantitatively analyzed the specificity, sensitivity, and toxicity of PRIME labeling in conjunction with a side-by-side comparison to FIAsh methodology. For the comparison, we prepared a single nucleus-targeted construct containing LAP2-mCherry, and the optimized 12-amino acid FIAsh recognition motif (7). In live HEK cells, PRIME labeling was performed first with \(^{W37V}\)LpaA for 10 min, followed by FIAsh for 30 min (Fig. S8A). To assess labeling specificity, we compared the nuclear coumarin or FIAsh signal (representing specific labeling, as the substrate is nuclear-localized) to the cytosolic signal (representing nonspecific labeling) in single cells. Fig. S8B shows that FIAsh gives more cytosolic background than PRIME. We also observed nonspecific FIAsh labeling in untransfected cells. A similar analysis of labeling sensitivity revealed that both methods can detect as little as 10 \(\mu\)M target protein with signal/background ratio >2:1 (Fig. S8C). An assay for mitochondrial respiration showed that FIAsh is more toxic than PRIME (Fig. S8D).

Imaging Nuclear Actin with PRIME. We used PRIME methodology to explore the biology of nuclear actin. Cytosolic \(\beta\)-actin has a well-recognized role in controlling cell migration and morphology, but biochemical studies have also suggested several roles for \(\beta\)-actin in the nucleus, including involvement in transcriptional regulation (24) and chromatin remodeling (25). Studies of nuclear actin have been hampered by the lack of methods to noninvasively detect all forms of actin in living cells. For example, antibody detection and phalloidin staining require cell fixation. Fluorescent proteins disrupt nuclear entry of actin, as shown in Fig. S9A. Phalloidin and Lifeact (26) are specific for polymeric actin over monomeric actin. Rhodamine-labeled actin can be visualized by fluorescent speckle microscopy, but this requires invasive microinjection (27).

We tested PRIME for labeling of nuclear actin in living cells. Fig. S9A (Left) shows confocal images of coumarin-labeled LAP2-\(\beta\)-actin in COS-7 cells also expressing a nuclear marker (histone 2B fused to YFP, or H2B-YFP). Coumarin signal is clearly present in the nucleus, which contrasts with confocal images of mApple-\(\beta\)-actin (Fig. S9A, Right). Though actin was overexpressed in these experiments, the diffuse coumarin labeling pattern suggests that nuclear actin may exist as monomers or short oligomers, consistent with previous studies (28).

Previous work has shown that cell stress, such as heat shock, can increase the abundance of nuclear actin (29). We therefore imaged actin after 30–90 min of heat shock (43 °C) and found that coumarin-labeled actin accumulates in the nucleus within 30 min, whereas mApple-actin accumulation is much slower (Fig. S9B). In a small fraction (<10%) of heat-shocked cells, we observed interesting coumarin-labeled nuclear actin structures (Fig. S9B), such as nucleolar clusters and filaments (which also stained with phalloidin, confirming F-actin morphology; Fig. S9C). This heterogeneity contrasted with the uniformly diffuse distribution of nuclear actin in nonstressed cells as well as mApple-actin in the nuclei of stressed cells. Perhaps filamentous actin, which has not previously been observed in the nucleus of living somatic cells, plays a role in transcriptional regulation (30) or nuclear stabilization (31).

The ability of PRIME methodology to detect nuclear pools of polymeric and nonpolymeric actin in living cells differentiates it from antibody, phalloidin, Lifeact, and fluorescent protein visualization techniques. PRIME may prove to be a useful tool for studying the biological function of nuclear actin.

Labeling Spatially Defined Protein Subpopulations Using Genetically Targeted Ligase. A unique feature of PRIME labeling is its ability to selectively highlight protein subpopulations via genetic targeting of the fluorophore ligase. To test this concept, we transfected HEK cells with both nuclear LAP2 and cytosolic LAP2. When cytosolic \(^{W37V}\)LpaA was coexpressed, we observed specific labeling of the cytosolic subpopulation of LAP2 (Fig. 4A, Left). Conversely, when nuclear-localized \(^{W37V}\)LpaA was coexpressed, we observed specific labeling of the nuclear subpopulation of LAP2 (Fig. 4B, Right). When enzyme and substrate were not colocalized, no coumarin labeling was observed. Similar results were obtained with genetically targeted \(^{V37L}\)LpaA (Fig. S7C).

We also tested this concept on neurexin-1-\(\beta\), a transmembrane adhesion protein that functions in synapse development and plasticity (32). We prepared a LAP2-neurexin-LAP2 double fusion construct to provide access to labeling reagents on both sides of the membrane. Fig. 4B shows PRIME labeling of the cell surface neurexin population by either purified \(^{W37V}\)LpaA applied to the cell medium or \(^{W37V}\)LpaA targeted to the inner leaflet of the plasma membrane. In contrast, untargeted \(^{W37V}\)LpaA labels the total intracellular pool of LAP2-neurexin-LAP2, including LAP2 present on the cytosolic face of the endoplasmic reticulum (ER) and the Golgi apparatus. Nuclear-targeted \(^{W37V}\)LpaA, as
expected, does not label neurexin. Similar results were obtained using targeted forms of W37ILplA (Fig. S7D).

We also experimented with W37VLplA and W37ILplA targeted to the ER using a KDEL sequence. We did not detect coumarin ligation onto LAP2-neurexin-LAP2 or a number of other LAP2 fusion proteins. This may be a consequence of high calcium concentration in the ER (33), which we found inhibits LplA activity.

To provide further evidence for nuclear actin, we performed compartmentalized labeling of LAP2-β-actin using W37VLplA-NLS (Fig. 4C). We initially did not detect any nuclear coumarin signal, perhaps because nuclear actin is bound within protein complexes and not sterically accessible to LplA. The nuclear actin signal seen when labeling with untargeted W37VLplA (Fig. S9A and Fig. 4C, Top) probably results from translocation of coumarin-labeled cytosolic actin to the nucleus, as suggested by a labeling experiment with W37VLplA-NES (Fig. 4C).

Therefore, we repeated the experiment with W37VLplA-NLS in the presence of leptomycin B to inhibit nuclear export of actin. In this case, we observed nuclear labeling, including actin structures similar to those seen in the nuclei of heat-shocked cells (Fig. 4C). Leptomycin B may cause nuclear actin levels to rise, saturating actin-binding proteins and producing free actin that is accessible to nuclear LplA.

Discussion

Using structure-guided mutagenesis of E. coli LplA, we have engineered a coumarin fluorophore ligase for PRIME labeling in living cells. Compared to other protein labeling methods, PRIME uniquely offers the combination of a peptide tag, high labeling specificity, and compatibility with the interior of living mammalian cells.

Aside from FlAsH, other peptide-based protein labeling methods are restricted to the cell surface either because the labeling reagents are not membrane-permeable [e.g., AcpS/Sfp (34) and sortase (35)] or because the chemistry is not orthogonal to intracellular functional groups [e.g., aldehyde tag (36) and ketone biotin (37)]. Unnatural amino acid mutagenesis stands apart as a
tagless protein labeling method. Though demonstrations of fluorophore incorporation in mammalian cells are promising (38), the prevalence of native amber stop codons in mammalian cells (38) and potential dominant negative effects from truncated protein by-products (39) could limit the utility of this method.

Several other protein labeling methods, like PRIME, use enzymes to catalyze probe conjugation to proteins (40). Here we explored genetic targeting of the labeling enzyme as a method to spatially restrict labeling to protein subpopulations of interest. With suitable sensitivity and labeling speeds, one might be able to use this strategy to study protein trafficking between subcellular compartments and organelles. The utility would be greatest for compartments that are too small to selectively illuminate with focused light, such as neuronal synapses.

Materials and Methods

In Vitro Coumarin Ligation. General reaction conditions: 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 NaHPO₄ pH 7.2 at 30 °C for various lengths of time. Reactions were quenched with EDTA (10 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 LAP2 peptide (GFEIDKVYWYDLOA), we used a 10-min gradient of 30–60% acetonitrile in water with 0.1% trifluoroacetic acid at 1 ml/min flow rate. Percent conversions were calculated by dividing the product peak area by the sum of product + starting material peak areas.

Coumarin Labeling of Mammalian Cell Lysate. 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 MgCl₂, 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 coumarin or lipoic acid, lysates were incubated with 200 mM LplA, 300 μM coumarin 4 or 250 μM lipoic acid, 1.5 mg/mL (pH 2.2) 5 and 5 mM Mg(OAc)₂, in 25 mM NaHPO₄ pH 7.2. After overnight incubation at 30 °C, reactions were boiled in protein loading buffer containing 2-mercaptoethanol for 7 min, then separated on a 12% SDS-PAGE gel. Coumarin fluorescence was visualized on an Alpha Innotech Chemilinker 5500 instrument using 365 nm UV light for excitation.

General Protocol for Coumarin PRIME Labeling in Living Cells. HEK cells were transfected at ~70% confluency with expression plasmids for 104-56-03 (20 ng for a 0.35 cm² dish) and the LAP2 fusion protein of interest (400–600 ng) using Lipofectamine 2000 (Invitrogen). 8–24 h after transfection, cells were incubated with 20 μM coumarin-AM, in serum-free DME (Dulbecco’s modified Eagle medium, Cellgro) for 10 min at 37 °C. 0.1% v/v Pluronic F-127 (Invitrogen) can be optionally added to the labeling solution to give more even coumarin distribution in cells. The media was replaced 3–4 times over 30–60 min at 37 °C with DME with 10% FBS (Fetal Bovine Serum), to wash out excess coumarin. Cells were imaged as described in SI Methods.

For further information on cloning, synthetic methods, cell culture, labeling, imaging, and other protocols, see SI Methods.

ACKNOWLEDGMENTS. We thank Amar Thyagarajan and Peng Zou for technical assistance. John Cronan (University of Illinois), Michael Davidson (Florida State University), Birka Herrmann (Göttingen), Daniel S. Liu (MIT), and Masahito Yamagata (Harvard) provided plasmids. Funding was provided by the National Institutes of Health (R01 GM072670, MIT, and the Dreyfus Foundation. S. T. was supported by an HHMI summer fellowship. M. F.-S. was supported by a fellowship from La Caixa Foundation (Spain).