

## Yeast Display Evolution of a Kinetically Efficient 13-Amino Acid Substrate for Lipoic Acid Ligase

Sujiet Puthenveetil, Daniel S. Liu, Katharine A. White, Samuel Thompson, and Alice Y. Ting\*

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

Received June 5, 2009; E-mail: ating@mit.edu

**Abstract:** *Escherichia coli* lipoic acid ligase (LplA) catalyzes ATP-dependent covalent ligation of lipoic acid onto specific lysine side chains of three acceptor proteins involved in oxidative metabolism. Our lab has shown that LplA and engineered mutants can ligate useful small-molecule probes such as alkyl azides (*Nat. Biotechnol.* **2007**, *25*, 1483–1487) and photo-cross-linkers (*Angew. Chem., Int. Ed.* **2008**, *47*, 7018–7021) in place of lipoic acid, facilitating imaging and proteomic studies. Both to further our understanding of lipoic acid metabolism, and to improve LplA's utility as a biotechnological platform, we have engineered a novel 13-amino acid peptide substrate for LplA. LplA's natural protein substrates have a conserved  $\beta$ -hairpin structure, a conformation that is difficult to recapitulate in a peptide, and thus we performed *in vitro* evolution to engineer the LplA peptide substrate, called "LplA Acceptor Peptide" (LAP). A  $\sim 10^7$  library of LAP variants was displayed on the surface of yeast cells, labeled by LplA with either lipoic acid or bromoalkanoic acid, and the most efficiently labeled LAP clones were isolated by fluorescence activated cell sorting. Four rounds of evolution followed by additional rational mutagenesis produced a "LAP2" sequence with a  $k_{\text{cat}}/K_m$  of  $0.99 \mu\text{M}^{-1} \text{min}^{-1}$ , >70-fold better than our previous rationally designed 22-amino acid LAP1 sequence (*Nat. Biotechnol.* **2007**, *25*, 1483–1487), and only 8-fold worse than the  $k_{\text{cat}}/K_m$  values of natural lipoate and biotin acceptor proteins. The kinetic improvement over LAP1 allowed us to rapidly label cell surface peptide-fused receptors with quantum dots.

### Introduction

Most proteins are evolved to interact with a multitude of cellular molecules and thus contain a number of distinct domains, binding sites, and activities. Often, it is useful to the biochemist to reduce a specific aspect of a protein's function to just a peptide fragment. This can help to determine the minimal features of a protein required for a specific function such as binding, recognition by an enzyme, translocation, or folding.<sup>1–4</sup> It may also be desirable to create a consensus peptide substrate for assay purposes,<sup>5,6</sup> or to use a peptide in place of a protein to facilitate crystallography of multiprotein complexes.<sup>7,8</sup> For therapeutic applications, replacement of protein drugs with peptides having similar activity can improve tissue penetration

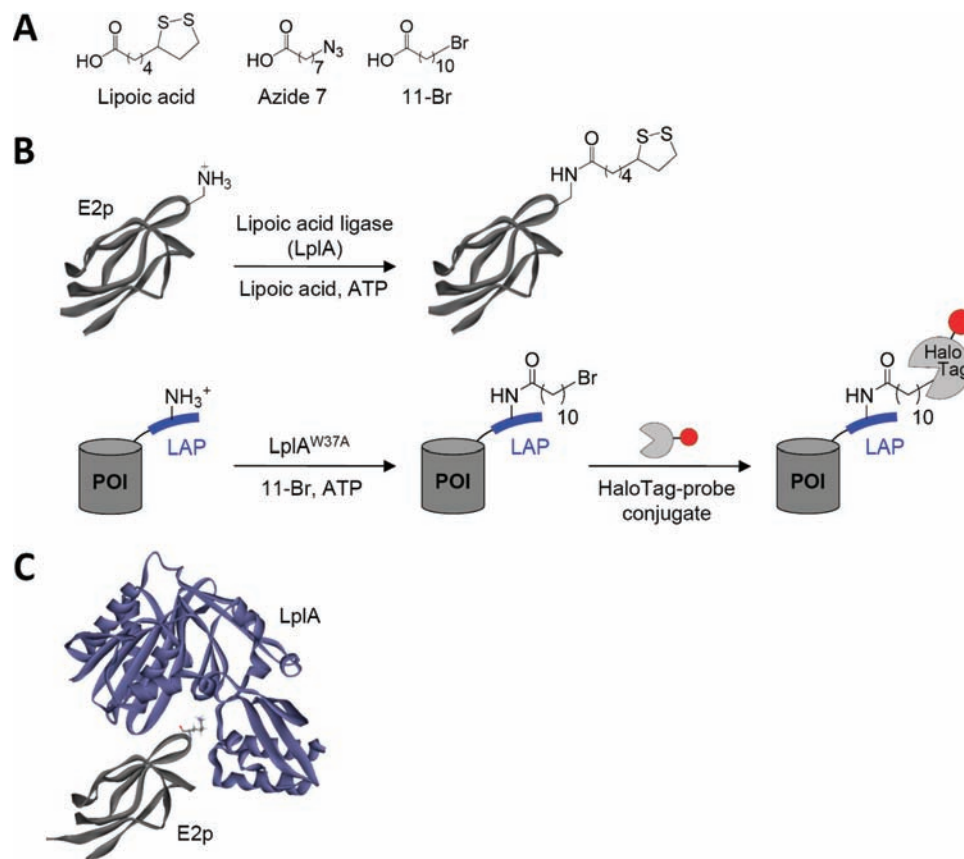
and reduce immunogenicity.<sup>9,10</sup> Our lab is interested in protein minimization to peptides for the purpose of developing new protein labeling technologies. Size minimization of protein tags that direct the targeting of fluorescent probes<sup>11</sup> can greatly reduce problems of tag interference with protein trafficking, folding, and interactions.

Conversion of proteins to peptides without loss of the function of interest, however, is challenging for a number of reasons. First, the function may require secondary structure that is difficult to recapitulate in a peptide. Second, the function may require contributions from multiple, noncontiguous regions of a protein. Third, structural information is not available for many proteins, and in some cases, even the regions that contribute to a protein's relevant activity are not known. Fourth, due to their more flexible structure, peptide binding is often associated with a greater entropic penalty than is protein binding,<sup>12</sup> making it more difficult to engineer high-affinity interactions.

Numerous methods have been used to reduce proteins to peptides. Simple truncation and/or rational design can be successful,<sup>13–15</sup> but is usually associated with at least a partial

- (1) Feng, S. B.; Kasahara, C.; Rickles, R. J.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 12408–12415.
- (2) Rowland, R. R. R.; Schneider, P.; Fang, Y.; Wootton, S.; Yoo, D.; Benfield, D. A. *Virology* **2003**, *316*, 135–145.
- (3) Hall, J. G.; Frieden, C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 3060–3064.
- (4) Zerella, R.; Chen, P. Y.; Evans, P. A.; Raine, A.; Williams, D. H. *Protein Sci.* **2000**, *9*, 2142–2150.
- (5) Toomik, R.; Ek, P. *Biochem. J.* **1997**, *322*, 455–460.
- (6) Cummings, R. T.; Salowe, S. P.; Cunningham, B. R.; Wiltsie, J.; Park, Y. W.; Sonatore, L. M.; Wisniewski, D.; Douglas, C. M.; Hermes, J. D.; Scolnick, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6603–6606.
- (7) Shamoo, Y.; Steitz, T. A. *Cell* **1999**, *99*, 155–166.
- (8) Ghose, R.; Shekhtman, A.; Goger, M. J.; Ji, H.; Cowburn, D. *Nat. Struct. Biol.* **2001**, *8*, 998–1004.

- (9) Borghouts, C.; Kunz, C.; Groner, B. *J. Pept. Sci.* **2005**, *11*, 713–726.
- (10) Sato, A. K.; Viswanathan, M.; Kent, R. B.; Wood, C. R. *Curr. Opin. Biotechnol.* **2006**, *17*, 638–642.
- (11) Lin, M. Z.; Wang, L. *Physiology* **2008**, *23*, 131–141.
- (12) Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. G. *Biochemistry* **1998**, *37*, 16839–16845.
- (13) Fernandez-Suarez, M.; Baruah, H.; Martinez-Hernandez, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.



**Figure 1.** LplA-catalyzed protein and peptide labeling reactions. (A) Natural (lipoic acid) and unnatural (Azide 7 and 11-Br) small-molecule substrates of LplA and its W37 mutant. (B) Natural and engineered ligation reactions. Top: LplA-catalyzed lipoylation of the 9 kD E2p domain of *E. coli* pyruvate dehydrogenase (structure from PDB 1QJO). Bottom: LplA-catalyzed 11-Br ligation onto an engineered LAP (“LplA Acceptor Peptide”), which is genetically fused to any protein of interest (POI). Ligated alkyl bromide can be specifically and covalently modified by HaloTag-fluorophore conjugates.<sup>31</sup> The red circle represents any probe. (C) Model for interaction between LplA (purple, from PDB 1X2H)<sup>33</sup> and E2p (gray, from PDB 1QJO).<sup>34</sup> The lipoylation site on E2p, Lys41, is rendered in stick.

loss of activity and/or specificity. Peptide scanning<sup>16</sup> or high-throughput screening<sup>17–19</sup> approaches are more exhaustive, but library sizes are limited (typically  $10^2$ – $10^5$ ), so it is difficult to identify optimal sequences. Peptide *selections*, on the other hand, can process libraries up to  $10^9$  in size, dramatically increasing the probability of identifying a successful sequence. Accordingly, selections on phage,<sup>20–23</sup> inside bacteria,<sup>24</sup> and on the surface of bacteria,<sup>25</sup> yeast,<sup>26</sup> and mammalian cells<sup>27</sup> have been used to evolve peptides with novel functions.

In this study, our goal was to identify novel, kinetically efficient peptide substrates for *Escherichia coli* lipoic acid ligase (LplA) (Figure 1). LplA is a cofactor ligase that our lab has harnessed for fluorescent protein labeling applications.<sup>13,28</sup> The natural function of LplA is to catalyze ATP-dependent, covalent ligation of lipoic acid (Figure 1A) onto specific lysine side chains of three *E. coli* proteins involved in oxidative metabolism: pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, and the glycine cleavage system.<sup>29</sup> Previously, we showed that LplA and engineered variants can ligate unnatural probes such as an alkyl azide (a functional group handle for fluorophore introduction; Figure 1A),<sup>13</sup> a fluorinated aryl azide photo-cross-linker,<sup>28</sup> bromoalkanoic acid (a ligand for HaloTag;<sup>30</sup> Figure 1A),<sup>31</sup> and a coumarin fluorophore<sup>32</sup> in place of lipoic acid. To utilize these ligation reactions for protein imaging applications, we prepared

- (14) Mitchell, R. D.; Glass, D. B.; Wong, C. W.; Angelos, K. L.; Walsh, D. A. *Biochemistry* **1995**, *34*, 528–534.  
 (15) Viguera, A. R.; Arrondo, J. L. R.; Musacchio, A.; Saraste, M.; Serrano, L. *Biochemistry* **1994**, *33*, 10925–10933.  
 (16) Rudiger, S.; Schneider-Mergener, J.; Bukau, B. *EMBO J.* **2001**, *20*, 1042–1050.  
 (17) Lam, K. S.; Wu, J. Z.; Lou, Q. *Int. J. Pept. Protein Res.* **1995**, *45*, 587–592.  
 (18) Reineke, U.; Volkmer-Engert, R.; Schneider-Mergener, J. *Curr. Opin. Biotechnol.* **2001**, *12*, 59–64.  
 (19) Marani, M. M.; Ceron, M. C. M.; Giudicessi, S. L.; de Oliveira, E.; Cote, S.; Erra-Balsells, R.; Albericio, F.; Cascone, O.; Camperi, S. A. *J. Comb. Chem.* **2009**, *11*, 146–150.  
 (20) Herman, R. E.; Badders, D.; Fuller, M.; Makienko, E. G.; Houston, M. E.; Quay, S. C.; Johnson, P. H. *J. Biol. Chem.* **2007**, *282*, 9813–9824.  
 (21) Chen, I.; Choi, Y. A.; Ting, A. Y. *J. Am. Chem. Soc.* **2007**, *129*, 6619–6625.  
 (22) Sidhu, S. S.; Koide, S. *Curr. Opin. Struct. Biol.* **2007**, *17*, 481–487.  
 (23) Zhou, Z.; Cironi, P.; Lin, A. J.; Xu, Y. Q.; Hrvatin, S.; Golan, D. E.; Silver, P. A.; Walsh, C. T.; Yin, J. *ACS Chem. Biol.* **2007**, *2*, 337–346.  
 (24) Schatz, P. J. *Biotechnology* **1993**, *11*, 1138–1143.

- (25) Dane, K. Y.; Chan, L. A.; Rice, J. J.; Daugherty, P. S. *J. Immunol. Methods* **2006**, *309*, 120–129.  
 (26) Krauland, E. M.; Peelle, B. R.; Wittrup, K. D.; Belcher, A. M. *Biotechnol. Bioeng.* **2007**, *97*, 1009–1020.  
 (27) Wolkowicz, R.; Jager, G. C.; Nolan, G. P. *J. Biol. Chem.* **2005**, *280*, 15195–15201.  
 (28) Baruah, H.; Puthenveetil, S.; Choi, Y. A.; Shah, S.; Ting, A. Y. *Angew. Chem., Int. Ed.* **2008**, *47*, 7018–7021.  
 (29) Cronan, J. E.; Zhao, X.; Jiang, Y. *Adv. Microb. Physiol.* **2005**, *50*, 103–146.  
 (30) Los, G. V.; et al. *ACS Chem. Biol.* **2008**, *3*, 373–382.  
 (31) Liu, D. S.; Phipps, W. S.; Howarth, M.; Puthenveetil, S.; Ting, A. Y. Unpublished work.  
 (32) Baruah, H.; Uttamapinant, C.; White, K. A.; Fernández-Suárez, M.; Puthenveetil, S.; Thompson, S.; Ting, A. Y. Unpublished work.

recombinant fusions of proteins of interest (POIs) to the 9 kD E2p domain of pyruvate dehydrogenase (Figure 1B top).<sup>13</sup> Such fusions could be labeled with high efficiency and specificity by our unnatural probes on the surface and in the cytosol of living mammalian cells.<sup>13,28,31,32</sup>

Even though 9 kD (85 amino acids) E2p is considerably smaller than Green Fluorescent Protein (27 kD) and other protein labeling tags such as HaloTag (33 kD)<sup>30</sup> and SNAP tag (20 kD),<sup>35</sup> we wanted to further reduce its size, to minimize steric interference with POI function. We previously attempted this by rational design of an “LplA Acceptor Peptide” (LAP1),<sup>13</sup> based on the sequence of LplA’s natural protein substrate 2-oxoglutarate dehydrogenase, with a few additional mutations. LAP1 is 17 amino acids long, or 22 amino acids with the recommended linker.<sup>13</sup> We found that LAP1 fusion proteins could be ligated by LplA to some probes (lipoic acid,<sup>13</sup> alkyl azide,<sup>13</sup> and aryl azide<sup>28</sup>) *in vitro* and in cell lysate, but not on the cell surface except under conditions of high LAP1-POI overexpression.<sup>13,28</sup> We could not detect LAP1 labeling in the cytosol (data not shown). Other probes (bromoalkanoic acid and coumarin) could not be ligated appreciably to LAP1 in any context.<sup>31,32</sup> In contrast, E2p fusions could be labeled by all probes on the cell surface and in the cytosol.<sup>13,28,31,32</sup> Since the measured  $k_{\text{cat}}$  values for azide 7 ligation, for instance, are similar for LAP1 and E2p ( $0.048 \pm 0.001 \text{ s}^{-1}$  and  $0.111 \pm 0.003 \text{ s}^{-1}$ , respectively<sup>13</sup>), we attribute the difference in labeling outcomes to the gap in their  $K_{\text{m}}$  values. H-protein of the glycine cleavage system has a  $K_{\text{m}}$  of  $1.2 \mu\text{M}$ ,<sup>36</sup> which is likely to be similar to E2p’s  $K_{\text{m}}$ , due to their sequence and structural similarity.<sup>37</sup> On the other hand, we estimate that the  $K_{\text{m}}$  for LAP1 is  $>200 \mu\text{M}$ , measured by HPLC (data not shown).

For this study, we selected yeast surface display<sup>38</sup> as our platform to evolve a novel peptide substrate for LplA (called “LAP2”), with kinetic properties comparable to those of LplA’s natural protein substrates. We preferred yeast display to other evolution platforms for a number of reasons. Selections in bacterial cytosol<sup>24</sup> do not allow fine adjustment of protein concentrations and selection conditions. Phage display has limited dynamic range, both due to displayed peptide copy number (3–5 on pIII or 2700 on pVIII<sup>39</sup>), and due to the all-or-nothing nature of affinity-based product capture. The limited dynamic range makes it very difficult to enrich kinetically efficient peptide substrates, as we discovered in our phage display evolution of yAP, a peptide substrate for yeast biotin ligase.<sup>21</sup> Mammalian cell surface display is challenging due to the need for viral transfection to control the multiplicity of infection, and the low viability of cells after fluorescence activated cell sorting (FACS).<sup>40</sup>

By careful library design, tuning of selection conditions with the help of a model selection, four rounds of selection with

decreasing LplA concentrations, and additional rational mutagenesis, we engineered a 13-amino acid LAP2 with a  $k_{\text{cat}}$  of  $0.22 \pm 0.01 \text{ s}^{-1}$  and a  $K_{\text{m}}$  of  $13.32 \pm 1.78 \mu\text{M}$  for lipoic acid ligation. The catalytic efficiency ( $k_{\text{cat}}/K_{\text{m}} = 0.99 \mu\text{M}^{-1} \text{ min}^{-1}$ ) is closer to that of LplA’s natural protein substrate H-protein ( $k_{\text{cat}}/K_{\text{m}} = 7.95 \mu\text{M}^{-1} \text{ min}^{-1}$ )<sup>33</sup> than that of LAP1 (est.  $k_{\text{cat}}/K_{\text{m}} < 0.0135 \mu\text{M}^{-1} \text{ min}^{-1}$  for azide ligation).<sup>13</sup> As a consequence of this improvement, we could easily lipoylate cell surface LAP2 fusion proteins, even at low expression levels. We also performed LplA-mediated quantum dot targeting to LAP2-fused LDL receptor. In comparison, quantum dot labeling was undetectable when using the same receptor fused to LAP1.

## Results

**Model Selections.** We designed the selection scheme shown in Figure 2A. A library of LAP variants is displayed on the C-terminus of Aga2p, a cell surface mating agglutinin protein commonly used for yeast display.<sup>38</sup> A c-Myc epitope tag is also introduced to allow measurement of LAP expression levels by immunofluorescence staining. Each of  $10^6$ – $10^8$  yeast cells expresses a single LAP mutant. Three hypothetical LAP mutants (LAP<sub>x</sub>, LAP<sub>y</sub>, and LAP<sub>z</sub>) with diminishing activity toward LplA are shown in Figure 2A. They are collectively labeled by LplA (e.g., with lipoic acid), and ligated probe is detected with a suitable fluorescent reagent (e.g., anti-lipoic acid antibody followed by phycoerythrin-conjugated secondary antibody). Since LAP<sub>x</sub> is the most active mutant in this scheme, yeast cells displaying this mutant should become brightly fluorescent. On the other hand, LAP<sub>y</sub> and LAP<sub>z</sub>-displaying yeast will be dimmer or unlabeled. To normalize for variations in expression level, the yeast pool is also collectively labeled with anti-c-Myc antibody, detected with a secondary antibody conjugated to Alexa Fluor 488, which is easily resolvable from phycoerythrin fluorescence. The double-labeled yeast cells are subjected to two-dimensional fluorescence activated cell sorting. Yeast cells displaying a high ratio of phycoerythrin intensity to Alexa Fluor 488 intensity (sorting gate shown in red in Figure 2A) represent the most efficiently labeled yeast, with the largest fraction of labeled LAPs, and are isolated by the FACS instrument. The captured yeast cells are amplified, sequenced, and subjected to further rounds of selection.

Before initiating selections on a LAP library, we tested and optimized our selection scheme using a model system consisting of mixtures of E2p-expressing yeast and LAP1-expressing yeast. Since LAP1 represents the best that we can achieve by rational design and E2p represents LplA’s natural substrate with evolutionarily optimized  $k_{\text{cat}}/K_{\text{m}}$ , we wished to design a selection that could maximally enrich E2p-yeast over LAP1-yeast. We performed lipoylation of E2p or LAP1 expressed on yeast surface by adding purified LplA, ATP, and lipoic acid to the media. FACS scanning showed that, for a 30 min reaction time, we could obtain the largest difference in signal between E2p-yeast and LAP1-yeast using 300 nM LplA (Figure 2B). Higher LplA concentrations increased LAP1 intensity without increasing E2p intensity, diminishing the difference between them (data not shown). To check the site-specificity of LplA labeling on the yeast surface, we also performed a negative control using an E2p-Aga2p construct with a Lys→Ala mutation at the lipoylation site, and observed no phycoerythrin staining (Figure 2B).

Using 300 nM LplA, we performed 30-min labeling on 1:10, 1:100, and 1:1000 mixtures of E2p-yeast and LAP1-yeast (E2p yeast in the minority). FACS was performed using the red gate

(33) Fujiwara, K.; Toma, S.; Okamura-Ikeda, K.; Motokawa, Y.; Nakagawa, A.; Taniguchi, H. *J. Biol. Chem.* **2005**, *280*, 33645–33651.

(34) Jones, D. D.; Stott, K. M.; Howard, M. J.; Perham, R. N. *Biochemistry* **2000**, *39*, 8448–8459.

(35) Gautier, A.; Juillerat, A.; Heinis, C.; Correa, I. R., Jr.; Kindermann, M.; Beauvais, F.; Johnsson, K. *Chem. Biol.* **2008**, *15*, 128–136.

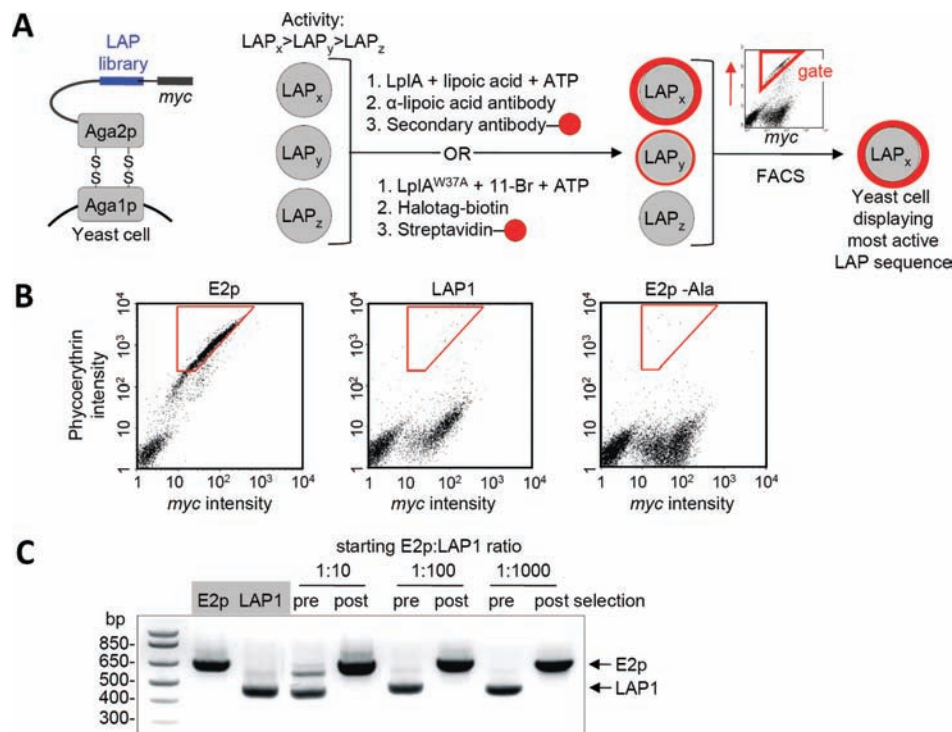
(36) Fujiwara, K.; Suzuki, M.; Okumachi, Y.; Okamura-Ikeda, K.; Fujiwara, T.; Takahashi, E.; Motokawa, Y. *Eur. J. Biochem.* **1999**, *260*, 761–767.

(37) Kim, D. J.; Kim, K. H.; Lee, H. H.; Lee, S. J.; Ha, J. Y.; Yoon, H. J.; Suh, S. W. *J. Biol. Chem.* **2005**, *280*, 38081–38089.

(38) Gai, S. A.; Wittrup, K. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 467–473.

(39) Paschke, M. *Appl. Microbiol. Biotechnol.* **2006**, *70*, 2–11.

(40) Martin, B. R.; Giepmans, B. N.; Adams, S. R.; Tsien, R. Y. *Nat. Biotechnol.* **2005**, *23*, 1308–1314.



**Figure 2.** Yeast display selection scheme and results of model selections. (A) Left: The LAP library (blue) is displayed on the yeast surface as a fusion to Aga2p protein. A C-terminal *myc* epitope is used to quantify LAP expression level. Right: Selection scheme. Yeast cells displaying three sample LAP sequences, with high ( $LAP_x$ ), moderate ( $LAP_y$ ), and low ( $LAP_z$ ) activity are shown. The yeast cells are collectively labeled with lipoic acid or 11-Br probe. The former is detected with anti-lipoic acid; the latter is detected with HaloTag-biotin,<sup>30</sup> followed by streptavidin-fluorophore conjugate. The yeast pool is then sorted (sorting gate in red) on the basis of both ligation extent (red probe intensity) and LAP expression level (c-Myc staining intensity), to enrich the most kinetically efficient LAP peptides. (B) Determination of labeling and sorting conditions for model selections. FACS scatter plots are shown for yeast displaying E2p, LAP1,<sup>13</sup> and E2p-Ala (E2p with a Lys41→Ala mutation), after lipoylation with 300 nM LplA for 30 min, and staining with anti-lipoic acid antibody. The plots show the distribution of single yeast cells as a function of phycoerythrin staining intensity (reflecting extent of lipoylation) and c-Myc staining intensity (reflecting expression level of the Aga2p-LAP fusion). A cell population on the lower left is present in all three samples, and represents untransformed yeast. Optimized sorting gates, used for the model selections, are shown in red. (C) Results of model selections. E2p-displaying yeast and LAP1-displaying yeast were mixed at ratios of 1:10, 1:100, or 1:1000, labeled with 300 nM LplA for 30 min, and sorted. PCR analysis gives the ratio of yeast populations pre- and post-selection. E2p enrichment factor was  $>10^3$ .

shown in Figure 2B. We used a PCR assay to determine the ratio of yeast before and after a single round of selection, capitalizing on the different sizes of the E2p and LAP1 genes. Figure 2C shows that for all starting mixtures, the selection protocol enriched E2p yeast and depleted LAP1 yeast so completely that the latter could not be detected. We conclude that our selection can enrich kinetically efficient LplA substrates (e.g., E2p) over active but inefficient substrates (e.g., LAP1) by  $>1000$ -fold in a single round.

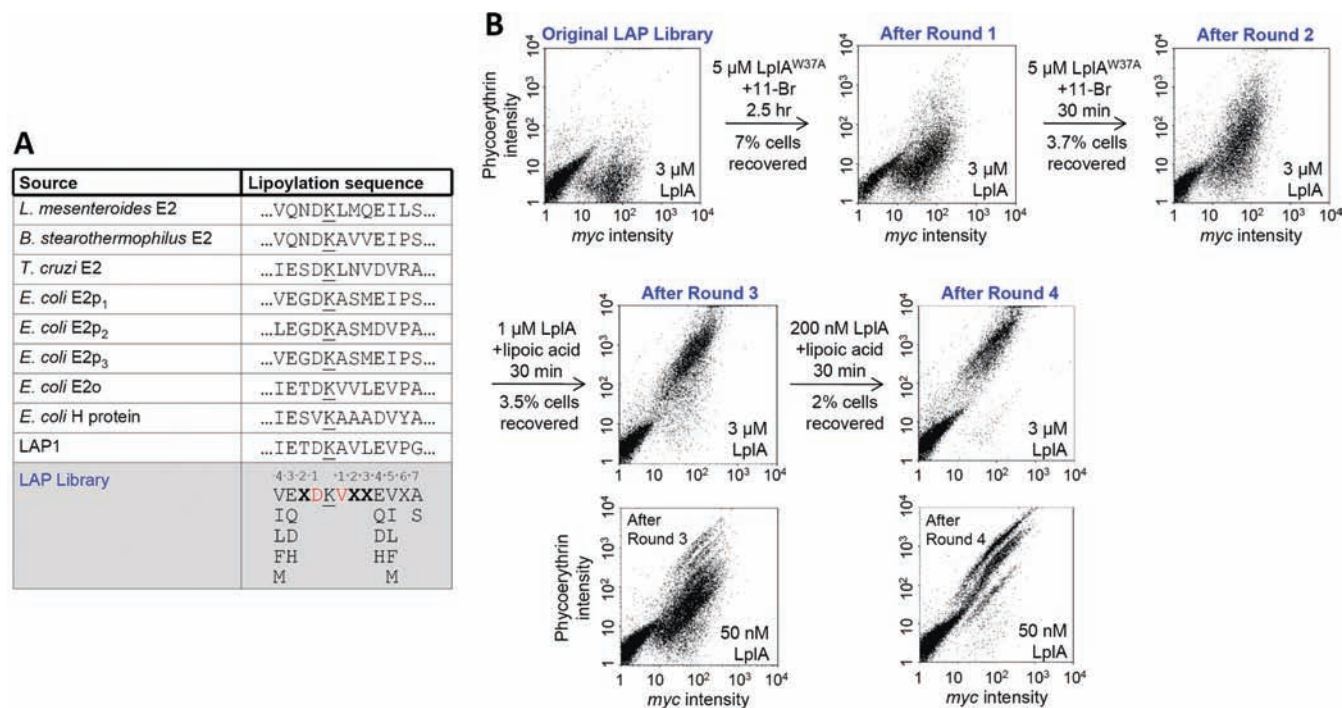
In addition to a selection based on lipoylation, we wished to develop a selection scheme based on ligation of an unnatural probe. This would serve two purposes. First, by using two different sets of probes and detection reagents in alternating rounds of selection, we could minimize the possibility of inadvertently isolating LAPs with affinity for one of our detection reagents. Second, we could increase the probability of isolating a LAP sequence that would be effective not just for lipoylation, but also for ligation of unnatural probes such as photo-cross-linkers and fluorophores.

In separate work,<sup>31</sup> we have identified mutants of LplA that catalyze ligation of bromoalkanoic acids. Once ligated to E2p or LAP, such probes can covalently react with the commercial protein HaloTag,<sup>30</sup> which is derived from a microbial dehalogenase. Thus, we have used 11-bromoundecanoic acid (11-Br, Figure 1A) to target HaloTag-conjugated fluorophores to specific cell surface proteins (Figure 1B, bottom).<sup>31</sup> For yeast display selections, we labeled cell surface E2p or LAP1 with the

Trp37→Ala mutant of LplA mutant ( $LplA^{W37A}$ ), ATP, and the 11-Br probe. We then detected ligated bromoalkane with HaloTag protein, conjugated to biotin, and detected that in turn with streptavidin conjugated to phycoerythrin (Figure 2A). As with the lipoylation assay, we detected a large difference in phycoerythrin staining between E2p-yeast and LAP1-yeast, using 500 nM mutant LplA, and no labeling of E2p (Lys→Ala)-yeast (data not shown). Thus, 11-Br probe is also suitable for LAP selections on yeast cells.

**Construction of LAP Library and Yeast Display Selections.** We wished to shorten LAP, from LAP1's 17-22 amino acids,<sup>13,28</sup> and thus opted for a 12-mer peptide library. With complete randomization of the 11 residues flanking the central Lys, the theoretical diversity would be  $\sim 10^{14}$ , far greater than the experimentally achievable library size, which is limited by yeast transformation efficiency to  $10^7$ – $10^8$ .<sup>41</sup> Thus, we decided to create a *partially* randomized 12-mer library, guided by alignments of natural lipoate acceptor protein sequences, the NMR structure of E2p,<sup>34</sup> and the structure of a related biotin acceptor domain in complex with biotin ligase.<sup>42</sup>

We aligned the sequences of 250 naturally lipoylated proteins ("lipoate acceptor proteins") from  $>100$  distinct species. The five lipoyl domains from *E. coli* (present in LplA's three acceptor proteins), along with lipoyl domains from three other species are shown in Figure 3A. Several trends were apparent from the alignment: (1) the  $-1$  Asp is highly conserved; (2) positions  $+1$ ,  $+5$ , and  $-4$  are usually hydrophobic; (3) Glu and



**Figure 3.** Library design and selection results. (A) Table showing sequences of natural LplA protein substrates, our previous rationally designed LAP1,<sup>13</sup> and our LAP library. Lysine modification sites are underlined. For the LAP library, we fixed positions  $-4$  and  $+5$  as hydrophobic amino acids (Val, Ile, Leu, Phe, Met), positions  $-3$  and  $+4$  as polar amino acids (Glu, Asp, Gln, His), and position  $+7$  as Ser or Ala. Red-colored positions are partially randomized (39% Asp or 49% Val). X represents any amino acid. (B) Results of four rounds of selection. Selection conditions, including small-molecule substrates used for labeling, are given above each arrow. To analyze amplified yeast pools following each round of selection, uniform lipoylation conditions were used (given in the lower right of each scatter plot). Yeast pools from rounds 3 and 4 were additionally analyzed under milder conditions, with 50 nM LplA.

Asp are enriched at positions  $-3$  and  $+4$ ; and (4) position  $+7$  is usually Ser or Ala. We introduced these preferences into our LAP library design, shown in Figure 3A.

In addition, we used structural data to inform our LAP library design. NMR structures are available for several lipoate acceptor domains.<sup>34,43–45</sup> All of them show that the lipoylated lysine is presented at the tip of a  $\beta$ -hairpin turn. Though this is a challenging structure to recapitulate in a peptide, we took a cue from the structure of *E. coli* E2p, which shows that the  $-1$  Asp side-chain hydrogen bonds with backbone amide N–H groups of both the central lysine and  $+1$  Ala (Figure S1, Supporting Information).<sup>34</sup> To promote this loop-favoring interaction, we installed Asp at the  $-1$  position with 39% frequency in our LAP library (Figure 3A).

There is no co-crystal structure of a lipoate acceptor domain with LplA, to indicate which residues might be important for interactions with the enzyme. However, lipoate domains are structurally similar to biotin acceptor domains,<sup>46,47</sup> and LplA is structurally related to biotin ligase as well.<sup>48</sup> The co-crystal structure of *Pyrococcus horikoshii* biotin ligase with its biotin acceptor protein shows a hydrogen bond between the  $+4$  Glu

of the acceptor and Lys27 of the enzyme.<sup>42</sup> In addition, the authors of the *Thermoplasma acidophilum* LplA structure created a computationally docked model of their enzyme with E2p.<sup>37</sup> The docked structure also predicts a hydrogen bond between the  $+4$  Glu of E2p and Lys155 of the enzyme, which corresponds to Lys143 of *E. coli* LplA. Figure 1C shows our docked model of *E. coli* LplA with its E2p lipoate acceptor. Because these structures and models suggest that  $+4$  Glu is important for interactions with LplA, we restricted the  $+4$  position of our LAP library to polar residues (Glu, Asp, Gln, and His) to promote intermolecular hydrogen bonding (Figure 3A).

The LAP library was cloned by Klenow-mediated fill-in of a synthetic oligonucleotide library. The insert was introduced into pCTCON2,<sup>41</sup> containing Aga2p and the c-Myc tag, by homologous recombination. Our yeast transformation efficiency was  $\sim 10^7$ ,  $10^3$ -fold under our theoretical diversity of  $\sim 10^{10}$ .

For reasons described above, we used both lipoic acid and bromoalkanoic acid (11-Br) probes for our selections. The latter was used for the first two rounds of selection, and lipoic acid was used for rounds 3 and 4 (Figure 3B). To successively increase selection stringency, we decreased LplA concentration throughout the selection, from 5  $\mu$ M in rounds 1 and 2, to 1  $\mu$ M in round 3, to 200 nM LplA in the final round. Reaction times were 2.5 h for the first round, and 30 min for all subsequent rounds.

To compare the activities of recovered yeast from each round of selection, we re-amplified the yeast pools and labeled them with lipoic acid under identical conditions. Figure 3B shows that c-Myc intensities remain constant, while phycoerythrin intensities gradually increase. With 3  $\mu$ M LplA, yeast recovered from rounds 3 and 4 looked identical; thus we also performed analysis under milder conditions (Figure 3B). With 50 nM LplA,

(41) Chao, G.; Lau, W. L.; Hackel, B. J.; Sazinsky, S. L.; Lippow, S. M.; Wittup, K. D. *Nat. Protoc.* **2006**, *1*, 755–768.

(42) Bagautdinov, B.; Matsuura, Y.; Bagautdinova, S.; Kunishima, N. *J. Biol. Chem.* **2008**, *283*, 14739–14750.

(43) Dardel, F.; Davis, A. L.; Laue, E. D.; Perham, R. N. *J. Mol. Biol.* **1993**, *229*, 1037–1048.

(44) Howard, M. J.; Chauhan, H. J.; Domingo, G. J.; Fuller, C.; Perham, R. N. *J. Mol. Biol.* **2000**, *295*, 1023–1037.

(45) Ricaud, P. M.; Howard, M. J.; Roberts, E. L.; Broadhurst, R. W.; Perham, R. N. *J. Mol. Biol.* **1996**, *264*, 179–190.

(46) Reche, P.; Perham, R. N. *EMBO J.* **1999**, *18*, 2673–2682.

(47) Cui, G. F.; Nan, B. Y.; Hu, J. C.; Wang, Y. P.; Jin, C. W.; Xia, B. *J. Biol. Chem.* **2006**, *281*, 20598–20607.

(48) Reche, P. A. *Protein Sci.* **2000**, *9*, 1922–1929.

it can be seen that yeast cells from round 4 are more extensively labeled by lipoic acid than yeast cells from round 3.

**Characterization of Selection Results.** The sequences of selected LAP clones from rounds 2, 3, and 4 are shown in Figure S2A,B, Supporting Information. In addition, graphical representations of amino acid frequencies are shown in Figure S2C, Supporting Information. We observed the following trends: (1) In general, selected LAP clones had interlaced hydrophobic and negatively charged side chains flanking the central lysine. (2) Position +2, which was fully randomized in the LAP library, became 100% Trp. This enrichment was apparent after just a single round of selection. (3) Position +3, which was also fully randomized, showed a preference for aromatic side chains. (4) Positions -3 and +4 were limited to one of 4 polar side chains in the LAP library. Position -3 became 100% Glu. Position +4 became exclusively Glu or Asp, already by round 2. (5) Positions -4 and +5 were limited to hydrophobic residues in the LAP library. Position +5 did not converge, but position -4 became 100% Phe. (6) Position +1, which was 49% Val in the library, became 100% Val. We note that after round 4, we observed only 4 distinct clones, and further rounds of selection did not reveal any additional diversity.

A powerful feature of FACS-based selection is its dynamic range. For a single round of selection, different sorting gates can be used, and the sequences of clones obtained via different gates can be compared, to infer sequence-activity relationships. For round 4, in addition to our standard high phycoerythrin gate ("Gate A"), we also collected yeast from a slightly lower gate ("Gate B"). Figure S2B, Supporting Information, shows that the major difference between Gate A clones and Gate B clones is the presence of Phe at the -4 position in Gate A clones. We surmised that the selection of -4 Phe may account for much of the jump in LAP activity between rounds 3 and 4. Indeed, when we mutated the -4 Phe of one of the Gate A clones, LAP4.1, to Val, its activity in a yeast surface lipoylation assay dropped to a level comparable to the Gate B clones (Figure S3, Supporting Information).

We utilized the information from Gate A and Gate B clones (Figure S2C, Supporting Information) to rationally design a new LAP sequence, called "LAP2". Since Gate A clones showed clear amino acid preferences at positions -4, -3, -2, +1, +2, +4, +5, and +7, we introduced these preferred residues into our LAP2 sequence. Positions -1, +3, and +6 did not show consensus in Gate A clones, so we based these amino acids in LAP2 on preferences seen in the Gate B clones. We characterized this rationally designed LAP2 alongside the four evolved LAP clones from round 4, in cell-based and *in vitro* assays, described below.

**Comparison of LAP Sequences.** To compare the round 4 LAP sequences and LAP2, we created genetic fusions to CFP-TM (cyan fluorescent protein fused to a transmembrane helix from PDGF receptor)<sup>13</sup> for mammalian cell surface expression, and HP1 (heterochromatin protein 1)<sup>13</sup> for bacterial expression. In all constructs, an N-terminal glycine from the Aga2p fusion was carried over, making the total LAP length 13 amino acids.

First, we compared the surface expression levels of the LAP fusions in HeLa mammalian cells. Whereas LAP4.1, LAP4.2, and LAP2 gave clear cell surface expression, both LAP4.3 and LAP4.4 showed poor expression (data not shown). We surmised that LAP4.3 expression might be hindered by its +6 Cys, due to intermolecular disulfide bond formation in the oxidizing secretory pathway. Since Gate B clones showed a preference for Asp at this position, we prepared a point mutant of LAP4.3

with a +6Cys→Asp mutation (LAP4.3D). Figure S4, Supporting Information shows that LAP4.3D gives improved cell surface expression compared to LAP4.3, as indicated by the pattern of CFP fluorescence. In addition, cell surface lipoylation with exogenous LplA gives a strong signal with LAP4.3D-CFP-TM, whereas little signal is detected under the same conditions with LAP4.3-CFP-TM. *E. coli* expression of the HP1 fusion protein also improved significantly upon introduction of the +6Cys→Asp mutation in LAP4.3. Based on these observations, we carried LAP4.3D into subsequent analyses, and we did not characterize LAP4.3 or LAP4.4 any further.

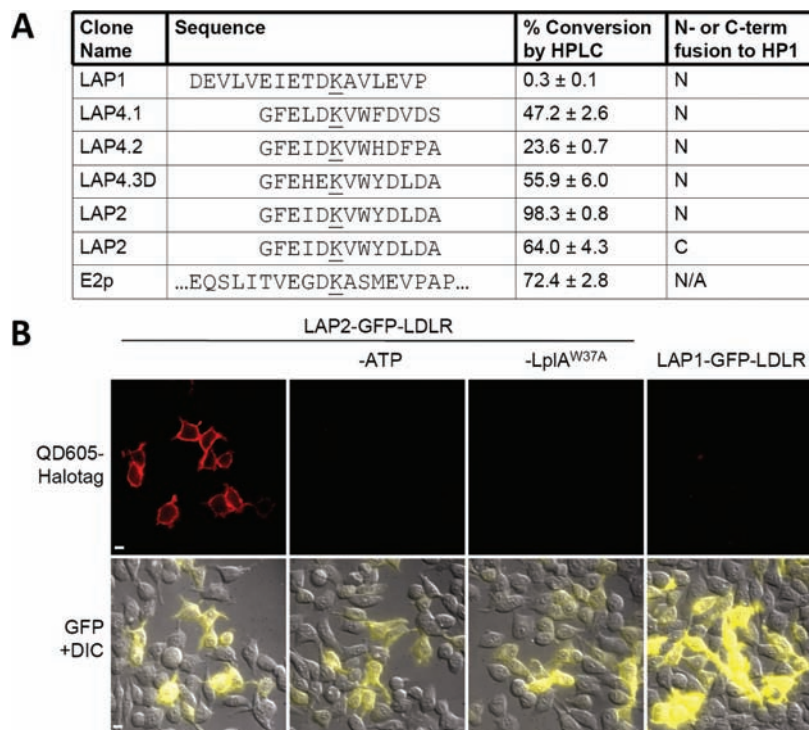
Second, we compared the LAPs in a cell surface lipoylation assay (Figure S5, Supporting Information). CFP-TM fusion constructs were expressed in human embryonic kidney 293 (HEK) cells, and lipoylation was carried out by purified LplA enzyme added to the media. After 10 min of labeling, lipoylated cell surface proteins were imaged using anti-lipoic acid antibody. Figure S5A, Supporting Information, shows representative images of labeled E2p, LAP2, and LAP1.<sup>49</sup> Whereas E2p and LAP2 are lipoylated to a similar degree, labeling is not detected under these conditions for LAP1. To quantitatively compare the labeling efficiencies of all the LAP sequences, we plotted lipoylation signal (as measured by antibody staining intensity) against CFP signal for single cells. Average signal ratios listed in Figure S5B, Supporting Information, indicate that LAP2 is labeled more efficiently than the other LAP sequences, and is comparable even to E2p.

Third, the LAP sequences were compared in an intracellular labeling assay. In separate work, we have engineered a coumarin fluorophore ligase for labeling of recombinant proteins in living mammalian cells.<sup>32</sup> To compare the LAP sequences using this assay, we prepared fusions to nuclear-localized yellow fluorescent protein (YFP), and labeled transfected cells with the coumarin probe for 10 min. Afterward, images were analyzed by plotting mean single cell coumarin intensities against mean single cell YFP intensities. Figure S6, Supporting Information, shows that LAP2 is labeled more efficiently than the other LAP sequences in the cytosol, and gives even higher signal intensities than E2p, at high expression levels.

Fourth, we compared the LAP sequences *in vitro* in an HPLC assay,<sup>13</sup> after expressing and purifying the HP1 fusion proteins<sup>13</sup> from bacteria. Figure 4A shows the percent conversion to lipoylated product under identical reaction conditions. As in the cellular assays, LAP2 is the best sequence. When fused to the C- rather than N-terminus of HP1, the activity of LAP2 decreased somewhat, but was still higher than all other LAP sequences at the N-terminus. We also performed HPLC assays using other probes (azide 7, 11-Br, and coumarin) and found that LAP2 was the best substrate for these (data not shown).

**Characterization of LAP2 and Application to Receptor Labeling.** Using HPLC to quantify product formation, we measured the  $k_{\text{cat}}$  and  $K_{\text{m}}$  values for LplA-catalyzed lipoylation of a synthetic LAP2 peptide (without an attached fusion protein). Figure S7, Supporting Information, shows that the  $k_{\text{cat}}$  is  $0.22 \pm 0.01 \text{ s}^{-1}$ , slightly lower than that of E2p ( $k_{\text{cat}} 0.253 \pm 0.003 \text{ s}^{-1}$ ). The  $K_{\text{m}}$  is  $13.32 \pm 1.78 \mu\text{M}$ , closer to that of LplA's natural substrate H-protein ( $K_{\text{m}} 1.2 \mu\text{M}$ <sup>33</sup>) than that of LAP1 (est.  $K_{\text{m}} >200 \mu\text{M}$ ; data not shown).

(49) We note that the surface expression levels of TM fusions to LAP peptides are ~2-fold lower than TM fusions to E2p. However, expression levels of intracellular proteins are similar, whether fused to a LAP sequence or E2p (Figure S6, Supporting Information).



**Figure 4.** Comparison of LAP clones and application to cell surface quantum dot tagging. (A) Various LAP sequences were compared to E2p protein, by lipoylation with 50 nM LplA for 1 h. Product was detected by HPLC. All LAPs were tested as fusions to the N- or C-terminus of carrier protein HP1, as indicated.<sup>13</sup> Error bars, ± 1 s.d. (B) HEK cells expressing LAP2 or LAP1-fused LDL receptor were labeled with LplA<sup>W37A</sup> and 11-Br for 5 min, followed by QD605-HaloTag<sup>31</sup> for 5 min. QD605 emission is shown in the top row. Merged GFP and DIC (differential interference contrast) images are shown in the bottom row. Negative controls are shown with ATP or LplA omitted. Scale bars, 10 μm.

To utilize LAP2 for receptor imaging, we prepared a fusion to the low density lipoprotein (LDL) receptor. LAP2-LDL receptor expressed in HEK cells was labeled with LplA<sup>W37A</sup> and 11-Br probe. Ligated bromoalkane was derivatized with HaloTag-conjugated quantum dot 605 (QD605). Figure 4B shows specific QD605 labeling of LAP2-LDL receptor at the cell surface. Omission of ATP or LplA eliminates labeling. The same experiment performed with LAP1-fused LDL receptor did not produce any detectable QD605 signal.

Often, we use LplA labeling in conjunction with biotin ligase (BirA) labeling, for two-color imaging applications.<sup>13,31</sup> We used HPLC to test the cross-reactivity of LAP2 with BirA and found no biotinylation after a 12 h reaction with 5 μM BirA (data not shown).

## Discussion

In summary, we have engineered a new peptide substrate for LplA using a novel selection platform based on yeast display. The peptide, LAP2, is lipoylated with a  $k_{cat}$  similar to that of LplA's protein substrate E2p, and has a  $K_m$  much closer to that of LplA's protein substrates than that of our previous rationally designed LAP1.<sup>13</sup> The consequence of this improvement in kinetic efficiency is the ability to label peptide-tagged cell surface receptors with unnatural probes, even at low or medium receptor expression levels. In other work, LAP2 also allows fluorophore tagging of intracellular proteins.<sup>32</sup> In contrast, LAP1 fusions are difficult to label at the cell surface,<sup>13,28</sup> and impossible to label inside of living cells.<sup>32</sup> LAP2 is also shorter than LAP1 (13 amino acids instead of 17–22 amino acids) and can be recognized by LplA at the N-terminus, C-terminus, and internally.<sup>32</sup>

Comparing LAP2 to LplA's natural protein substrates, the negatively charged residues at positions −1, −3, and +4, and

the hydrophobic residues at positions −4 and +5 are shared. Since −1 Asp of E2p may promote loop formation (Figure S1, Supporting Information), and +4 Glu in E2p may interact with Lys143 in LplA's binding pocket (see above), LAP2 may interact with LplA in a manner similar to E2p. When overlaying the LAP2 sequence onto the E2p NMR structure (Figure S1, Supporting Information),<sup>34</sup> the −4 Phe and the +3 Tyr are positioned to interact in an intramolecular manner. We speculate that this interaction may help to stabilize LAP2 in a loop conformation that promotes high affinity binding to LplA. Interestingly, the engineered 15-amino acid acceptor peptide for biotin ligase<sup>50</sup> also contains aromatic side chains at these two positions. We also noticed that the +2 Trp that emerged in our selections may be positioned to interact with a hydrophobic patch on the LplA surface that includes Phe24.

Our study also introduces a new selection scheme for evolution of peptide substrates. Previously, yeast display has been used to evolve enzyme specificity,<sup>35,51</sup> binding peptides,<sup>26</sup> and binding proteins,<sup>38</sup> but, to our knowledge, no enzymatic substrates have been evolved by this method. The appeal of yeast display for enzyme substrate evolution lies in its dynamic range: up to 10<sup>4</sup>–10<sup>5</sup> copies of peptide can be displayed on the surface of each yeast cell,<sup>41</sup> and FACS sorting allows fractionation of yeast into distinct pools based on the extent of surface peptide modification. In contrast, phage display has far more limited dynamic range due to the low copy number of displayed peptides, and the all-or-nothing nature of affinity-based product capture. As a consequence, we previously used two generations

(50) Beckett, D.; Kovaleva, E.; Schatz, P. J. *Protein Sci.* **1999**, *8*, 921–929.

(51) Antipov, E.; Cho, A. E.; Wittrup, K. D.; Klivanov, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17694–17699.

of phage display selections (as opposed to the single generation of selections used here) to produce a peptide substrate for yeast biotin ligase with a  $k_{cat}/K_m$  of only  $0.00078 \mu\text{M}^{-1} \text{min}^{-1}$ ,<sup>21</sup> >1000-fold worse than the  $k_{cat}/K_m$  we obtained here for LAP2. Yin et al. have also used phage display to evolve peptide substrates for phosphopantetheinyl transferases, and obtained  $K_m$  values in the 51–117  $\mu\text{M}$  range, with  $k_{cat}/K_m$  in the range of  $0.015\text{--}0.19 \mu\text{M}^{-1} \text{min}^{-1}$ .<sup>23</sup> Again, these values are poorer than the corresponding values for LAP2. Our selection scheme should be generalizable to other classes of enzyme substrates, such as those for kinases and glycosyltransferases, as long as the enzymatic products can be detected by fluorescence.

Future work will involve the engineering of even shorter LAP sequences, performing biochemical assays and crystallography to determine the mode of LAP binding to LplA, and evolving orthogonal LAP/LplA pairs for multicolor imaging applications.

## Materials and Methods

**Cloning of Aga2p Fusions to LAP1 and E2p for Yeast Display (Figure 2).** The E2p gene was amplified from E2p-CFP-TM<sup>13</sup> using the primers E2p-NheI-PCR (5'GCATCGCTAGCATGGCTATC-GAAATCAAAGTACCGG; incorporates an *NheI* site) and E2p-BamHI-PCR (5'GGTGAGGATCCCGCAGGAGCTGCCGCGAG; incorporates a *BamHI* site). The resulting PCR product was digested with *NheI* and *BamHI* and ligated in-frame to *NheI/BamHI*-digested pCTCON2 vector.<sup>41</sup> To clone the Aga2p fusion to LAP1, we hybridized the oligos LAP1-NheIBamHI-F (5'CTAGCGA-CGAAAGTACTGGTTGAAATCGAAACCGACAAAGCAGTTC-TGGAAGTACCGGGCGGTGAGGAGGAGG) and LAP1-NheIBamHI-R (5'GATCCCTCCTCTCACCGCCGGTACTTCCAGAAGTCTTT-GTCGGTTTCGATTTCAACCAGTACTTTCGTTCG). The annealed oligos encode the 22-amino acid LAP1 sequence DEVLVEIETD-KAVLEVPGGEEE.<sup>13</sup> We then ligated the duplex DNA in-frame to *NheI/BamHI*-digested pCTCON2 vector. The E2p-Ala mutant was generated by Lys40→Ala mutagenesis using the QuikChange oligo 5' GATCACCGTAGAAGGCGACGCTGCTTCTATGGA-AGTTCGGC and its reverse complement.

**Model Selections on Yeast with LAP1 and E2p (Figure 2).** Aga2p-E2p and Aga2p-LAP1 plasmids were transformed into *Saccharomyces cerevisiae* EBY100 using the Frozen-EZ Yeast Transformation II kit (Zymo Research). After transformation, cells were grown in SDCAA media<sup>41</sup> at 30 °C with shaking for 20 h. The culture was then diluted to a cell density of  $10^6$  cells/mL in SGCAA media<sup>41</sup> to induce protein expression for 20 h with shaking at room temperature. Cells were harvested by centrifugation and washed with PBSB (phosphate buffered saline, pH 7.4 + 0.5% BSA).

To lipoylate the yeast,  $10^6\text{--}10^7$  cells were pelleted at 14,000g for 30 s in a 1.5 mL Eppendorf tube, then resuspended in 100  $\mu\text{L}$  of PBSB. To these cells, 750  $\mu\text{M}$  ( $\pm$ )- $\alpha$ -lipoic acid, 300 nM LplA, 3 mM ATP, and 5 mM magnesium acetate were added. The cells were incubated on a rotator for 30 min at 30 °C. After washing the cells once with PBSB, cells were incubated with rabbit anti-lipoic acid antibody (1:300 dilution, Calbiochem) and mouse anti-c-Myc antibody (1:50 dilution, Calbiochem) for 40 min at 4 °C. The cells were washed again with PBSB followed by incubation with phycoerythrin-anti-rabbit antibody (1:100 dilution, Invitrogen) and Alexa Fluor 488-anti-mouse antibody (1:100 dilution, Invitrogen) for 40 min at 4 °C. Finally, cells were rinsed twice with PBSB and resuspended in 600  $\mu\text{L}$  of PBSB for FACS analysis on a FACScan instrument, or FACS sorting on an Aria FACS instrument, both from BD Biosciences, and housed in the Koch Institute flow cytometry core facility.

We note that for c-Myc tag detection, we initially used a chicken anti-c-Myc antibody. However, we found that anti-chicken antibody cross-reacts with rabbit antibodies, and thus we switched to mouse anti-c-Myc antibody, which gives a lower signal, but does not bind to the rabbit anti-lipoic acid antibody.

To implement the model selections, E2p-displaying yeast and LAP1-displaying yeast were combined in various ratios. A total of  $10^7$  cells were lipoylated as described above in 100  $\mu\text{L}$  PBSB. Following labeling, cells were sorted using a typical polygonal gate as shown in Figure 2B. We recovered ~5% of cells from the 1:10 mixture of E2p:LAP1, 0.5% of cells from the 1:100 mixture, and <0.1% of cells from the 1:1000 mixture. Collected cells were amplified in SDCAA media for 24–48 h. Plasmids were isolated using Zymoprep II (Zymo Research). For PCR analysis of enrichment factors, the primers pctPCR•F (5'GCGGTTCTCACCCCT-CAACAAC) and pctPCR•R (5'GTATGTGTAAGTTGGTAACG-GAACG) were used.

**Cloning of LAP Library (Figure 3A).** We ordered from IDT (Integrated DNA Technologies) a partially randomized oligo with the following sequence: 5'AAATAAGCTTTTGTTCGGATCCNG-MMNNNANNTSMNNMNNAACTTTATCMNNNTSNANTCC-GCTAGCCGACCCCTCC. Underlined nucleotides were synthesized from mixtures containing 70% of the indicated base + 10% of each of the other bases. N designates an equimolar mixture of all bases. S designates a 1:1 mixture of G and C. M designates a 1:1 mixture of A and C.

This oligo was annealed with another oligo, Con2For•F (5'CTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCG-GAGGCGAGGGTCCGCTAGCGGA), which overlaps with both pCTCON2 vector and the library oligo. The 5' overhangs were filled in using Klenow polymerase. The resulting product was PCR-amplified using the primers Con2For•F and Con2Rev•R (5'TAT-CAGATCTCGAGCTATTACAAGTCTCTTCAGAAATAAGC-TTTTGTTCGGATCC). Meanwhile, pCTCON2 vector was prepared by digestion with *NheI* and *BamHI*, and gel-purified. PCR insert and pCTCON2 vector were transformed together into *S. cerevisiae* EBY100 (Invitrogen) by electroporation as described by Colby et al.<sup>52</sup> Homologous recombination occurred inside the yeast. Serial dilutions of transformed yeast were plated on SDCAA plates and colonies were counted, to determine transformation efficiency.

**Yeast Display Selection on LAP Library (Figure 3B).** Yeast displaying the LAP library were prepared as described above (see Model Selections). The cells ( $\sim 7 \times 10^7$ ) were washed and resuspended in 700  $\mu\text{L}$  of PBSB. For the first round, HaloTag labeling was performed. Cells were combined with 1 mM 11-Br, 5  $\mu\text{M}$  LplA(W37A), 3 mM ATP, and 5 mM magnesium acetate for 2.5 h at 30 °C. After washing with PBSB, 700 nM biotinylated-HaloTag protein<sup>31</sup> was incubated with the cells in 50  $\mu\text{L}$  of PBSB for 30 min at 30 °C. HaloTag protein was biotinylated by EZ-Link Sulfo-NHS-LC-Biotin (sulfosuccinimidyl-6-(biotinamido)hexanoate) (Thermo Fisher Scientific) as described by the manufacturer. Then, cells were rinsed once with PBSB and labeled with streptavidin-phycoerythrin (1:100 dilution, Jackson ImmunoResearch) for 40 min at 4 °C. For detection of the c-Myc tag, chicken anti-c-Myc antibody (1:200 dilution, Invitrogen) and Alexa Fluor 488-anti-mouse antibody (1:100 dilution, Invitrogen) were used. Labeled cells were rinsed twice with PBSB and resuspended in 1 mL of PBSB for FACS sorting. After sorting, collected yeast cells were amplified in SDCAA media at 30 °C for 36–48 h and induced with SGCAA media at 30 °C for 20 h, for the next round.

Rounds 2–4 were implemented with 11-Br or lipoic acid labeling, under the conditions indicated in Figure 3B. Lipoylation was carried out as described above under Model Selections.

**Analysis of Yeast Pools after Each Round of Selection (Figure 3B).** Yeast harvested from each round of selection were amplified and induced as described above. All pools were then treated identically with 3  $\mu\text{M}$  LplA or 50 nM LplA, 750  $\mu\text{M}$  ( $\pm$ )- $\alpha$ -lipoic acid, and 3 mM ATP for 30 min. 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

(52) Colby, D. W.; Kellogg, B. A.; Graff, C. P.; Yeung, Y. A.; Swers, J. S.; Wittrup, K. D. *Protein Eng.* **2004**, *388*, 348–358.



concentration, LAP genes were PCR-amplified from plasmid using the primers PctPCR-F and PctPCR-R (sequences under Model Selections). Sequencing was completed using the primer PctSeq (5'GGCAGCCCCATAAACACAC).

**Cloning and Expression of LAP-HP1 Fusion Proteins (Figure 4A).** First, an MfeI restriction site was introduced into our previously described<sup>13</sup> LAP1-HP1 expression plasmid, at the C-terminal end of the LAP1 sequence, using the QuikChange primer 5' AAG-CAGTTCTGGAAGTACCGCAATTGGGCGGTGAGGAGGAG-TACGCC and its reverse complement. We then annealed the forward and reverse oligos shown in Table S1, Supporting Information, and ligated the duplex DNA in-frame into *NheI/MfeI*-digested LAP1-(MfeI)-HP1 vector. The vector introduced a C-terminal His<sub>6</sub> tag. Bacterial expression and purification were carried out as previously described.<sup>13</sup>

C-terminal fusion of LAP2 to HP1 was performed by annealing LAP2-C forward and reverse oligos (Table S1, Supporting Information), and ligating the duplex in-frame to *NdeI/BamHI* digested pET15b vector, which introduces an N-terminal His<sub>6</sub> tag.

**Comparison of LAP Clones by HPLC Assay (Figure 4A).** To compare the labeling efficiencies of the different LAP-HP1 fusion proteins, we assembled labeling reactions as follows: 50 nM LplA, 60 μM LAP-HP1 or E2p, 750 μM (±)-α-lipoic acid, 3 mM ATP, and 5 mM magnesium acetate in Dulbecco's phosphate buffered saline (DPBS). Reactions were incubated at 30 °C for 1 h, and then quenched with 180 mM EDTA (final concentration). The extent of conversion to lipoylated product was determined by HPLC as described in previous work.<sup>13,28</sup>

**Cloning of LAP Fusion Proteins for Mammalian Expression.** Three QuikChange mutations were made on the published pEGFP-LAP-LDLR construct.<sup>13</sup> 5'GAAGTACCATCAGCAGACGGC-CAATTGACTGTGAGCAAGGGCGAGG and its reverse complement were used to introduce MfeI site to 3' end of LAP1. Subsequently, 5'GCACCTCGTTCTATCGATAACGCGTACCATG-GGGCCTGGGGC and its reverse complement were used to mutate upstream (outside of the gene) *NheI* site to *MluI*. A new *NheI* site was then introduced to 5' end of LAP1 using 5'CTG-CAGTTGGCGACAGAAGTGCTAGCGACGAAGTACTGGTT-GAAATC and its reverse complement. This expression vector was named LAP1-GFP-LDLR. LAP2-GFP-LDLR was obtained by annealing LAP2 forward and reverse oligos used for LAP2 HP1 fusion protein and ligating the duplex DNA in-frame into *NheI/MfeI*-digested LAP1-GFP-LDLR.

LAP2-CFP-TM was generated by annealing LAP2-*BgIII*AscI-F (5'GATCTGGCTTCGAGATCGACAAGGTGTGGTACGACCTG-GACGCCGG) and LAP2-*BgIII*AscI-R (5'CGCGCCGGCGTCCAG-GTCGTACCACACCTTGTCTCGATCTCGAAGCCA) and ligating the duplex DNA in-frame into *BgIII/AscI* digested LAP-CFP-TM (renamed as LAP1-CFP-TM).<sup>13</sup> E2p-CFP-TM has previously been described.<sup>13</sup>

**Cell Surface Quantum Dot Labeling of LAP2 with LplA (Figure 4B).** HEK 293T cells were transfected with LAP2-GFP-LDLR plasmid using Lipofectamine 2000. After 24 h in growth media (Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS)) at 37 °C, enzymatic ligation of 11-Br was performed in DPBS containing 10 μM LplA(W37A), 500 μM 11-Br, 1 mM ATP, 5 mM Mg(OAc)<sub>2</sub> and 1% (w/v) BSA (Fraction V, EMD) as a blocking agent for 5 min at room temperature. Cells were then rinsed three times with DPBS followed by treatment with 50 nM HaloTag-QD605<sup>31</sup> in DPBS containing 1% BSA for 5 min at room temperature. After another three rinses with DPBS, cells were imaged in the same buffer on a Zeiss Axio Observer.Z1 inverted epifluorescence microscope using a 40X oil-immersion lens. GFP (493/16 excitation, 525/30 emission, 488 dichroic, 300 ms exposure), QD605 (400/120 excitation, 605/30 emission, 488

dichroic, 200 ms exposure), and DIC images were collected and analyzed using Slidebook software (Intelligent Imaging Innovations). Fluorescence images were normalized to the same intensity ranges.

**Cloning and Analysis of -4 Phe→Val Mutant of LAP4.1 (Figure S3, Supporting Information).** pCTCON2 plasmid carrying LAP4.1 was isolated from yeast clone using the Zymoprep Yeast Plasmid Miniprep kit. Phe at position -4 was mutated to Val using the QuikChange primer 5'GGAGGGTCCGGTAGCGGAGTG-GAAGTGTGATAAAGTATGGTTGATGTCG and its reverse complement primer. This construct was subsequently transformed into *S. cerevisiae* EBY100, grown and induced as described above (see "Model selections"). To compare the yeast cell surface lipoylation of the Phe→Val mutant with the original LAP4.1 clone, clones from Gate A and the clones from Gate B, cells were lipoylated as described above except that 200 nM LplA was used.

**Cell Surface Lipoylation of LAP Constructs (Figures S4 and S5, Supporting Information).** HEK 293T or HeLa cells were transfected with LAP4.1-, LAP4.2-, LAP4.3D-, E2p-, LAP2-, or LAP1-CFP-TM<sup>13</sup> plasmids using Lipofectamine 2000. After 24 h in growth media (DMEM with 10% FBS) at 37 °C, lipoylation was performed in growth media containing 1 μM LplA, 100 μM (±)-α-lipoic acid, 1 mM ATP, 5 mM Mg(OAc)<sub>2</sub> and 1% (w/v) BSA for 10 min at room temperature. Cells were then rinsed three times with DPBS followed by incubation with rabbit anti-lipoic acid antibody (1:300 dilution, Calbiochem) in DPBS containing 1–2% BSA for 10 min at room temperature. Fluorescence staining was achieved by treatment with either fluorescein-conjugated goat anti-rabbit anti-body (1:100 dilution, Calbiochem) or Alexa Fluor 568-conjugated goat anti-rabbit antibody (1:100 dilution, Invitrogen) for 10 min at room temperature in DPBS with 1–2% BSA. Cells were imaged as described above using CFP (420/20 excitation, 475/40 emission, 450 dichroic, 500 ms exposure), fluorescein (493/120 excitation, 525/30 emission, 488 dichroic, 100 ms exposure) and Alexa Fluor 568 (570/20 excitation, 605/30 emission, 585 dichroic, 200 ms exposure) filter sets. Slidebook software was used for emission intensity ratio quantitation. Average across-cell fluorescein and CFP intensities were used, after background subtraction.

**Measurement of LAP2 Kinetics (Figure S7, Supporting Information).** Synthetic LAP2 peptide (sequence GFEIDK<sub>V</sub>VWYDLDA) was prepared by the Tufts University Core Facility. To measure the  $k_{cat}$  and  $K_m$  values for lipoylation, 50 nM LplA was combined with 750 μM lipoic acid, 2 mM ATP, and 5 mM magnesium acetate in DPBS. Varying concentrations of LAP2 (5.5, 11, 22, 44, 88, 176, or 352 μM) were used. Sixty-microliter aliquots were removed from the 30 °C reactions at 5 min intervals, up to 20 min, and quenched with 180 mM EDTA (final concentration). HPLC was used to determine the amount of product in each aliquot, and kinetic parameters were extracted using the Michaelis–Menten equation as described previously.<sup>13,28</sup>

**Acknowledgment.** Funding was provided by the NIH (R01 GM072670 and PN2 EY018244), MIT, and the Sloan and Dreyfus Foundations. We thank Karishma Rahman, Dr. Irwin Chen, Yoon-Aa Choi, Prof. Mark Howarth, Dr. Benjamin J. Hackel, and Prof. K. Dane Wittrup for their advice and assistance. We acknowledge the MIT Koch Institute for use of flow cytometry facilities.

**Supporting Information Available:** Figures S1–S7, Table S1, and the complete author list for ref 30. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA904596F