Molecular Engineering of New Protein Labeling Methodology Based on Rational Design and *In Vitro* Evolution

by
Yoon-Aa Choi

B.S., Chemistry (2004)
Korea Advanced Institute of Science and Technology

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

at the

Massachusetts Institute of Technology

February 2010

Signature of the Author: _____
Department of Chemistry
January 22, 2010

Certified by: _____
Alice Y. Ting
Associate Professor of Chemistry
Thesis Supervisor

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

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

Barbara Imperiali
Class of 1922 Professor of Chemistry and Professor of Biology
Chair

Alice Y. Ting
Associate Professor of Chemistry
Thesis Supervisor

Sarah E. O'Connor
Latham Family Career Development Associate Professor of Chemistry

Molecular Engineering of New Protein Labeling Methodology Based on Rational Design and *In Vitro* Evolution

by

Yoon-Aa Choi

Submitted to the Department of Chemistry on January 22, 2010 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

Site-specific labeling using *E. coli* biotin ligase (BirA) and its 15-amino acid "acceptor peptide" (AP) has been applied to study the function of various cellular proteins. In order to extend the capabilities of biotin ligase-based labeling, we engineered key elements of the labeling platform. First we characterized a novel peptide substrate (called "yeast acceptor peptide" (yAP)) for yeast biotin ligase (yBL) that had been evolved by phage display. Assays performed *in vitro* and on the yeast surface showed that the yBL/yAP pair was orthogonal to the BirA/AP pair, allowing two-color labeling of different proteins on cells with differently-colored probes. Second, to improve the kinetic efficiency of yAP, we developed a novel selection scheme based on yeast display. Model selections demonstrated up to 1000-fold enrichment, and three rounds of selection on a randomized peptide library were performed.

Third, we attempted to improve the kinetic efficiency of BirA through evolution by *in vitro* compartmentalization (IVC). Because the original IVC protocol based on bead-linked DNA had many technical problems, we developed a novel bead-less IVC protocol. An enrichment factor of 25 was obtained in a model selection. Due to the single-turnover nature of the selection, however, this scheme was not able to enrich highly active catalysts over moderately active ones.

In separate work, we turned our attention to the streptavidin-biotin pair. Again using bead-less IVC, we performed a selection for streptavidin mutants that could bind a ketone analog of biotin with high affinity. Two rounds of selection were performed but characterization of enriched clones was not completed.

Finally, we helped to discover a mutant ligase that could catalyze attachment of a fluorinated aryl azide photocrosslinker to proteins fused to a 17-amino acid peptide tag. The aryl azide probe was tested and shown to be accepted by a W37V mutant of *E. coli* lipoic acid ligase (LplA).

Thesis Supervisor: Alice Y. Ting

Title: Associate Professor of Chemistry

To my dad

Acknowledgements

First of all, I would like to thank my research advisor, Professor Alice Y. Ting, for her incredible support and guidance throughout last five and a half years. She was not only a "research" advisor but also a great life mentor who taught me how to survive as a scientist in this competitive world. Although she is well-known as a demanding advisor, they do not know how supportive she is. Without her endless support and faith for me, I would not be able to finish the graduate school. I greatly appreciate her being my advisor.

I also thank members of my thesis committee, Professor Imperiali and Professor O'Connor, for educating me outside the Ting Lab. I was able to become strict with myself because they provided highly objective view as academic referees.

Apart from research, I thank Dean Blanche Staton of the Graduate Student Office, for helping me manage very difficult situations. I am also grateful for staffs in MIT medical, who helped me not to give up when I encountered a personal crisis.

Most of all, I was lucky to work with wonderful people in the Ting Lab. I would like to thank Irwin Chen for teaching me basic biochemical techniques when I joined the group without much experience in bio-related research. Working with him for the yAP project and IVC selections was always full of inspiration. I also thank Dr. Hemanta Baruah for being there with me in the lonely journey of BirA IVC selection. It was also nice to work with Dr. Dan Chinnapen and Dr. Guobin Luo for streptavidin IVC evolution. Although it was the darkest time of the Ting Lab, I appreciate the fact that I could work with them who are exceptionally bright and hard-working.

I also thank people who proofread my thesis: Peng Zou (Chapter 1), Dr. Sujiet Puthenveetil (Chapter 2), Dan Liu (Chapter 3), Tao Uttamapinant (Chapter 3), and Katie White

(Chapter 4). I thank Sarah Slavoff and Jen Yao for helpful discussion preparing thesis. I thank Chi-Wang Lin, Marta Fernandez-Suarez, and Jackie Chan for personal support.

I would like to thank my korean friends, Seungjib Choi, Hyun A Kang, Changsik Song, Yunmi Lee, Mi Hee Lim, Hyun-Ji Song, Changhoon Lee, and Sunkyu Han, for helping me settle down here in Cambridge. I also thank Samsung Scholarship for the financial support throughout the graduate school. I thank Professor Sukbok Chang for mentoring me even after I graduated from KAIST.

I would like to thank my family: my dad, my younger brother Changgyu, and my lovely dog Doong-ee. I know how much they have worried about me living here alone. I wish they would know how much I love them.

Lastly, I would like to thank my fiancé, Yongwon Jung for his endless support on me. For me, meeting with him was the best thing happened at MIT. He always cheered me up whenever I was disappointed with bad results. As an alumnus of MIT chemistry department, he always understood my anxieties about cumes, second year oral exam, original proposal, and thesis defense. He helped me proofread the thesis so much, and I think he almost deserved his second Ph.D. I deeply appreciate his dedication to my life at MIT.

Table of contents

Title Page.....1

Signature Page.....2

Abstract.....3

Dedication Page.....4

Acknowledgements.....5

Table of Contents.....7

List of Figures.....11

List of Tables.....14

List of Abbreviations.....15

Chapter 1: Introduction.....19

Part 1. Current methodologies for site-specific labeling of proteins in live cells.....21

Small molecule labeling using protein handles.....21

Small molecule labeling with peptide handles using binding interaction.....23

Small molecule labeling with peptide handles using enzyme-catalyzed covalent ligation.....24

Genetic incorporation of unnatural amino acids.....26

Part 2. Directed evolution of proteins for site-specific protein labeling.....30

Cell-surface display (bacterial display and yeast display).....32

Phage display.....34

In vitro compartmentalization (IVC).....35

Part 3. Biotin ligase and lipoic acid ligase in *E. coli*.....35

E. coli biotin protein ligase (BirA).....38

E. coli lipoic acid protein ligase (LplA).....39

Applications of BirA and LplA.....41

Future directions.....43

References.....43

Chapter 2: Improvement of Yeast Biotin Ligase Acceptor Peptide by Yeast Display Evolution....56

Introduction....57
Results and Discussion....59
Part I. Characterization and applications of phage-evolved yAP peptide....59
Summary of Irwin Chen's work....59
Biotinylation of yAP fusion proteins....61
Orthogonality of the yAP/yBL pair....62
Part II. Yeast display evolution of improved yAP sequences....66
Model selection (1)....66
Model selection (2)....68
Library design....71
Selections with AP2.2 library....71
Conclusions....73
Experimental....75
References....83

Chapter 3: Development of a Biotin Ligase Evolution Platform Based on *In Vitro* Compartmentalization....87

Introduction....88
Library construction....88
In vitro compartmentalization: linking genotype to phenotype....89
Direct selection for catalytic activity....90
Results and Discussion....91
Design and construction of a BirA mutant library....91
BirA evolution strategy: bead-based *in vitro* compartmentalization (IVC)....93
BirA evolution strategy: bead-less IVC....95
Model selection with wild type BirA and kinetically impaired G115S BirA....97
Conclusions....100
Experimental....102
References....112

Chapter 4: Development of a Streptavidin Evolution Platform Based on *In Vitro* Compartmentalization.....117

Introduction.....118

Results and Discussion.....122

Model selection with wild type (WT) and S45A streptavidin 124

Design and construction of a streptavidin mutant library 1.....126

Selection with the library (1) 128

Functional assay: model selection with BirA.....128

Design and construction of a streptavidin mutant library 2.....131

Selection with the library (2) 132

Functional assay: gel shift assays.....133

Conclusions.....135

Experimental.....136

References.....141

Chapter 5. An Engineered Aryl Azide Ligase for Site-Specific Mapping of Protein-Protein Interactions via Photocrosslinking.....144

Introduction.....145

Results and Discussion.....147

Design, synthesis, and incorporation of aryl azide by wild type and mutants of LplA.....147

Engineering of LplA mutants to use aryl azide probe **1**.....148

Specificity of aryl azide ligation by W37V LplA 150

Application of aryl azide ligase for photocrosslinking proteins.....150

Conclusions.....153

Experimental.....155

References.....158

Appendix.....163

Characterization of AP conjugated oligonucleotides:

HPLC and MALDI-TOF (Chapter 3).....164

Characterization of ketone biotin conjugated oligonucleotides:

MALDI-TOF (Chapter 4).....166
Spectral characterization of aryl azide **1** (Chapter 5).....168

Curriculum Vitae.....171

List of Figures

Chapter 1: Introduction

Figure 1 Overview of directed evolution of proteins.....31

Figure 2 Strategies for genotype and phenotype linkage by (A) cell- surface display, (B) phage display, and (C) *in vitro* compartmentalization. 33

Figure 3 Reactions catalyzed by (A) biotin protein ligase, and (B) lipoic acid protein ligase.....37

Chapter 2: Improvement of Yeast Biotin Ligase Acceptor Peptide by Yeast Display Evolution

Figure 1 Design of the peptide library for yeast display selection based on AP2.2 library.....60

Figure 2 Biotinylation of yAP fusion proteins. 62

Figure 3 Site-specific biotinylation of yAP expressed on the surface of live HeLa cells.....63

Figure 4 Selective labeling of live HeLa cells expressing yAP or AP fusion proteins (domain structures shown at top) with quantum dots. 64

Figure 5 Orthogonal labeling on yeast cell surface. 65

Figure 6 Schematic illustration of yeast display evolution of yAP. 66

Figure 7 Model selection (1). 67

Figure 8 Model selection (2). 69

Figure 9 Sequences of the AP2.2 library for yeast display.....71

Figure 10 Myc expression of AP2.2 library after transformed to yeast cells. 72

Figure 11 Three rounds of yeast display selection with the AP2.2 library. 73

Chapter 3: Development of a Biotin Ligase Evolution Platform Based on *In Vitro* Compartmentalization

Figure 1 Positions of randomized residues of BirA. 92

Figure 2 Selection scheme for bead-based IVC..... 94

Figure 3 Selection scheme for bead-less IVC..... 96

Figure 4 Results of bead-less IVC model selection using BirA and folA 97

Figure 5 Bead-less IVC model selections using wild type (WT) BirA and kinetically-impaired G115S mutant 98

Figure 6 Results of model selections with wild type (WT) BirA and G115S using bead-less IVC.....99

Figure 7 The BirA library was created using overlap extension PCR. 103

Chapter 4: Development of a Streptavidin Evolution Platform Based on *In Vitro* Compartmentalization

Figure 1 Crystal structure of the biotin binding site of streptavidin. 119

Figure 2 A proposed orthogonal pair of streptavidin and biotin analogs. 119

Figure 3 Structure of biotin analogs that have been investigated for streptavidin engineering. 120

Figure 4 Proposed bead-less IVC selection scheme for streptavidin mutants that bind to ketone biotin.....123

Figure 5 IVC model selection scheme with wild type (WT) and S45A streptavidin.....124

Figure 6 Results of model selections with wild type (WT) and S45A streptavidin. 125

Figure 7 Randomized positions in streptavidin mutant library 1.....127

Figure 8 Functional assays of selected streptavidin mutants from each round, via model selection with BirA. 130

Figure 9 Randomized positions in streptavidin mutant library 2.....132

Figure 10 Functional assays on selected streptavidin (SA) mutants from each round of selection, via gel shift. 134

Figure 11 HPLC traces of ketone biotin modified oligonucleotides.....140

Chapter 5. An Engineered Aryl Azide Ligase for Site-Specific Mapping of Protein-Protein Interactions via Photocrosslinking

Figure 1 Scheme illustrating site-specific aryl azide ligation to LAP fusion protein, followed by photocrosslinking to interacting protein. 147

Figure 2 Synthetic scheme for aryl azide probe **1**.....148

Figure 3 The active site of *E. coli* LplA 149

Figure 4 HPLC assay of W37V-catalyzed ligation of aryl azide **1** onto LAP fused HP1 protein (LAP-HP1).....149

Figure 5 Specificity of aryl azide ligation. 151

Figure 6 Aryl azide-mediated photocrosslinking of FKBP and FRB in mammalian cell lysate.....152

Appendix

Figure 1 HPLC analysis of the purified AP conjugated oligonucleotides (Chapter 3).....164

Figure 2 MALDI-TOF analysis of the purified AP conjugated oligonucleotide.....165

Figure 3 MALDI-TOF analysis of the purified ketone biotin conjugated oligonucleotide (PIVB-1).....166

Figure 4 MALDI-TOF analysis of the purified ketone biotin conjugated oligonucleotide (LMB2-1).....167

Figure 5 ^{1}H-NMR spectra of aryl azide **1**.....168

Figure 6 ^{19}F-NMR spectra of aryl azide **1**.....169

Figure 7 HR-MS spectra of aryl azide **1**.....170

List of Tables

Chapter 1: Introduction

Table 1a. Different methods for labeling recombinant proteins with small molecule probes in live cells and their relative attributes. 28

Table 1b. Different methods for labeling recombinant proteins with small molecule probes in live cells and their relative attributes... 29

Chapter 3: Development of a Biotin Ligase Evolution Platform Based on *In Vitro* Compartmentalization

Table 1 Sequences of selected clones from the BirA mutant library.....92

Chapter 4: Development of a Streptavidin Evolution Platform Based on *In Vitro* Compartmentalization

Table 1 Sequences of random clones from streptavidin mutant library 1.....127

Table 2 Sequences of random clones from streptavidin mutant library 2.....132

List of Abbreviations

AcOH	glacial acetic acid
ACP	acyl carrier protein
AP	acceptor peptide
AP Q →E	The mutant AP peptide where the Gln right before the biotinylation site is mutated to Glu
B4F	biotin-4-fluorescein
BCA	bicinchoninic acid
BCCP	*E. coli* biotin carboxyl carrier protein
BirA	*E. coli* biotin ligase
biotin-AMP	biotin adenylate ester
bp	DNA base pairs
BPL	biotin protein ligase
B-PER	bacterial protein extraction reagent
BS	bovine serum
BSA	bovine serum albumin
bsBL	*B. subtilis* biotin ligase
BWT	binding and washing buffer plus Triton X-100
CFP	cyan fluorescent protein
cfu	colony-forming units
CHO	Chinese hamster ovary
CoA	coenzyme A
dBSA	dialyzed bovine serum albumin
DHFR	dihydrofolate reductase
DIC	differential interference contrast
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethyl sulfoxide
DPBS	Dulbecco's phosphate-buffered saline
DTT	dithiothreitol
EDC	1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride
EDTA	ethylenediamine tetraacetic acid
EGFR	epidermal growth factor receptor
EGTA	ethylene glycol tetraacetic acid
ELISA	enzyme-linked immunosorbent assay

ESI-MS	electrospray ionization-mass spectrometry
EtOH	ethanol
Et2O	diethyl ether
EtOAc	ethyl acetate
FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FKBP12	FK506 binding protein 12
FlAsH	Roger Tsien's fluorogenic biarsenicals
folA	*E. coli* dihydrofolate reductase
FP	fluorescent protein
FRET	fluorescence resonance energy transfer
GFP	green fluorescent protein
GHCl	guanidinium hydrochloride
hAGT	human alkylguanine transferase
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hex	hexanes
HP1	heterochromatin protein 1
HPLC	high-performance liquid chromatography
HR-MS	high resolution mass spectrometry
HRP	horseradish peroxidase
IPTG	isopropyl-β-D-thiogalactopyranoside
IVC	*in vitro* compartmentalization
IVTM	*in vitro* transcription and translation mixture
LAP	LplA acceptor peptide, first generation
LAP2	LplA acceptor peptide, second generation
LB	Luria-Bertani broth
LPL	lipoic acid protein ligase
LplA	*E. coli* Lipoic acid ligase
Lypoyl-AMP	lipoate adenylate ester
MALDI-TOF	matrix-assisted laser desorption/ionization time of flight
Me	methyl
MeOH	methanol
MES	2-(*N*-morpholino)ethanesulfonic acid
mjBL	*M. jannaschii* biotin ligase

NaOAc	sodium acetate
NEt3	triethylamine
NHS	*N*-hydroxysuccinimide
Ni-NTA	nickel nitriloacetic acid
NMR	nuclear magnetic resonance
OD	optical density
PA	photoaffinity
PBS	phosphate-buffered saline
PBS-B	phosphate-buffered saline + 0.1% BSA
PBS-T	phosphate-buffered saline + 0.1% Tween
PCA	protein fragment complementation assay
PCP	peptide carrier protein
PCR	polymerase chain reaction
PDB	Protein Data Bank
PE	(*R*)-phycoerythrin
PEET	PBS-T + 1 mM EDTA + 1 mM EGTA
pFLAG	phosphine-FLAG peptide
pfu	plaque-forming units
PhBL	*P. horikoshii* biotin ligase
PMSF	phenylmethylsulfonyl fluoride
PPI	protein-protein interaction
PPtase	4′-phosphopantetheinyl transferase
psBCCP	P. shermanii transcarboxylase 1.3S subunit
QD	quantum dot
Rf	retention factor
rt	room temperature
RT-PCR	real time polymerase chain reaction
SA	streptavidin
SCD	single chain-dimeric
SD –W –U	synthetic dextrose media lacking tryptophan and uracil
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SG –W –U	synthetic galactose media lacking tryptophan and uracil
SLF′	synthetic ligand for FKBP12
SMCC	succinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate
TA	5 mM Tris-acetate pH 8.0 buffer

TBS	Tris-buffered saline
TBS-T	Tris-buffered saline + 0.1% Tween
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin-layer chromatography
TM	transmembrane domain
TMS	trimethysilyl
yAP	yeast acceptor peptide
yBL	yeast biotin ligase
YFP	yellow fluorescent protein
WT	wild type

Chapter 1: Introduction

Understanding the structure and function of cellular components such as proteins, nucleic acids, and other biomolecules is the main goal of biochemistry. Powerful tools such as crystallography or nuclear magnetic resonance have been developed to reveal the structure of biomolecules. Investigating the function of biomolecules also requires various experimental tools that would allow for quantitative and experimentally well-defined monitoring at the molecular level of the spatial and/or temporal cellular processes[1]. Many experimental strategies such as immuno-precipitation have been developed for this purpose. Among these strategies, cellular fluorescence imaging techniques have provided novel and important information (e.g. protein localization and interaction with other biomolecules) in cell biology, where most of it would not be obtainable with any other biological techniques.

Key elements of these cellular imaging techniques are fluorescent proteins (FPs) or reporter proteins such as β-lactamase, β-galactosidase, and luciferase, which can be genetically introduced into proteins of interest and provide high spatial or temporal information of the target cellular proteins[2]. Although they have been successfully used in biological studies to visualize, track, and quantify proteins in live cells, a consensus drawback of FPs has been their large size (Green fluorescent protein has 238 amino acids), which can make it difficult to study the innate structure and function of the protein of interest[3, 4]; for example, Baens and coworkers demonstrated that EGFP (Enhanced Green Fluorescent Protein) fusion proteins inhibit Lys 63-and Lys 48-linked polyubiquitination[5].

Compared to FPs, small molecule probes, on the other hand, display a variety of properties and functions such as small size, photoreactivity, stable fluorescence, and versatile functional group handles for modification. These facets of small molecule probes can be utilized to develop more sophisticated strategies to explore cellular proteins in a non-invasive

manner. For example, protein substrates derivatized with photocaging groups can be used to temporally and spatially regulate protein activities[6]. Site-specifically labeled photoaffinity probes can offer detailed molecular information about protein-protein interactions. Tissues in a live animal can be imaged by using magnetic resonance imaging or positron emission tomography probes. Moreover, electron microscopy probes provide much higher resolution for cellular spatial information after cell fixation. One critical requirement for developing these applications, however, is to target small molecule probes to specific proteins in the cellular context. In the first part of this introductory chapter, current protein labeling methods with small molecules will be discussed, followed by a brief review on protein evolutions for site-specific labeling in Part 2. Lastly, two small molecule ligases, which are main focuses of this thesis, will be discussed in depth.

Part 1. Current methodologies for site-specific labeling of proteins in live cells

Multiple methods have emerged for site-specific protein labeling with small molecules in recent years[7, 8]. Common strategies of these methods are to employ a protein or peptide handle, which is genetically fused to the protein of interest. Small molecule probes can be subsequently recruited to this handle through either a covalent linkage or a high affinity non-covalent binding interaction. In this section, currently reported labeling methods are evaluated with particular focus on their application to live cell imaging. Table 1 at the end of this section summarizes the relative advantages and drawbacks of the methods.

Small molecule labeling using protein handles

A number of proteins can specifically bind certain small molecules, which are often

enzyme substrates or ligands for specific receptors. These highly specific interactions have been well exploited to guide various chemical probes to proteins of interest. Protein handles using binding interactions include *Escherichia coli* dehydrofolate reductase (DHFR), FK506-binding protein (FKBP12, as a F16V mutant), and single chain antibodies. Israel and coworkers used fluorescent derivatives of methotrexate, an inhibitor of DHFR, to visualize cellular proteins fused to mammalian DHFR in DHFR-deficient Chinese hamster ovary (CHO) cells[9]. Similarly, FK506 analog SLF' (synthetic ligand for FKBP12 F36V mutant) was constructed to bind to this mutant form of FKBP12 (F36V) with a good affinity (94 pM) but not to the endogenous FKBP12 in NIH-3T3, COS-7, Jurkat, and HeLa mammalian cell lines[10]. SLF' has a fluorescein moiety, which enables specific fluorophore labeling of FKBP12 (F36V) fusions in cells. Farinas and Verkman also demonstrated that single-chain antibodies fused to localization signal sequences can be used to target various hapten-fluorophore conjugates to specific subcellular compartments of the secretory pathway in live cells[11]. Due to their non-covalent properties, however, they are unsuitable to long-term labeling. Dissociation of ligands from the protein handles inevitably decreases the signals over time.

Covalent linkage between proteins and small molecules is superior in that perspective, and protein handles such as *E. coli* dehalogenase (commercialized as Halotag, Promega), human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT), and cutinase have been successfully utilized to covalently attach chemical probes via their corresponding substrate derivatives. For example, a modified 33-kDa bacterial dehalogenase domain has recently been developed by Promega[12]. Wild type dehalogenases break down haloalkanes in a two-step reaction, involving alkyl transfer to an active site Asp side chain (and the loss of the halide ion) followed by hydrolysis of the alkyl-enzyme bond (releasing alkyl probes)[13]. An

active site His residue is necessary for the hydrolysis, mutation of which preserves the alkyl-enzyme conjugates. Since eukaryotes lack dehalogenases, His mutated dehalogenase fusion proteins can be labeled with haloalkane derivatives with a very low background staining in cells. Johnsson and coworkers also demonstrated that cell permeable O^6-benzylguanine derivatives containing a fluorophores, biotin, haptens, or other protein ligands can be linked to Cys 145 of hAGT[14]. The labeling was performed with nuclear, cytosolic, cytoskeletal, and cell surface proteins in *E. coli*, yeast, and hAGT-deficient mammalian cells. The fungal enzyme cutinase and its suicide inhibitor p-nitrophenyl phosphate are used to attach a variety of labels to the integrin lymphocyte-associated antigen-1 (LFA-1) on the surface of live cells[15]. The size of most protein handles, however, is rather large (80 - 297 amino acids), which is a major drawback of these methods.

Small molecule labeling with peptide handles using binding interaction

Reducing the size of targeting tags using smaller peptides, however, brings a new challenge; the strength and specificity of peptide interactions with small molecule probes are mostly lower than those of proteins. As long as these peptides can be recognized by their interaction partners at a high signal to noise ratio, however, peptides are clearly more favorable since they can be fused not only to termini but also to loops of protein of interest, without perturbing the structure and function of the protein. His_6-tag[16], which has been widely used for protein purification, was applied to label proteins in living cells, and the tetracysteine FlAsH binding sequence[17] was also used for imaging connexins in gap junctions. Furthermore, ReAsH, a red fluorophore derivative of FlAsH, can generate reactive oxygen species for light-assisted protein inactivation in functional studies or diaminobenzidine precipitation in electron

microscopy[18]. These peptide handles, however, also show several limitations in protein labeling. The His_6-tag method suffers from inadequate specificity and a non-covalent linkage which has a rather fast off rate. The FlAsH method also has fairly low sequence specificity (monothiols are labeled by FlAsH as well)[19, 20], a slow labeling kinetics[21], an inability to work in oxidizing cellular compartment such as secretory pathways[22], and no demonstrated successes on structures other than xanthene-based fluorophores. New complex of polyhistidine-zinc-zinc chelating dye (called HisZiFit) has been developed by Tsien and coworkers[23]. Their binding is fairly tight (K_d = 40 nM), but the dye shows suboptimal characteristics such as low photostability and high background staining. Another zinc chelator DpaTyr has been shown to bind polyaspartate residues with low K_d (1.4 μM)[24]. These zinc chelating dyes, however, are not membrane permeable, and the concentration of zinc necessary for labeling is much higher than physiological concentration of zinc. A lanthanide binding aptamer is also available for luminescence imaging[25], and a Texas Red binding aptamer was reported but yet needs to be shown in cellular context[26].

Small molecule labeling with peptide handles using enzyme-catalyzed covalent ligation

Enzymatic coupling of chemical probes to peptide handles offers a potentially high signal to noise ratio due to the selectivity of the enzyme. Our laboratory developed new methodologies of cellular protein labeling which address many of the deficiencies of FlAsH and other labeling methods[27]. A 15-amino acid acceptor peptide ("AP", GLNDIFEAQ**K**IEWH) is fused to the protein of interest, and then a ketone analog of biotin is specifically ligated to a lysine side chain (underlined) within this sequence using the *E. coli* enzyme biotin ligase (BirA). The reaction depends on ATP, which is used to activate the biotin

as an AMP ester. The labeling reaction is extremely sequence-specific because BirA is highly specific for the AP peptide and does not biotinylate any endogenous mammalian proteins. As ketones are absent from cell surfaces, the enzymatically introduced ketone can be selectively derivatized with diverse hydrazide- or hydroxylamine-functionalized probes. BirA/ketone labeling is currently limited to cell surface applications because of the imcompatibility of ketone-hydrazide (or hydroxylamine) ligation with intracellular applications. Sarah Slavoff in the lab reported that biotin ligase from *Pyrococus horikoshii* can incorporate an azide or alkyne analog of biotin (DTB-Az or PB, respectively) onto p67, which is one of the endogenous biotinyl domains of human biotin ligase[28]. Unfortunately, the reaction rates of these analogs are too low to be used for live cell labeling at this point. Also, *P. horikoshii* biotin ligase did not use AP as a substrate, which is another barrier for the biotin ligase system.

In light of this result, Marta Fernandez-Suarez in the lab developed a new labeling methodology using *E. coli* lipoic acid ligase (LplA), which is structurally related to biotin ligase and can attach a lipoic acid prosthetic group to pyruvate dehydrogenase[29]. Instead of lipoic acid, 7-azidoheptanoic acid was incorporated to a lipoic acid acceptor peptide tag (22 amino acids) by LplA, and then the azide was derivatized with a cyclooctyne-conjugated fluorophore. Although the methodology was only demonstrated on cell surface proteins, the absence of azide and cyclooctyne inside cell will potentially help the methodology being extended to intracellular labeling. Moreover, we have found that LplA is more tolerable to point mutations in active site and thereby accepts many different substrate probes such as aryl azide analog[30], coumarin analog, and so on. A detailed procedure for the development of the methodology with an aryl azide analog is discussed in Chapter 5 with its advantages, limitations, and suggestions for future development.

Another method that enables single-step labeling is *Bacillus subtilis* phosphopantetheine transferase (PPtase; called Sfp)-catalyzed ligation of diverse CoA derivatives to the ybbR peptide tag[31]. Originally, the labeling was performed with a protein recognition sequence (~80 amino acids) derived from either peptide carrier protein (PCP)[32, 33] or acyl carrier protein (ACP)[34]. More recently, Yin and Walsh have identified a 12-amino acid sequence that serves as a specific substrate for Sfp (called "ybbR"). PPtase possesses wide small molecule substrate tolerance, and it also shows high specificity for its peptide substrate comparable to those of BirA and LplA. PPtase-based labeling is currently limited to extracellular tagging because CoA is cell impermeant and intracellular use would require proper conjugation of functional groups to CoA by endogenous biosynthetic pathways.

Other enzymes used for extracellular tagging include the bacterial enzyme sortases, which cleave the Thr-Gly amide bond in LPXTG sequences and ligate the resulting COOH terminus to pentaglycine peptide derivatives[35]. The large polar surface areas of these peptide analogs, however, prevent cellular entry. The use of transglutaminase (which has been harnessed by our laboratory) is limited by low specificity to its peptide substrate, although the enzyme can handle diverse amine substrates[36]. Farnesyltransferases can mediate attachment of a variety of farnesyl derivatives to the cysteine residue in the COOH-terminal sequence CVIA, but the presence of endogenous targets again precludes their uses inside cells[37].

Genetic incorporation of unnatural amino acids

A different set of techniques allow specific protein labeling with synthetic functional groups with single amino acid precision. Genetic incorporation of unnatural amino acids was demonstrated in living cells[38], whereas other methods such as chemical modification[39] or

native chemical ligation[40, 41] can be used only *in vitro* or with microinjections to cells.

Unnatural amino acid mutagenesis via amber codon suppression has been used to label proteins with biophysical probes such as benzophenone[42] and aryl azide crosslinkers[43], as well as fluorophores such as dansylalanine[44] and 7-hydroxycoumarin[45]. In addition to introducing such modifications directly, amino acids with ketone[46], azide[43], acetylene[47], and thioester groups can be incorporated[48], providing new labeling reaction moieties that are orthogonal to those of endogenous amino acids. The perfect specificity with minimum possible structural perturbation to the protein of interest is a big advantage of the methodology. However, the method cannot be used to incorporate unnatural amino acids that are toxic or incompatible with the protein biosynthesis machinery (e.g., D-amino acids)[21]. Also, a new orthogonal tRNA synthetase/tRNA pair needs to be evolved for every new probe. Another serious concern is that a large amount of truncated protein product is inevitably generated as a result of the competition between unnatural amber suppressor tRNA and the release factors that normally terminate protein synthesis at the amber stop codon. This can lead to a dominant negative effect, especially if the protein of interest is multimeric in its functional state.

Method	Tag size	Sources of background labeling	restrictions	Other comments
Protein handles				
DHFR[9]	157	Endogenous DHFR	-	Ligand dissociation
FKBP12 (F36V)[10]	108	-	-	Ligand dissociation
Single-chain antibody[11]	~230	-	Oxidizing cellular compartments	Ligand dissociation
HaloTag™ [12]	~300	Unknown endogenous 28kDa protein[49]	-	-
hAGT[14]	207	Endogenous hAGT	-	-
Cutinase[15]	~200	-	-	-
Peptide handles using binding interactions				
His_6-tag[16]	6-10	-	Generally cell surface	Ni^{2+} quenches fluorescence and is toxic, ligand dissociation
Tetracysteine[17,18]	6-10	Significant affinity for monothiols	Reducing environment	Fluorogenic, long-labeling times, possible residual arsenic toxicity
HisZiFit[23]	6	-		
Polyaspartate[24]				
Lanthanide binding tag[25]	15-25	-	unknown	
Texas-red binding aptamer[26]	23-38	Texas-red localization to mitochondria	Texas-red only	

Table 1a. Different methods for labeling recombinant proteins with small molecule probes in live cells and their relative attributes. The tag size is given as the number of amino acids.

Method	Tag size	Sources of background labeling	Restrictions	Other comments
Peptide handles using enzyme reaction				
Biotin ligase/ketone[27]	15	One natural substrate in *E. coli*	Cell surface	Two-step labeling
Lipoic acid ligase/ azide[29]	22	Three natural substrates in *E. coli*	Cell surface	Two-step labeling, potential for intracellular labeling
Sfp PPtase[31]	11	One natural substrate in B. subtilis	Cell surface	Single step labeling
Sortase[35]	5+linkers	Two natural substrates in *S. aureus*	Cell surface	Single step labeling
Transglutaminase[36]	6-7	Can label endogenous proteins under forcing conditions	Cell surface	Single step labeling
Farnesyltransferase[37]	4	Endogenous substrates in mammalian cells	Not with cells	Two-step labeling
Other methods				
Unnatural amino acid mutagenesis[21,38-48,50]	0	-	-	Protein truncations

Table 1b. Different methods for labeling recombinant proteins with small molecule probes in live cells and their relative attributes. The tag size is given as the number of amino acids.

Part 2. Directed evolution of proteins for site-specific protein labeling

As discussed in Part 1, many site-specific protein labeling methods employ enzymes to get covalent linkage between a small molecule probe and the protein of interest and to benefit from the innate specificity of the enzyme reaction, which reduces background labeling. However, many enzymes are very specific to their original small molecule substrates and do not use biochemically interesting probes (e.g. fluorophores) as substrates. One solution to this problem is to use substrate analogs, which possess intended functional moieties, as demonstrated in ketone biotin labeling by BirA. However, in most cases the analogs are not accepted by the original enzyme due to their structural difference from the native substrate. Site-directed mutagenesis of an enzyme in the vicinity of the substrate binding site has been used to overcome the changes in the substrate structure, and significant success of this 'rational design' has been demonstrated[51, 52]. Proper structural information for rational design, however, is not available for all proteins. Moreover, successful design of an enzyme mutant with new substrate specificity is often highly challenging. Directed enzyme evolution has become a more general strategy as the method does not necessarily require structural information on the enzyme and can cover a wide range of mutations, providing a better chance to find properly mutated enzymes for the substrate analogs.

Biochemists have been actively trying to change the properties of biological molecules to give them novel functions in a directed fashion (often called molecular evolutions)[53, 54]. In evolution experiments in the laboratory, the gene encoding a protein of interest is first (partially) randomized as described in Figure 1. From these mutated genes, many mutant proteins are produced, forming a so called protein library. Appropriate screening or selection methods may then be applied to select members of the resulting libraries of

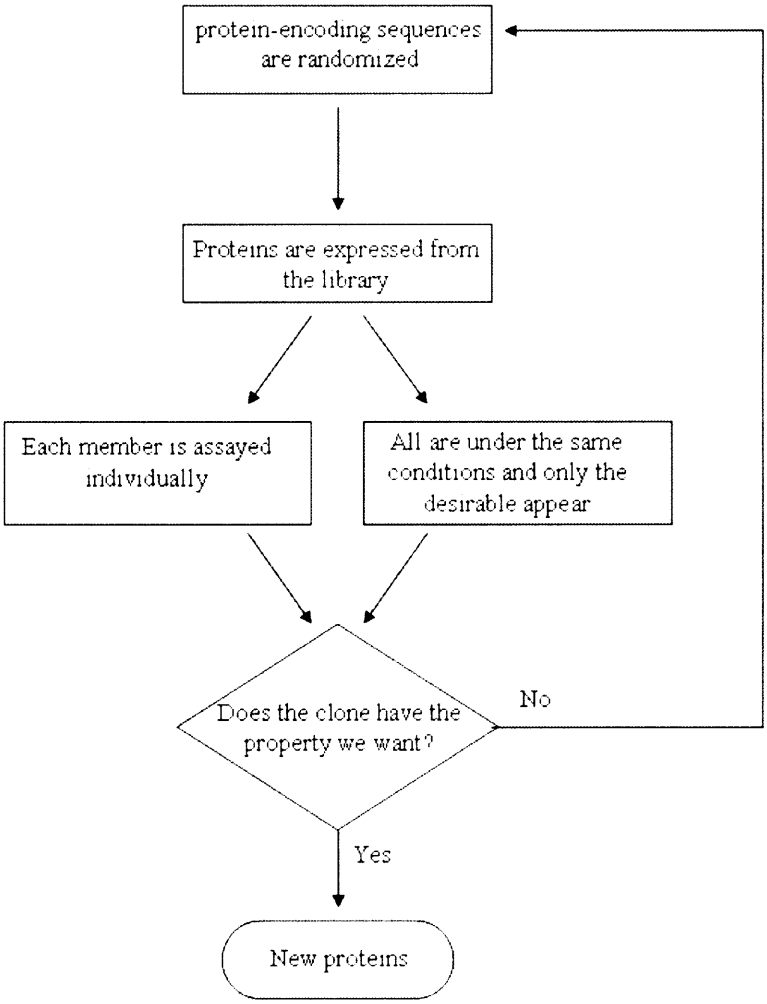


Figure 1 Overview of directed evolution of proteins

protein mutants that have particular properties, such as the ability to bind a small molecule (e.g. fluorophore) or to catalyze a chemical transformation (e.g. ligation of a fluorophore). The genes for selected mutant proteins are collected and amplified for further mutations and selections. Through iterative cycles of mutagenesis and amplification of selected mutant protein members, beneficial mutations accumulate. In this way, populations of macromolecules may be carefully evolved toward having new functions. Although successful

progress has been made toward these goals[55], evolution of biomolecules in the laboratory remains a challenging task because functional sequences are rare compared with the almost unimaginably large number of theoretical sequences to be sampled. In Chapter 2, engineering of a peptide substrate will be deliberated, and in Chapter 3, Chapter 4, and Chapter 5 we will discuss directed evolution of several proteins.

Since proteins cannot be amplified themselves, genotype (the genes encoding proteins) and phenotype (the expressed proteins) must be linked physically to get the sequence information. *In vitro* protein evolution can be categorized by strategies to generate the genotype-phenotype linkage (Figure 2): cell-surface display[56], phage display[57], mRNA display[58], ribosome display[59], and *in vitro* compartmentalization (IVC)[60]. Here we will compare these selection techniques with particular focuses on their advantages and disadvantages. Chapter 2 will cover the yeast display methodology to evolve better peptide substrate for yeast biotin ligase, and detailed discussion of IVC for evolution of BirA and Streptavidin will be in Chapter 3 and 4.

Cell-surface display (bacterial display and yeast display)

In cell-surface display, proteins of interest are fused to a membrane protein so that they are presented on the cell surface. Various cells, including bacteria[61] and yeast[62], can be used for this purpose, where genotype and phenotype are contained in each cell. The advantage of cell-surface display is that high throughputs (~10^5 clones/second) can be obtained with techniques like fluorescence activated cell sorting (FACS)[63]. However, compared with other methods (e.g. *in vitro* compartmentalization), the size of library is limited in cell-surface display due to the gene transformation step, which generates ~10^{10} clones for *E. coli* and ~10^7 clones for yeast.

Displaying antibody libraries on the surface of *E. coli* and binding with fluorescence-labeled antigens resulted in antibodies with high affinities[64, 65]. The selection of novel enzymes was also reported using a positively charged synthetic FRET substrate[66].

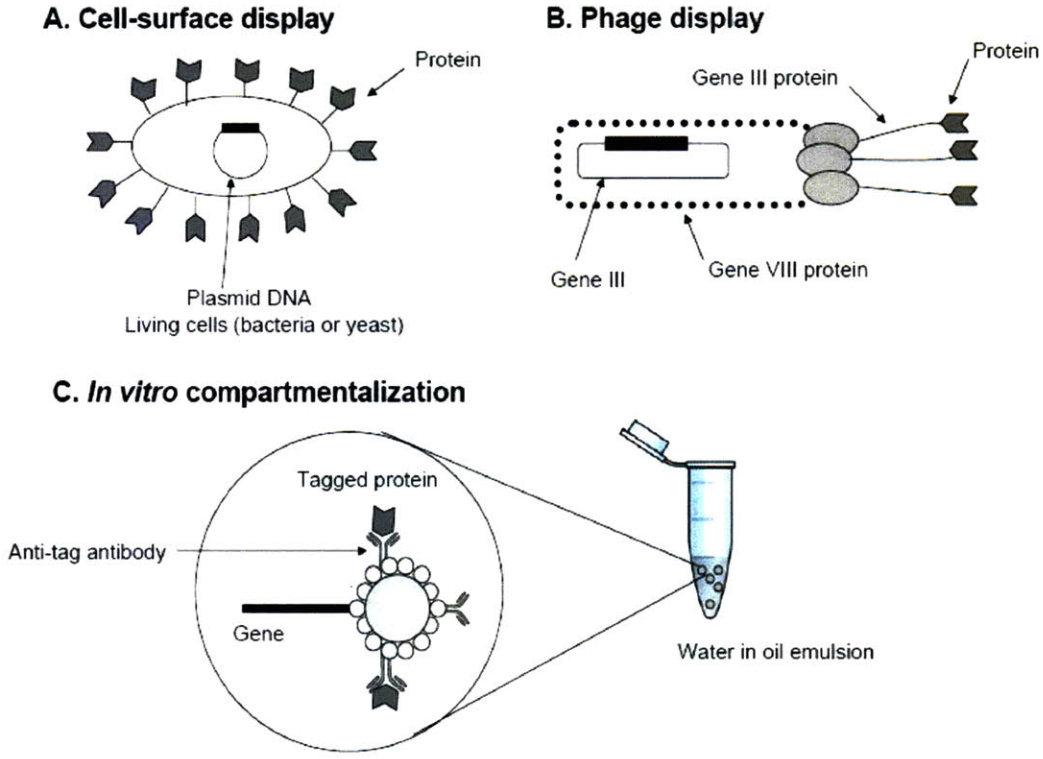


Figure 2 Strategies for genotype and phenotype linkage by (A) cell-surface display, (B) phage display, and (C) *in vitro* compartmentalization. The figure was modified from Matsuura, T.; Yomo, T.; Journal of bioscience and bioengineering, 2006, 101, 449-456.

In yeast display, a protein of interest is presented as a fusion to the Aga2p mating agglutinin protein on the surface of yeast[67]. Recent advances in applications of yeast display include affinity maturation, protein engineering for improved production and stability, as well as novel applications in cell-based selections, epitope mapping, cDNA library screening, cell adhesion molecule engineering, and selections against non-biological targets[68]. Particular advantages of yeast display over other *in vitro* evolution methods include eukaryotic

expression and processing, quality control mechanisms (ensuring that only properly folded proteins reach the cell surface), minimal avidity effects, and also quantitative screening through FACS[69]. One drawback of yeast display is differential glycosylation in yeast compared to mammalian cells. Systemic and high throughput examples of enzyme evolution has not yet to be described[70], but recently horseradish peroxidase has been evolved using yeast display to show slightly higher enantioselectivity[71].

Phage display

One of the most common protein evolution methods is phage display, which has been widely used to engineer antibodies[72] and protein binders[73]. The protein of interest is displayed on the surface of the phage and the encoding gene is encapsulated inside the phage. One advantage of phage display is the relative robustness of the phage particle under variety of selection conditions, which is different from yeast display. However, poor display of some proteins on the phage surface and poor production of phage clones displaying certain proteins have been main disadvantages of the method[74].

Numerous enzymes have been displayed on the phage surface in active conformations[75]. Directed enzyme evolution by phage display requires a technique that couples the catalytic ability of an enzyme with the binding ability acquired as a result of the reaction[76]. Several examples using enzyme inhibitors or designed substrates have been reported. Recently, novel catalytic antibodies that hydrolyze an aryl phosphate were identified by a selection with 2-difluoromethylphenyl phosphate as substrate[77]. By design, the hydrolysis product subsequently rearranged to an electrophilic quinonemethide, which immediately formed a covalent bond with the enzyme and its associated phage particle.

In vitro compartmentalization (IVC)

In IVC, the genotype-phenotype linkage is provided by synthetic compartments of aqueous solutions within a water-in-oil emulsion[78]. Each compartment contains a single gene on average. By virtue of allowing single genes per compartment, the desired activity secures the survival of the fittest genes directly[79]. Reactions in these compartments can be controlled without compromising the chemical integrity of the droplets. Currently, IVC is the only *in vitro* evolution method for enzymes because it allows multiple-turnover selection. Recent examples of IVC include *in vitro* evolution of streptavidin binding desthiobiotin[80] and of Ebg enzyme (originally with negligible β-galactosidase activity) with high β-galactosidase activity[60].

The expression of the protein of interest is a prerequisite for the evolution methods described above. Differential protein expression may cause biases in selection. In order to circumvent this problem, *in vitro* translation system is used in IVC, and thus protein libraries with a higher diversity (~10^{14}) can be constructed (Other methods such as mRNA display and ribosome display also use *in vitro* translation system.) However, since the selected enzymatic reaction takes place in the same compartment where *in vitro* transcription and translation are carried out, complications may arise as a result of the specific environment (pH, temperature, ionic strength, etc.)[79] Also, substrates are immobilized in most cases, and thus it will drive the evolved enzymes to a low K_M[79].

Part 3. biotin ligase and lipoic acid ligase in *E. coli*

Ideal enzyme candidates for site-specific protein labeling would preferably catalyze

small molecule incorporation onto a small peptide substrate, where the reaction shows high specificity for the target peptide and the enzyme can tolerate some degree of structural changes of the small molecule substrate. Two *E. coli* protein ligases (BirA and LplA) have been extensively studied in our lab. BirA shows high sequence specificity toward its peptide substrate and LplA has high tolerance against various small molecule substrates.

Biotin and lipoic acid are the covalently bound cofactors of several enzyme complexes involved in key metabolic reactions (Figure 3A and 3B)[81]. Their attachment to a protein called biotinyl- or lipoyl domain is catalyzed by biotinyl protein ligase (BPL) or lipoyl protein ligase (LPL), respectively[82]. BPL and LPL have a homologous catalytic module, thus their reaction mechanism is very similar[82]; biotin and lipoic acid are first activated to an adenylated intermediate by BPL or LPL, and then transferred to a specific lysine residue on biotinyl- or lipoyl domain. The reaction mechanism is analogous to that of acyl-tRNA synthetases, suggesting a common ancestral relationship[81].

In *E. coli,* ligation reactions by biotin ligase (BirA) and lipoic acid ligase (LplA) are both extremely sequence specific. Biotin carboxyl carrier protein (BCCP) is the only protein biotinylated by BirA out of >4000 proteins in *E. coli*[83], and LplA has only four substrates (pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, branched-chain 2-oxoacid dehydrogenase, and the glycine cleavage system) [84]. One difference between biotinylation and lipoylation in *E. coli* is existence of the second lipoylation enzyme LipB, which uses lipoyl-acyl carrier protein (lipoyl-ACP) as the donor of the lipoyl group[85].

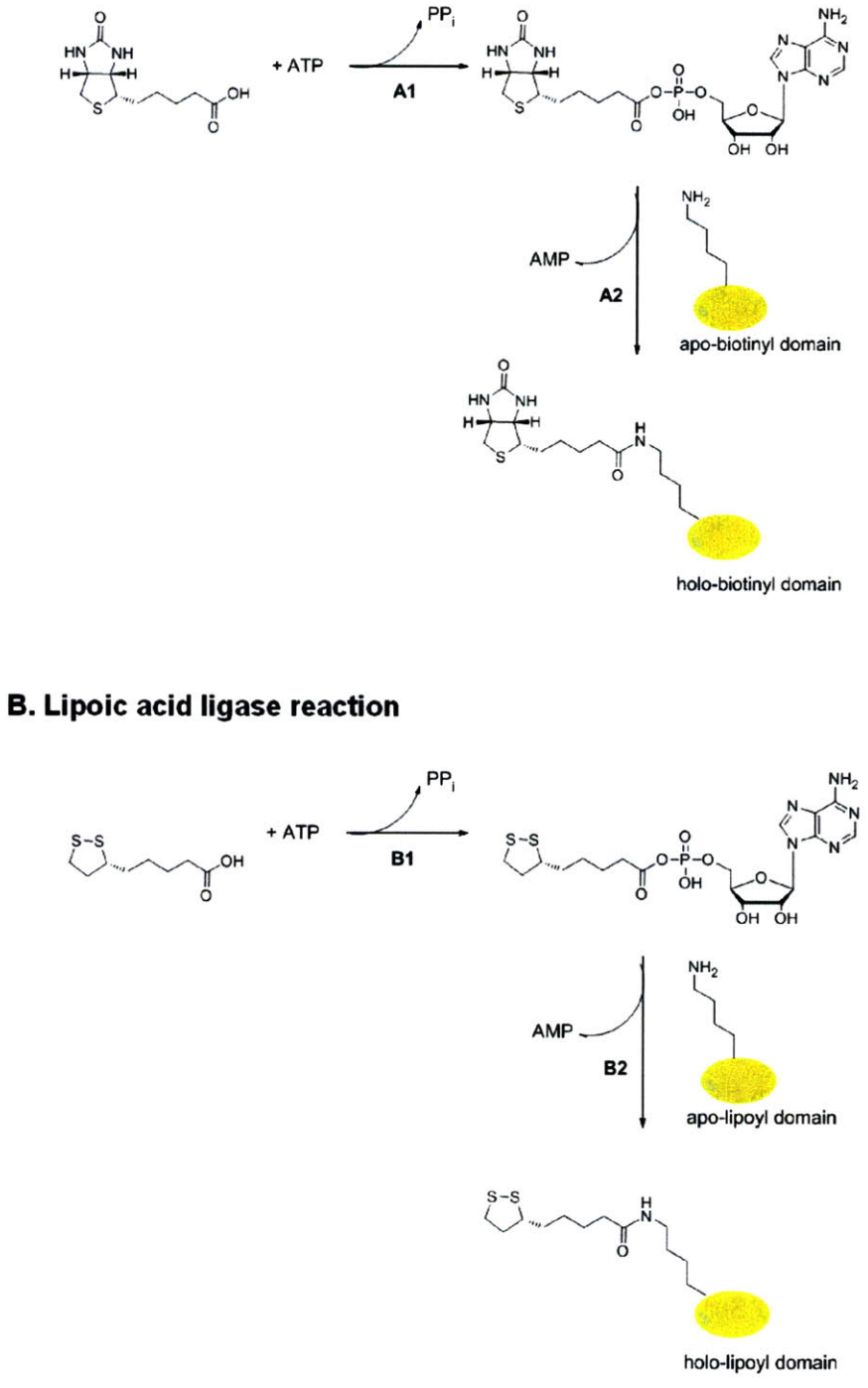


Figure 3 Reactions catalyzed by (A) biotin protein ligase, and (B) lipoic acid protein ligase.

E. coli biotin protein ligase (BirA)

Biotin is a cofactor for many biotin-dependent carboxylases, where biotin is required for the transfer of carbon dioxide from bicarbonate to organic acid metabolites[86]. Biotin is covalently attached to this family of enzymes by BPL. As the biotin-dependent enzymes are turned over, biotin is hydrolyzed off the degraded peptide fragments by biotinidase, allowing recycling of the cofactor[86]. The biotin-dependent enzymes are all functional carboxylases and are essential due to their involvement in fatty acid synthesis, gluconeogenesis and amino acid catabolism[87]. BirA is a BPL of *E. coli* (35.5 kDa), and it also regulates the *E. coli* biotin synthesis by binding to the biotin operator transcription site in the form of a BirA-biotinyl-5'-adenylate complex, thereby repressing transcriptions of genes related to biotin synthesis[88].

Crystal structure of biotin bound BirA shows that BirA monomer is organized in three domains[89]. The N-terminal domain, which contains a helix-turn-helix motif, is responsible for binding DNA when BirA forms dimers. The C-terminal domain functions in the transfer of biotin onto BCCP. The final domain Ⅱ has biotin- and ATP-binding sites and contains several poorly defined loop regions, which are associated with conformational changes in BirA upon ligand binding. The biotin-binding loop (residues 116-124) becomes well-ordered after biotin binds BirA. In addition to packing interactions with the thiophene ring and the aliphatic tail of biotin, several hydrogen bonds are formed between the main chain atoms of Arg 116 and the ureido ring of biotin. The carbonyl oxygen of the biotin forms hydrogen bonds with the side chains of Arg 116 and Ser 89, and ureido nitrogens interact with the carbonyl oxygen of Arg 116 and the side chain oxygen atoms of Thr 90 and Gln 112. These hydrogen bonds are responsible for specific binding of BirA to biotin. The adenylate binding site is fully formed only after the biotin-binding loop is ordered[88]. Arg 118 and Arg 121 make the critical

phosphate binding interactions[88].

The kinetic analysis of biotin ligation reaction by BirA under steady state conditions demonstrate that the K_M values for apo BCCP87 (87-residue C-terminal fragment of BCCP, functionally identical to BCCP in biotin transfer), biotin, and ATP are 4.39 ± 0.37 μM, 0.49 ± 0.07 μM, and about 0.3 mM, respectively[90]. The k_{cat} value is 0.16 ± 0.09 s^{-1} [90].

BirA can biotinylate BPL protein substrates derived from other organisms[91]. BPL-catalyzed reaction system for biotin addition is highly conserved, particularly in the region of target lysine residues of biotinyl domains. A Met-Lys-Met motif, where the middle lysine is biotinylated, is essentially invariant in all biotinyl domains[92]. Screening for a novel peptide substrate, which can be biotinylated by BirA, has led to the identification of acceptor peptide (AP; peptide sequence: GLNDIFEAQKIEWHE)[93,94]. The sequence of AP bears little similarity to the known biotinylated sequences of BCCP subunits, but the specificity constant (k_{cat}/K_M) is nearly identical to that measured for the natural protein substrate BCCP (11,900 ± 400 M^{-1} s^{-1} for BCCP where 10,000 ± 500 M^{-1} s^{-1} for AP)[94]. The K_M value of AP is around 25 μM[94], which is higher than BCCP, and due to its small size AP will bind BirA in a different manner where BCCP binds BirA (It would be hard to mimic the structural conformation of the protein substrate with a 15-mer peptide.)

E. coli lipoic acid protein ligase (LplA)

Lipoic acid is an extremely widely distributed cofactor which is essential for activity of a variety of enzymes catalyzing oxidative decarboxylations[95]. The pyruvate dehydrogenase complex of *E. coli* consists of three enzymes; pyruvate dehydrogenase (E1), dihydrolipoamide acetyl transferase (E2), and lipoamide dehydrogenase (E3)[96]. The reaction intermediate is

transferred between these three types of active sites via linkage to the lipoyl-lysine 'swinging arms' that are located in the E2p chains[97]. The specific attachment of lipoic acid to a single lysine of E2 domains is catalyzed by LplA (38 kDa).

Crystal structures of native and lipoic acid-bound forms of *E. coli* LplA are initially reported[98], but structural information, providing more insights for lipoylation by LplA, was obtained from lipoyl-AMP complex of LplA of *Thermoplasma acidophilum* (Ta)[99]. The lipoyl group of lipoyl-AMP is buried completely inside the bifurcated pocket. The dithiolane ring and the aliphatic chain of lipoyl-AMP are surrounded by hydrophobic residues of Ta LplA. The side chains of two histidine residues (His 81 and His 161) and the aliphatic side chains of Leu 18, Ile 46, Arg 72, and Ala 163 form hydrophobic environment for the dithiolane ring. Since no hydrogen bond exists between LplA and the lipoyl group, the substrate specificity of LplA is apparently lower than that of BirA, which forms multiple hydrogen bonds with biotin. In addition to its preferred substrate D-lipoic acid, LplA can also incorporates L-lipoic acid and octanoic acid[100,101].

The K_M(app) value for lipoic acid incorporation by LplA under steady state conditions is 1.7 μM[100] or 4.5 μM[98]. The k_{cat} value is 0.253 ± 0.003 s^{-1} [29].

The structural analysis shows that the lipoyl and biotinyl domains are closely related[81]. BirA and LplA distinguish their protein substrates by the existence of the protruding thumb region in biotinyl domains. When this thumb is removed, the biotinyl domain is almost as efficient a substrate as the native E2p lipoyl domain for lipoylation. In contrast to the rather free lipoyl-lysine residue in E2p, the biotinyl-lysine residue is tightly bound in *E. coli* BCCP by the interactions between the biotin moiety and residues in the thumb region. Although a large excess of LplA can lipoylate the biotinyl domain with a Met-Lys-Met motif, an Asp-Lys-

Ala motif of the lipoyl domain is not accepted by BirA for biotinylation. The peptide substrate of LplA (Lipoic acid Acceptor Peptide, LAP) has been designed and validated by others in the lab. BirA Acceptor Peptide AP was not lipoylated by LplA, and LAP in turn was not biotinylated by BirA[102]

Applications of BirA and LplA

Biotinylation has been widely used for affinity-based protein coupling in many biological assays, since biotin binds Avidin (or streptavidin) with exceptionally strong affinity ($K_d = 10^{-15}$ M)[103]. Either endogenous BPL or exogenous BPL (BirA in most cases) has been utilized to biotinylate the protein of interest which is fused to the biotinyl domain. Due to the small size of AP (compare to native BCCP substrate protein) and the great specificity of this peptide for the BirA reaction, biotinylation of AP fused proteins with BirA has been widely investigated in bacteria[104] and mammalian cells[105]. BirA also specifically incorporates desthiobiotin, a sulfur missing form of biotin, onto biotinyl domain, where the K_M value for desthiobiotin is about 40-fold higher than that for biotin[106]. More surprisingly, others in our lab found that BirA can incorporate a ketone isostere of biotin to the target biotinyl sequence with an only 3.7-fold lower initial reaction rate than biotin[27]. Ketones are absent from native cell surface. Proteins on cell surfaces can be site-specifically labeled with the ketone biotin by this BirA system and subsequently labeled with fluorophore by using the highly specific chemical reaction between with fluorophore hydrazide and the ketone biotin. Our lab has also developed a quantum dot (QD) targeting method using BirA/AP system and streptavidin QDs[107].

Unlike biotinylation, lipoylation has not been utilized for bioassays since no strong binder for lipoic acid is available. As mentioned in Part 1 of this chapter, however, our lab has

demonstrated site-specific protein modification with azide by LplA[29].

The major goal of my thesis work was to develop new site-specific small molecule labeling methodologies by engineering key elements of the biotin and lipoic acid ligase platforms. In Chapter 2, I discuss yeast display selection for kinetically improved peptide substrates of yeast biotin ligase. The selection produced new peptide substrates for yeast biotin ligase, but they likely have activity comparable to the previous yeast acceptor peptide discovered by phage display[108]. Establishment of the yeast display platform for kinetically improved peptide substrates, however, has led to the improvement of LplA's acceptor peptide, achieved by others in our lab[102]. We also attempted to engineer BirA to have faster kinetics for AP biotinylation by using *in vitro* compartmentalization (IVC) as discussed in Chapter 3. A novel bead-less IVC scheme for BirA evolution was successfully developed. Faster BirA selection was, however, restricted by technical difficulties of IVC and the nature of the selection (single turn-over selection). We further expanded the IVC selection strategy and developed a platform for streptavidin evolution using IVC. The development of a new streptavidin and biotin analog pair will extend the range of current streptavidin-biotin technology. Selection for streptavidin variants which show high affinities toward ketone biotin was performed with two rationally designed libraries (Chapter 4). Streptavidin evolution for ketone biotin binding was, however, inconclusive due to the lack of suitable assays to evaluate the progress of the selection.

We also have engineered the small molecule substrate specificity of LplA. At the start of our work, lipoic acid, selenolipoic acid, and octanoic acid were the only known small molecule substrates of *E. coli* LplA[101,109]. Among many biophysical probes, this thesis focuses

on the development of photoaffinity labels that can be incorporated onto proteins in a site-specific manner. Compared to fluorophore labeling of proteins, targeting methods for photoaffinity labels are highly limited. Photoaffinity labeling, however, offers additional information that fluorophores cannot provide, such as the identity of endogenous interaction partners for a protein of interest (even transient interactions)[110]. In Chapter 5, engineering of LplA to incorporate a photoaffinity probe (aryl azide analog) is demonstrated.

Future directions

Many methods of site-specific cellular protein labeling have been developed during the last decade, and we now have a better tool box to study cell biology than ever before. Some challenges for the future include the following. First, it would be desirable to have labeling methods that tag endogenous, rather than recombinant proteins. Recombinant proteins are often overexpressed, and this together with their tags, can alter biological function. Some methods exist to tag endogenous proteins, such as fluorescent ligands and antibodies, but these are usually labile on the timescale of minutes. Second, major improvements in enzyme-based technologies such as those developed by our lab will require successful application of *in vitro* evolution and/or computational techniques. *In vitro* evolution has so far been most successful for discovering novel binders, such as new antibodies. Evolving new catalytic activity is far more challenging, but some notable examples exist, particularly with the aid of computational methods[111]. We note that methods to evolve novel enzymes will have wide-ranging impact, not just in cellular imaging, but also in the field of organic chemistry, such as by enabling simpler and more environment-friendly synthesis[112].

References

[1] RSAS The green fluorescent protein: discovery, expression and development in *Scientific Background on the Nobel Prize in Chemistry 2008* (Stockholm, 2008).

[2] Lippincott-Schwartz, J. & Patterson, G., Development and use of fluorescent protein markers in living cells. *Science* 300 (5616), 87-91 (2003).

[3] Marguet, D., 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* 11, 231-240 (1999).

[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* 4, 491-501 (2003).

[5] Baens, M., The Dark Side of EGFP: Defective Polyubiquitination. *PLpS. ONE* 1, e54 (2006).

[6] Young, D. & Deiters, A., Photochemical control of biological processes. *Org Biomol Chem* 5 (7), 999-1005 (2007).

[7] Chen, I. & Ting, A.Y., Site-specific labeling of proteins with small molecules in live cells. *Curr. Opin. Biotechnol.* 16, 35-40 (2005).

[8] Marks, K.M. & Nolan, G.P., Chemical labeling strategies for cell biology. *Nat. Methods* 3, 591-596 (2006).

[9] Israel, D. & Kaufman, R., Dexamethasone negatively regulates the activity of a chimeric dihydrofolate reductase/ glucocorticoid receptor protein. *Proc Natl Acad Sci USA* 90, 4290-4294 (1993).

[10] Marks, K.M., Braun, P., & Nolan, G., A general approach for chemical labeling and rapid, spatially controlled protein inactivation. *Proc Natl Acad Sci USA* 101, 9982-

9987 (2004).

[11] Farinas, J. & Verkman, A.S., Receptor-mediated targeting of fluorescent probes in living cells. *J. Biol. Chem.* 274, 7603-7606 (1999).

[12] Los, G. & Wood, K., The HaloTag: a novel technology for cell imaging and protein analysis. *Methods Mol Biol* 356, 195-208 (2007).

[13] Pries, F. *et al.*, Histidine 289 is essential for hydrolysis of the alkyl-enzyme intermediate of haloalkane dehalogenase. *J. Biol. Chem.* 270, 10405-10411 (1995).

[14] Keppler, A. *et al.*, Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods* 32, 437-444 (2004).

[15] Bonasio, R. *et al.*, Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots. *Proc Natl Acad Sci USA* 104, 14753-14758 (2007).

[16] Guignet, E.G., Hovius, R., & Vogel, H., Reversible site-selective labeling of membrane proteins in live cells. *Nat. Biotechnol* 22, 440-444 (2004).

[17] Griffin, B.A., Adams, S.R., & Tsien, R.Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281, 269-272 (1998).

[18] Gaietta, G. *et al.*, Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296 (5567), 503-507 (2002).

[19] Griffin, B.A., Adams, S.R., Jones, J., & Tsien, R., Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Methods Enzymol.* 327, 565-578 (2000).

[20] Langhorst, M.F., Genisyuerek, S., & Stuermer, C.A.O., Accumulation of FlAsH/Lumio Green in active mitochondria can be reversed by β-mercaptoethanol for specific staining of tetracysteine-tagged proteins. *Histochem. Cell Biol.* 125, 743-747 (2006).

[21] Lin, M.Z. & Wang, L., Selective labeling of proteins with chemical probes in living

cells. *Physiology* 23, 131-141 (2008).

22 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* 124, 6063-6076 (2002).

23 Hauser, C. & Tsien, R., A hexahistidine-Zn^{2+}-dye label reveals STIM1 surface exposure. *Proc Natl Acad Sci USA* 104, 3693-3697 (2007).

24 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* 128, 10452-10459 (2006).

25 Franz, K., Nitz, M., & Imperiali, B., Lanthanide-binding tags as versatile protein coexpression probes. *Chembiochem* 4, 265-271 (2003).

26 Marks, K., Rosinov, M., & Nolan, G., In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem Biol* 11, 347-356 (2004).

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

28 Slavoff, S., Chen, I., Choi, Y., & Ting, A.Y., Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *J. Am. Chem. Soc* 130, 1160-1162 (2008).

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

30 Baruah, H., Puthenveetil, S., Choi, Y.A., Shah, S., & Ting, A.Y., An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* 47 (37), 7018-7021 (2008).

31 Yin, J. *et al.*, Genetically encoded short peptide tag for versatile protein labeling by Sfp

phosphopantetheinyl transferase. *Proc Natl Acad Sci USA* 102 (44), 15815-15820 (2005).

[32] Meier, J.L., Mercer, A.C., Rivera, H.J., & Buckart, M.D., Synthesis and evaluation of bioorthogonal pantethein analogues for in vivo protein modification. *J. Am. Chem. Soc* 128, 12174-12184 (2006).

[33] Yin, J., Liu, F., Li, X.H., & Walsh, C.T., Labeling proteins with small molecules by site-specific posttranslational modification. *J. Am. Chem. Soc* 126, 7754-7755 (2004).

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

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

[36] Lin, C. & Ting, A., Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc* 128, 4542-4543 (2006).

[37] Duckworth, B.P., Zhang, Z., Hosokawa, A., & Distefano, M.D., Selective labeling of proteins by using protein farnesyltransferase. *Chembiochem* 8 (98-105) (2007).

[38] Wang, L. & Schultz, P.G., Expanding the genetic code. *Chem. Commun. (Camb.)*, 1-11 (2002).

[39] Wang, L. & Schultz, P.G., Expanding the genetic code. *Angew Chem Int Ed Engl* 44, 34-66 (2004).

[40] Dawson, P.E., Muir, T.W., Clark-Lewis, I., & Kent, S.B., Synthesis of proteins by native chemical ligation. *Science* 266, 776-779 (1994).

[41] Muir, T.W., Sondhi, D., & Cole, P.A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95, 6705-6710 (1998).

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

[43] Chin, J.W. *et al.*, Addtion of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc* 124, 9026-9027 (2002).

[44] Summerer, D. *et al.*, A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci USA* 103, 9785-9789 (2006).

[45] Wang, J., Xie, J., & Schultz, P.G, A genetically encoded fluorescent amino acid. *J. Am. Chem. Soc* 128, 8738-8739 (2006).

[46] Zhang, Z. *et al.*, A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* 42, 6735-6746 (2003).

[47] Deiters, A. *et al.*, Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J. Am. Chem. Soc* 125, 11782-11783 (2003).

[48] Xie, J. & Schultz, P.G, A chemical toolkit for proteins: an expanded genetic code. *Nat. Rev. Mol. Cell. Biol.* 7, 775-782 (2006).

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

[50] Lin, M.Z. & Wang, L. *Physiology* 23, 131-141 (2008).

[51] Cedrone, F., Menez, A., & Quemeneur, E., Tailoring new enzyme functions by rational redesign. *Curr. Opin. Struct. Biol.* 10, 405-410 (2000).

[52] Bolon, D.N., Voigt, C.A., & Mayo, S.L., De novo design of biocatalysts. *Curr. Opin.*

Chem. Biol. 6, 125-129 (2002).

[53] Arnold, F.H., Design by directed evolution. *Acc. Chem. Res.* 31, 125-131 (1998).

[54] Arnold, F.H., Combinatorial and computational challenges for biocatalyst design. *Nature* 409, 253-257 (2001).

[55] Wilson, D.S. & Szostak, J.W., In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* 68, 611-647 (1999).

[56] Georgiou, G. *et al.*, Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat. Biotechnol* 15, 29-34 (1997).

[57] Kehoe, J.W. & Kay, B.K., Filamentous phage display in the new millennium. *Chem. Rev.* 105, 4056-4072 (2005).

[58] Roberts, R.W. & Szostak, J.W., RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci USA* 94, 12297-12302 (1997).

[59] Hanes, J. & A., P., In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 94, 4937-4942 (1997).

[60] 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* 12, 1291-1300 (2005).

[61] Francisco, J.A., Campbell, R., Iverson, B.L., & Georgiou, G., Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. *Proc Natl Acad Sci USA* 90, 10444-10448 (1993).

[62] Boder, E.T. & Wittrup, K.D., Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol* 15, 553-537 (1997).

[63] Aharoni, A., Griffith, A.D., & Tawfik, D.S., High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* 9, 210-216 (2005).

[64] Chen, G. *et al.*, Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nat. Biotechnol* 19, 537-542 (2001).

[65] Harvey, B.R. *et al.*, Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries. *Proc Natl Acad Sci USA* 101, 9193-9198 (2004).

[66] Olsen, M.J. *et al.*, Function-based isolation of novel enzymes from a large library. *Nat. Biotechnol* 18, 1071-1074 (2000).

[67] Boder, E.T. & Wittrup, K.D., Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol.* 328, 430-444 (2000).

[68] Pepper, L.R., Cho, Y.K., Boder, E.T., & Shusta, E.V., A decade of yeast surface display technology: Where are we now? *Comb. Chem. High Throughput Screen.* 11, 127-134 (2008).

[69] Colby, D.W. *et al.*, Engineering antibody affinity by yeast surface display. *Methods Enzymol.* 388, 348-358 (2004).

[70] Gai, S.A. & Wittrup, K.D., Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* 17, 467-473 (2007).

[71] Antipov, E., Cho, A.E., Wittrup, K.D., & Klibanov, A.M., Highly L and D enantioselective variants of horseradish peroxidase discovered by an ultrahigh-throughput selection method. *Proc Natl Acad Sci USA* 105, 17694-17699 (2008).

[72] Winter, G., Griffiths, A.D., Hawkins, R.E., & Hoogenboom, H.R., Making antibodies by phage display technology. *Annu. Rev. Immunol.* 12, 433-455 (1994).

[73] Hoess, R.H., Protein design and phage display. *Chem. Rev.* 101, 3205-3218 (2001).

[74] Sidhu, S.S., *Phage display in biotechnology and drug discovery*. (CRC Press, 2005).

[75] Fastrez, J., Investigation of phage display for the directed evolution of enzymes in *Directed Molecular Evolution of Proteins*, edited by S. Brakmann & K. Johnsson (Wiley VCH, Weinheim, 2002), pp. 79-110.

[76] Fernandez-Gacio, A., Uguen, M., & Fastrez, J., Phage display as a tool for the directed evolution of enzymes. *Trends. Biotechnol.* 21, 408-414 (2003).

[77] Cesaro-Tadic, S. *et al.*, Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library. *Nat. Biotechnol* 21, 679-685 (2003).

[78] Tawfik, D.S. & Griffith, A.D., Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol* 16, 652-656 (1998).

[79] Griffith, A.D. & Tawfik, D.S., Man-made enzymes—from design to in vitro compartmentalisation. *Curr. Opin. Biotechnol.* 11, 338-353 (2000).

[80] Levy, M. & Ellington, A.D., Directed evolution of streptavidin variants using in vitro compartmentalization. *Chem Biol* 15, 979-989 (2008).

[81] Reche, P. & Perham, R.N., Structure and selectivity in post-translational modification: attaching the biotinyl-lysine and lipoyl-lysine swinging arms in multifunctional enzymes. *EMBO J.* 18, 2673-2682 (1999).

[82] Reche, P., Lipoylating and biotinylating enzymes contain a homologous catalytic module. *Protein Sci.* 9, 1922-1929 (2000).

[83] Chapman-Smith, A., Mulhern, T.D., Whelan, F., Cronan, J.E., & Wallace, J.C., The C-terminal domain of biotin protein ligase from E. coli is required for catalytic activity. *Protein Sci.* 10, 2608-2617 (2001).

[84] 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* 10, 1293-1302 (2003).

[85] Morris, T.W., Reed, K.E., & Cronan, J.E., Lipoic acid metabolism in Escherichia coli: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. *J. Bacteriol.* 177, 1-10 (1995).

[86] Pendini, N.R. *et al.*, Microbial biotin protein ligases aid in understanding holocarboxylase synthetase deficiency. *Biochim. Biophys. Acta* 1784, 973-982 (2008).

[87] Wolf, B., Disorders of biotin metabolism in *The metabolic and molecular basis of inherited diseases*, edited by C. R. Scriver, A. L. Beavder, & W. S. Sly (Mcgraw Hill, New York, 1995), Vol. 7, pp. 3151-3177.

[88] Wood, Z.A., Weaver, L.H., Brown, P.H., Beckett, D., & Matthews, B.W., Co-repressor Induced Order and Biotin Repressor Dimerization: A Case for Divergent Followed by Convergent Evolution. *J. Mol. Biol.* 357, 509-523 (2006).

[89] Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J., & Matthews, B.W., Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc Natl Acad Sci USA* 89, 9257-9261 (1992).

[90] Chapman-Smith, A., Morris, T.W., Wallace, J.C., & Cronan, J.E., Molecular Recognition in a Post-translational Modification of Exceptional Specificity. *J. Biol. Chem.* 274, 1449-1457 (1999).

[91] Chapman-Smith, A. & Cronan, J.E., The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends. Biochem. Sci.* 24, 359-363 (1999).

[92] Weaver, L.H., Kwon, K., Beckett, D., & Matthews, B.W., Competing protein:protein interactions are proposed to control the biological switch of the E coli biotin repressor. *Protein Sci.* 10, 2618-2622 (2001).

[93] Schatz, P.J., Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. *Biotechnology* 11, 1138-1143 (1993).

[94] Beckett, D., Kovaleva, E., & Schatz, P.J., A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 8, 921-929 (1999).

[95] Cronan, J.E., Zhao, X., & Jiang, Y., Function, attachment and synthesis of lipoic acid in Escherichia coli. *Adv. Microb. Physiol.* 50, 103-146 (2005).

[96] Reed, L.J., Multienzyme complexes. *Acc. Chem. Res.* 7, 40-46 (1974).

[97] Bleile, D.M., Munk, P., Oliver, R.M., & Reed, L.J., Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from Escherichia coli. *Proc Natl Acad Sci USA* 76, 4385-4389 (1979).

[98] Fujiwara, K. *et al.*, Crystal Structure of Lipoate-Protein Ligase A from Escherichia coli. *J. Biol. Chem.* 280, 33645-33651 (2005).

[99] Kim, D.J. *et al.*, Crystal Structure of Lipoate-Protein Ligase A Bound with the Activated Intermediate. *J. Biol. Chem.* 280, 38081-38089 (2005).

[100] Green, D.E., Morris, T.W., Green, J., Cronan, J.E., & Guest, J.R., Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem. J.* 309, 853-862 (1995).

[101] Morris, T.W., Reed, K.E., & Cronan, J.E., Identification of the gene encoding lipoate-protein ligase A of Escherichia coli. Molecular cloning and characterization of the lplA

gene and gene product. *J. Biol. Chem.* 269, 16091-16100 (1994).

[102] Puthenveetil, S., Liu, D.S., White, K.A., Thompson, S., & Ting, A.Y., Yeast display evolution of novel kinetically efficient peptide substrates for lipoic acid ligase. *J. Am. Chem. Soc* 131, 16430-16438 (2009).

[103] Green, N.M., Avidin. *Adv. Protein Chem.* 29, 85-133 (1975).

[104] Cull, M.G. & Schatz, P.J., Biotinylation of proteins in vivo and in vitro using small peptide tags. *Methods Enzymol.* 326 (430-440) (2000).

[105] de Boer, E. *et al.*, Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc Natl Acad Sci USA* 100, 7480-7485 (2003).

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

[107] 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 USA* 102, 7583-7588 (2005).

[108] 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* 129, 6619-6625 (2007).

[109] Reed, K.E., Morris, T.W., & Cronan, J.E., Mutants of Escherichia coli K-12 that are resistant to a selenium analog of lipoic acid identify unknown genes in lipoate metabolism. *Proc Natl Acad Sci USA* 91, 3720-3724 (1994).

[110] Choi, S.K., Kalivretenos, A.G., Usherwood, P.N.R., & Nakanishi, K., Labeling studies

of photolabile philanthotoxins with nicotinic acetylcholine receptors: mode of interaction between toxin and receptor. *Chem. Biol.* 2, 23-32 (1995).

[111] Jiang, L. *et al.*, De novo computational design of retro-aldol enzymes *Science* 319, 1387-1391 (2008).

[112] Hult, K. & Berglund, P., Engineered enzymes for improved organic synthesis. *Curr. Opin. Biotechnol.* 14, 395-400 (2003).

Chapter 2: Improvement of Yeast Biotin Ligase Acceptor Peptide by Yeast Display Evolution

This chapter describes the characterization and utilization of yeast biotin ligase acceptor peptide (yAP), and subsequent efforts to evolve an improved yAP by yeast display. The first part of this chapter has been published in Chen, I., Choi, Y. A., and 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.* 129, 6619-6625 (2007).

Description of my specific contribution: I joined the project after yAP was evolved via phage display by Irwin Chen, a former graduate student in the Ting laboratory. While he characterized yAP using phage ELISA assays, I investigated the orthogonality of yAP and AP (the peptide substrate for *E. coli* biotin ligase) on yeast cell surface. I also characterized yAP and AP orthogonality on purified AP and yAP fusions to cyan fluorescent protein (CFP) and heterochromatin protein 1 (HP1). I helped to develop the labeling protocol for orthogonal QD targeting to yAP and AP fusion proteins on the mammalian cell surface.

The project involving yeast display evolution of yAP was performed independently, without collaborators.

Introduction

Biotinylation is widely used for protein labeling, immobilization, and purification as reviewed in Chapter 1. This is mostly due to the strong interaction between biotin and (strept)avidin[1] and to the existence of a peptide substrate (AP)[2,3] for biotin ligase. AP is a 15 amino acid peptide which is specifically biotinylated by BirA, the *E. coli* biotin ligase. It has been desirable to develop a protein labeling strategy with similar specificity using small tags. Moreover, if the reactivity of the new method is independent from that of the BirA reaction, both methods can be simultaneously utilized for multiple protein labeling. For example, many biological processes such as signal transduction involve several proteins, and by labeling (and thus monitoring) multiple proteins at the same time, we can get more insightful information on the biological event.

In order to achieve this goal, Irwin Chen in our lab tested biotinylation of AP using biotin ligases from other organisms, and observed that biotin ligases from *Bacillus subtilis*, *Methanococcus jannaschii*, yeast (*Saccharomyces cerevisiae*)[4] and humans[5] do not use AP[6]. He then used phage display to select peptide substrates that are active for biotin ligases from *B. subtilis*, *M. jannaschii*, and *S. cerevisiae* (bsBL, mjBL, and yBL, respectively). There was an increase in biotinylation activity over three rounds, and only yBL was found to be responsible for the peptide biotinylation. The selected peptides contained a common Met-(Thr/Glu)-Phe motif immediately following the biotinylation site lysine. Thus, the second generation selection was performed with a new library including the Met-(Thr/Glu)-Phe motif (34% of the library contained the sequence of this motif). The best peptide derived from the selection was named as yeast acceptor peptide (yAP). Characterization of yAP and its application to mammalian and yeast cell surface labeling will be discussed in Part I of the present chapter.

Although we were able to demonstrate orthogonal labeling of AP- and yAP-fusion proteins on cell surface, the catalytic efficiency for the biotinylation of the yAP by yBL was still 780 times worse than that reported for the biotinylation of the AP by BirA[3]. It was mostly due to a poor k_{cat} of yAP, which indicated the need of multiple turnover selection condition to improve the kinetic property of yAP. A model phage display selection was performed with yAP and the 2-3-5 peptide which is biotinylated by yBL about ten fold worse than yAP. With a 1:20 mixture of yAP-phage and the 2-3-5 phage, Irwin Chen could not see any enrichment of yAP over the 2-3-5-phage, indicating that the phage display system used for the selection does not provide the necessary dynamic range to distinguish between the two substrates for yBL.

Yeast surface display is a powerful tool for protein engineering, by which a protein can be evolved to get improved or novel binding affinity, catalytic activity, and/or structural properties such as high stability and better expression[7]. Developed by Boder and Wittrup about ten years ago[8], numerous examples of yeast display application have been reported[9], particularly in therapeutic antibody engineering. Other utilization for protein characterization, including the mapping of functional epitopes mediating protein-protein interactions[10], is emerging as well. Furthermore, proteins that target insoluble or unknown molecules are also engineered using yeast display[11]. Yeast cells can also serve as a physical support where proteins are immobilized, replacing solid resins[12].

Yeast display has several advantages over the phage display. First of all, the use of fluorescence-activated cell sorting (FACS) allows quantitative screening where one can observe the statistics of sample directly during the process[13]. Using double labeling of yeast cells, the antigen-binding signal can be normalized with surface protein expression, eliminating artifacts due to expression bias[14]. Recently, Bowley and coworkers compared the

two methods (the phage and yeast display) using identical antibody libraries and antigens[15]. The study showed that yeast display samples the members of a library considerably more fully[15]. In addition, the number of protein copies displayed on the surface is 10^4 - 10^5 for yeast whereas 3 - 5 for pIII protein or 2700 for pVIII protein of wild type M13 phage[16]. We envisioned that yeast display is more suitable to evolution of our peptide substrates than phage display since it will allow multiple turnover selection and finer discrimination between mutants.

In Part II of the present chapter, our attempts to improve the catalytic activity of the current yAP peptide substrate of yBL will be discussed. By performing a model selection with kinetically repressed peptides, we have successfully confirmed that yeast display actually provides a multiple-turnover condition for selection of yBL peptide substrates. The developed yeast display method for turnover-based yBL substrate selection will offer an excellent platform for further investigation of yAP/yBL labeling methodologies. In our knowledge, this is the first example of yeast display for peptide substrate evolution. Indeed, Dr. Sujiet Puthenveetil showed that LAP2 (LplA acceptor peptide 2), which has a >70 times better k_{cat}/K_M than the rationally engineered LAP[17], could be evolved using the yeast display system with a covalent product capture method[18]. LAP2 has almost same catalytic efficiency as a natural protein substrate for LplA.

Results and Discussion

Part I. Characterization and applications of phage-evolved yAP peptide

Summary of Irwin Chen's work Irwin Chen performed a model selection with a native AP peptide (GLNDIFEAQ**K**IEWHE) and an Ala mutant (GLNDIFEAQ**A**IEWHE), where the

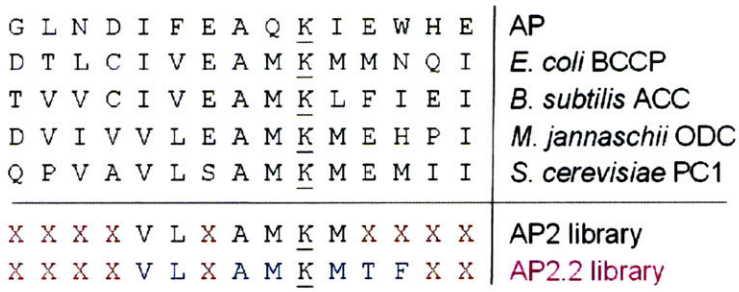


Figure 1 Design of the peptide library for yeast display selection based on AP2.2 library. The sequences of the AP, biotin ligase substrate proteins from four different species (*E. coli* BCCP (biotin carboxyl carrier protein), *B. subtilis* ACC (acetyl CoA-carboxylase), *M. jannaschii* ODC (oxaloacetate decarboxylase), and *S. cerevisiae* PC1 (pyruvate carboxylase 1)) and the first-generation AP2 library are shown for comparison. The lysine biotinylation sites are underlined. Red positions are completely randomized, while blue positions are partially randomized (~34% the indicated amino acid). This figure was reproduced from reference 6.

lysine biotinylation site is mutated to alanine. The AP and Ala peptides are displayed as a fusion to the pIII coat protein of M13 bacteriophage. The 1:100 mixture of AP-phage and Ala-phage was biotinylated by BirA, and the biotinylated phages were purified by streptavidin beads. Re-amplification of the collected phage gave a 2000-fold enrichment of AP over Ala.

The new peptide library was designed as follows. Five residues that are well-conserved in the protein substrates of biotin ligases in other organisms were fixed, and other residues except the lysine biotinylation site were fully randomized (Figure 1, AP2 library). The first generation of selections was then performed with 100 nM of three biotin ligases (bsBL, mjBL, and yBL), and the *E. coli* strain BM4092F' was used to produce phages in biotin-depleted media for the first two rounds of selection. This strain does not biosynthesize biotin, and it expresses a mutant of BirA with a higher K_M value for biotin. Use of this strain, therefore, removed the background biotinylation by endogenous BirA in *E. coli,* and a few members of the library which are biotinylated by the mixture of three biotin ligases could be

maintained at earlier rounds of selection. Subsequent ELISA assays with each biotin ligase showed that only yBL was responsible for the biotinylation activity in the selection. Evaluating individual clones from the third and fourth rounds by sequencing revealed that all peptides biotinylated by yBL contained a Met-(Thr/Glu)-Phe motif immediately following the lysine. Therefore, the second generation of selections was performed with a new library that is biased to the Met-(Thr/Glu)-Phe motif (Figure 1, AP2.2 library). Previously fixed residues were mutated so that 34% of the sequences would contain the designed amino acid, while the remaining 66% would contain any of the other 19 amino acids or stop codons (Figure 1).

Despite randomization in the design of the second generation library, the Met-(Thr/Glu)-Phe motif re-emerged largely intact even after the second generation of selections. The best isolated peptide was biotinylated by yBL ten-fold better than the best peptide from the first generation selection. He named this sequence the "yeast acceptor peptide" (yAP), and all subsequent characterizations were performed with this peptide.

Biotinylation of yAP fusion proteins I created yAP fusions to heterochromatin protein 1 (HP1) and cyan fluorescent protein (CFP), in order to test if yAP could be recognized by yBL in various different contexts. Thus, both N- and C-terminal fusions were prepared. Biotinylation was detected by blotting with streptavidin-horse radish peroxidase (HRP) (Figure 2). yAP was successfully biotinylated by yBL in an ATP-dependent manner, with some context-dependence. Biotinylation level of yAP at N-terminus of CFP was lower than that of yAP at C-terminus of CFP. Nevertheless our results demonstrate that the yAP can be recognized at either the N- or C-terminus of fusion proteins.

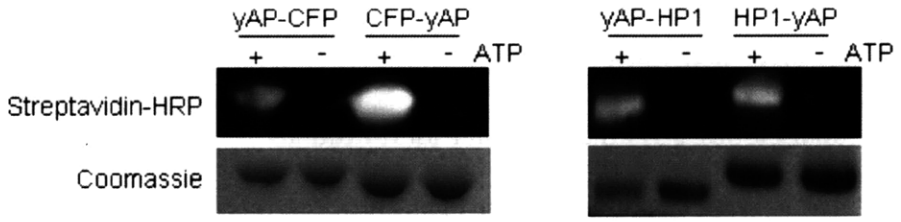


Figure 2 Biotinylation of yAP fusion proteins. N-and C-terminal yAP fusions to CFP and HP1 were biotinylated using yBL and analyzed by blotting with streptavidin-horse radish peroxidase (HRP) conjugate. Coomassie staining confirms equal loading of protein. This figure was reproduced from reference 6.

Orthogonality of the yAP/yBL pair We tested the orthogonality of the new yAP/yBL pair using three assays.

First, phage ELISA assays were performed by Irwin Chen. Irwin found that yAP is not biotinylated by BirA, and conversely that AP is not biotinylated by yBL (data not shown).

Second, working together with Irwin, we together performed orthogonal labeling of AP and yAP fusion proteins on the surface of live HeLa cells. Each peptide was fused to a distinct fluorescent protein marker expressed on the extracellular surface. First, we showed that yAP-YFP-TM construct could be biotinylated with yBL, then labeled with streptavidin-conjugated quantum dot (QD) 655. The QD targeting was specific for transfected cells (Figure 3); neighboring untransfected cells remained unlabeled. Negative controls showed that the biotinylation was ATP-dependent and site-specific at the lysine of the yAP sequence (Figure 3).

Next, HeLa cells that were individually transfected with yAP-YFP-TM or AP-CFP-TM were mixed together to populate the same dish. The cells on the dish were then biotinylated by yBL for 1 h at 30 °C, stained with streptavidin-QD655, biotinylated by BirA for 5 min, and stained with streptavidin-QD565. Figure 4 shows that yAP-displaying cells were only labeled

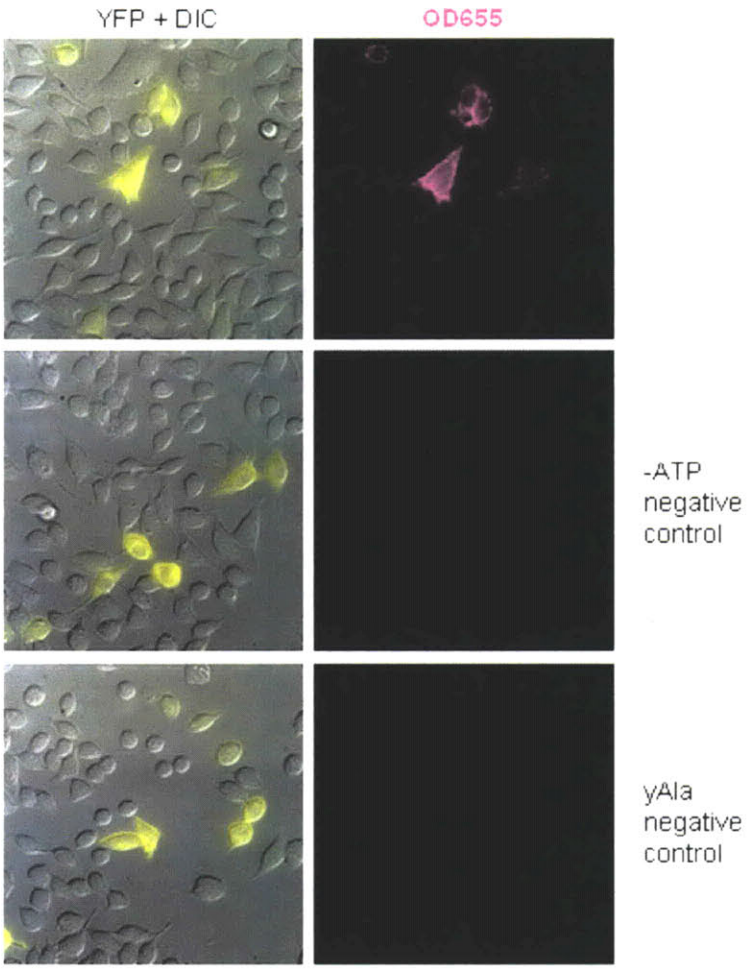


Figure 3 Site-specific biotinylation of yAP expressed on the surface of live HeLa cells. HeLa expressing the yAP-YFP-TM construct were biotinylated with yBL for 1 h at 32 °C and detected by staining with streptavidin-QD655 for 5 min. Negative controls with ATP omitted, or yAP-YFP-TM replaced by its yAla-YFP-TM point mutant (where the lysine biotinylation site is mutated to alanine), are shown. YFP images were merged and overlaid with the differential interference contrast (DIC) image. This figure was reproduced from reference 6.

with QD655 and AP-displaying cells were exclusively labeled with QD565, demonstrating that the orthogonality of yAP/yBL was also preserved on the mammalian cell surface.

Third, I performed an assay to test yAP and AP labeling on the surface of live yeast

cells. Irwin Chen prepared AP and yAP fusions to the Aga2p mating protein of yeast. Each peptide was successfully biotinylated only by the matching biotin ligase, as confirmed by streptavidin-phycoerythrin (PE) labeling and subsequent fluorescence-activated cell sorting (FACS) analysis (Figure 5). Both fusion proteins were displayed at the same levels on yeast cell surface, as indicated by anti-myc tag staining. No fluorescence signal was observed when the central lysine residue in yAP is mutated to alanine or when ATP is not available.

A technical note is that while performing orthogonality tests, we observed that yBL samples were often contaminated by endogenous BirA from *E. coli* cells, in which the yBL was overexpressed. To remove this contaminating BirA, we incubated yBL samples with

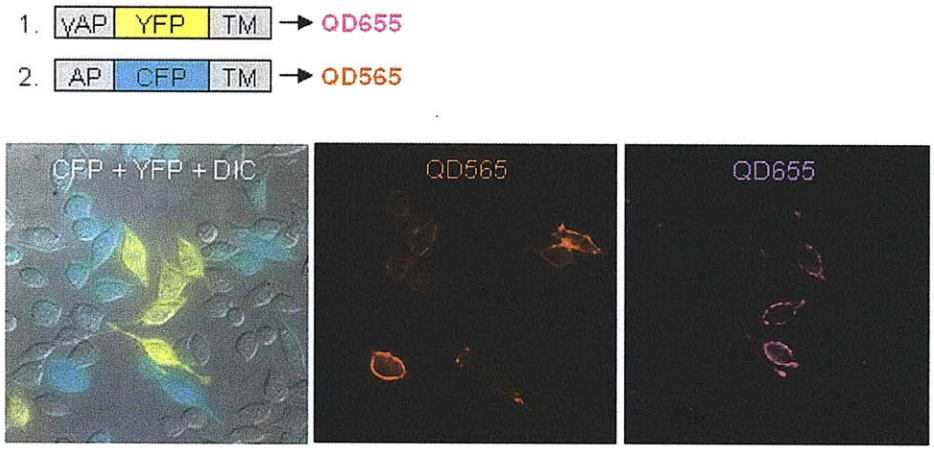


Figure 4 Selective labeling of live HeLa cells expressing yAP or AP fusion proteins (domain structures shown at top) with quantum dots. First, yAP-expressing cells (indicated by YFP fluorescence) were selectively labeled with streptavidin-QD655 conjugate, using yBL biotinylation. Then, AP-expressing cells (indicated by CFP fluorescence) in the same dish were labeled with streptavidin-QD565 conjugate, using BirA biotinylation. CFP and YFP images were merged and overlaid with the DIC image. This figure was reproduced from reference 6.

anti-BirA antibody and pulled down with secondary antibody-conjugated beads.

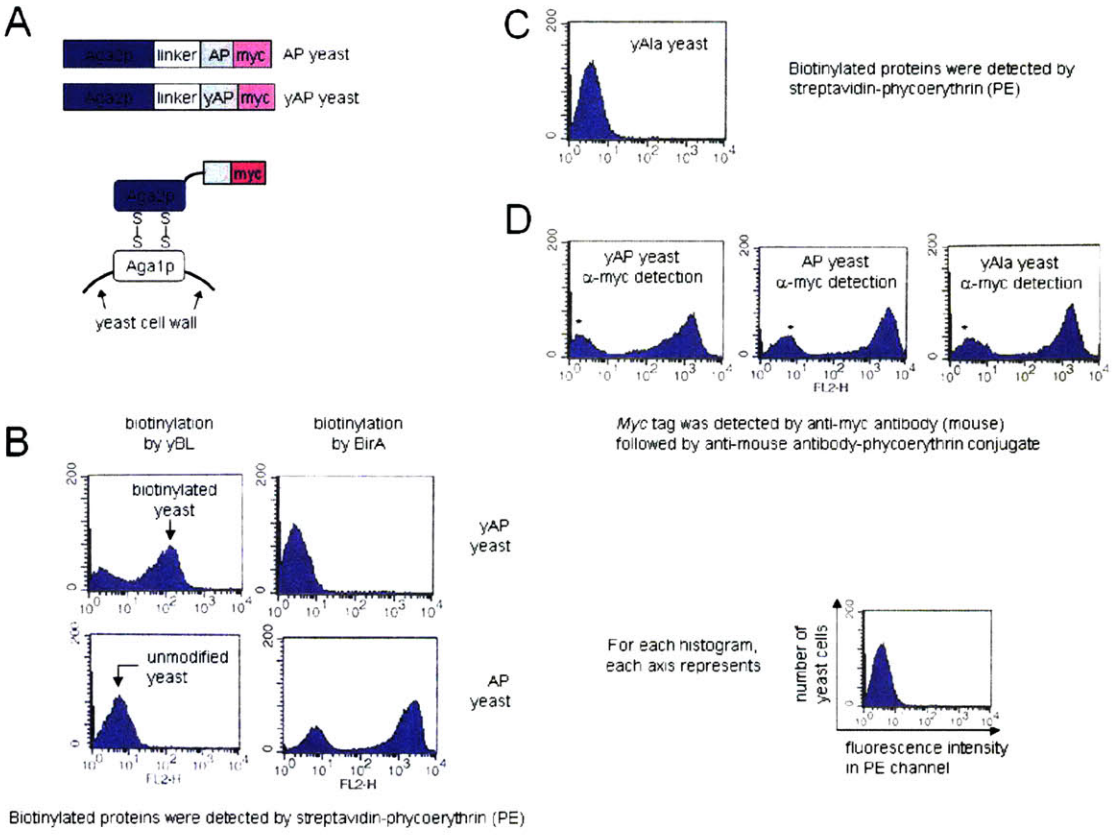


Figure 5 Orthogonal labeling on yeast cell surface. (A) Constructs used for peptide display on yeast surface. The Aga2p fusions were anchored to the yeast cell surface via disulfide bonds to the Aga1p protein. (B) Yeast cells expressing the yAP or AP on their surface were biotinylated with either yBL or BirA. After staining with streptavidin-phycoerythrin conjugate, yeast cells were analyzed by flow cytometry. The histograms show the number of cells (y-axis) counted at different fluorescence intensities (FL2, x-axis). (C) The alanine mutant of yAP (yAla) was used as a negative control. The fluorescence histogram is shown after treatment with yBL as in (B). (D) To verify similar levels of surface expression, the yeast cells were stained with an anti-myc antibody, followed by anti-mouse antibody conjugated to phycoerythrin. Typically, 20 - 25% of the yeast cells did not display any Aga2p fusion, and thus did not become labeled (*). This figure was reproduced from reference 6.

Part II. Yeast display evolution of improved yAP sequences

Model selection (1) Two different model selections were performed to confirm that yeast display selection can provide multiple turnover selective pressure. First, model selections were performed to enrich yAP over yAla (where the active lysine residue is mutated to alanine) (Figures 6 and 7). Since the genes encoding these constructs have the same sizes, we created a restriction enzyme site (BstXI) on the DNA sequence of yAla in order to distinguish it from yAP by the restriction digestion pattern.

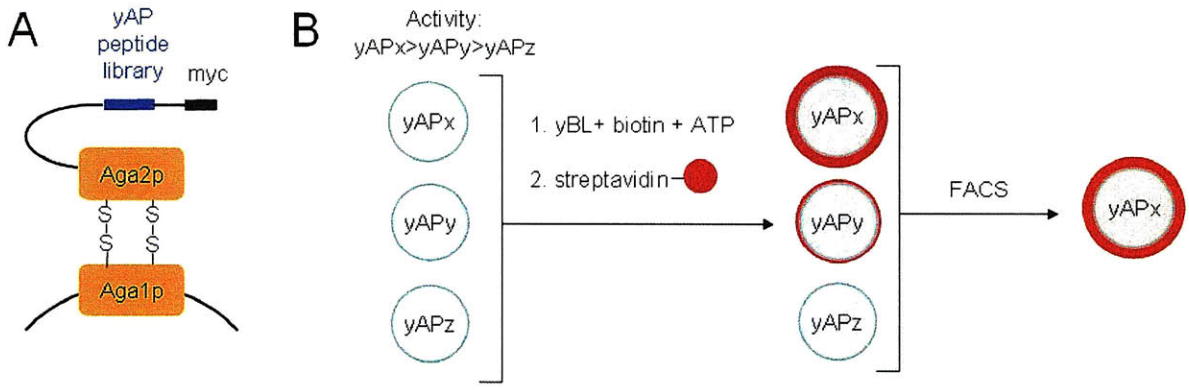


Figure 6 Schematic illustration of yeast display evolution of yAP. (A) The yAP library (blue) is displayed on the yeast surface as a fusion to Aga2p protein. A C-terminal *myc* epitope is used to quantify yAP expression level. (B) Yeast cells displaying three sample yAP sequences, with high (yAPx), moderate (yAPy), and low (yAPz) activity are shown. After reaction with yBL, the yeast cells are collectively labeled with streptavidin-phycoerythrin. The yeast pool is then sorted on the basis of both ligation extent (red probe intensity) and yAP expression level (*myc* staining intensity), to enrich the most kinetically efficient yAP peptides (such as yAPx).

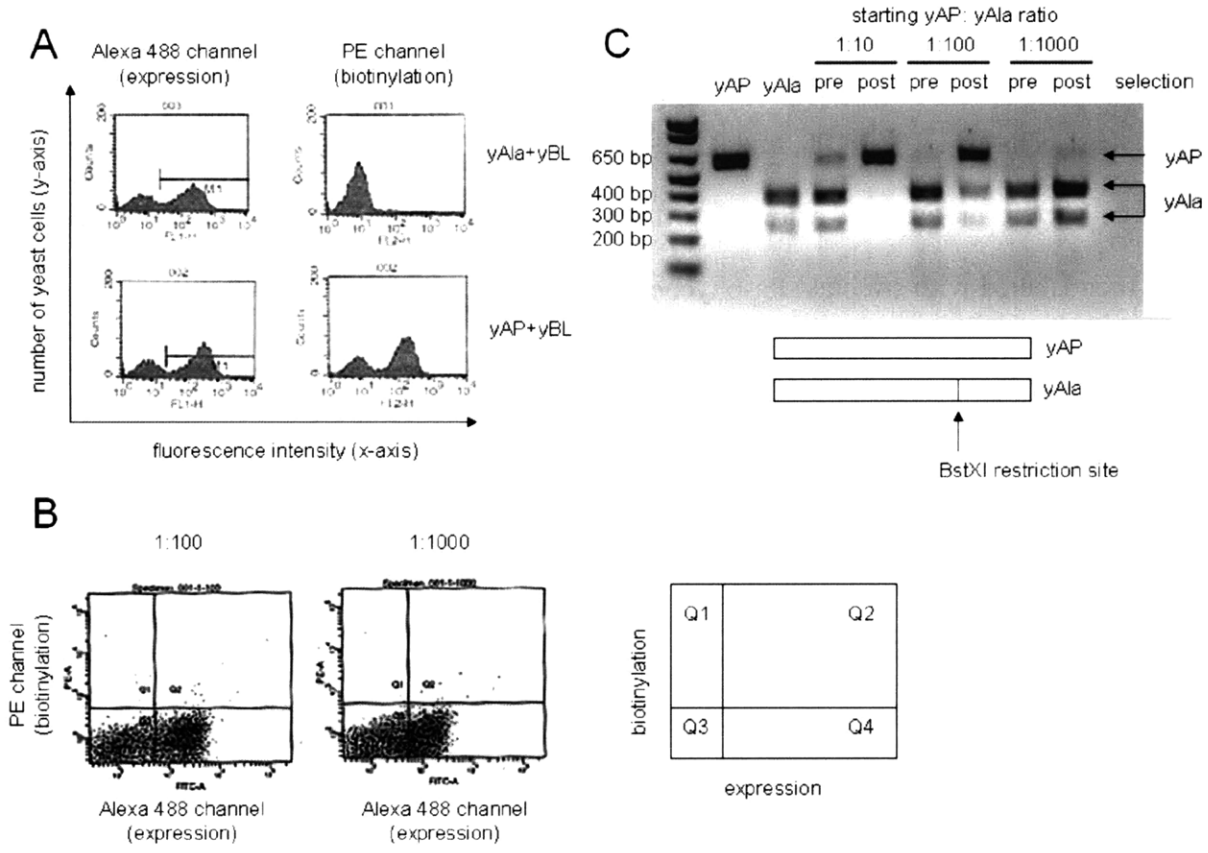


Figure 7 Model selection (1). (A) Yeast cells expressing yAP or yAla on their surface were biotinylated with yBL. After staining with streptavidin-phycoerythrin (PE) conjugate, yeast cells were analyzed by FACS. To verify similar levels of surface expression, the yeast cells were also stained with an anti-myc antibody (mouse), followed by anti-mouse antibody conjugated to Alexa 488. (B) Determination of sorting conditions for model selections. Yeast cells displaying yAP or yAla were mixed in 1:100 or 1:1000 ratios (yAP: yAla), and biotinylated with 4.7 μM yBL at room temperature for 3 h. After staining as in (A), cells were sorted using FACS. The FACS scatter plots show the distribution of yeast cells as functions of PE staining intensity (reflecting extent of biotinylation; y-axis) and *myc* staining intensity (reflecting expression level of the Aga2p yAP fusion; x-axis). A cell population in Q3 is present in all samples, and represents untransfected yeast. Yeast cells in Q2 were collected for the model selections. (C) Results of model selections. PCR reaction followed by digestion with BstXI gives the ratio of yeast populations pre- and post-selection. yAP enrichment factor was ~80-fold.

Yeast cells separately transformed with yAP and yAla plasmids were grown under identical conditions. Prior to the induction of fusion protein expression, we mixed yAP and yAla yeast cells in 1:10, 1:100, and 1: 1000 ratios (yAP: yAla). After 24 h of induction, the yeast cells were biotinylated with 4.7 μM yBL at room temperature for 3 h. The yeast cells were then labeled with streptavidin-phycoerythrin and anti-myc antibody (mouse) followed by anti-mouse secondary antibody conjugated to Alexa 488 fluorophore. The doubly labeled cells were sorted using FACS based on their fluorescence intensity in both phycoerythrin and Alexa 488 channels. The collected yeast cells were grown in selective media with glucose as a carbon source and antibiotics to prevent any contamination by bacteria. DNA was extracted from the yeast using the standard 'zymo-prep' protocol. The genes were then PCR-amplified, and the BstXI restriction enzyme was added to the PCR products. Finally, the reaction mixture was analyzed by an agarose gel. As shown in Figure 7, yAP is fully enriched over yAla after selection from the 1:10 mixture and we observed partial enrichments of yAP in the 1:100 and 1:1000 mixtures as well. The overall enrichment factor for the selection was calculated to be ~80-fold, which indicates that yeast display can be used to discriminate active substrates from completely inactive ones.

Model selection (2) We performed a different model selection to determine if yeast display could enrich highly active peptides over moderately active ones. We used the BirA system instead of yBL. AP peptide was mixed together with AP Gln (-1) →Glu mutant (AP Q→E). With this mutation (the glutamine residue is located right before the biotinyl lysine site), the biotinylation by BirA is slightly reduced compared to AP. When biotinylated with 0.3 μM BirA at room temperature for 1 h, the fluorescence intensity in phycoerythrin channel (the

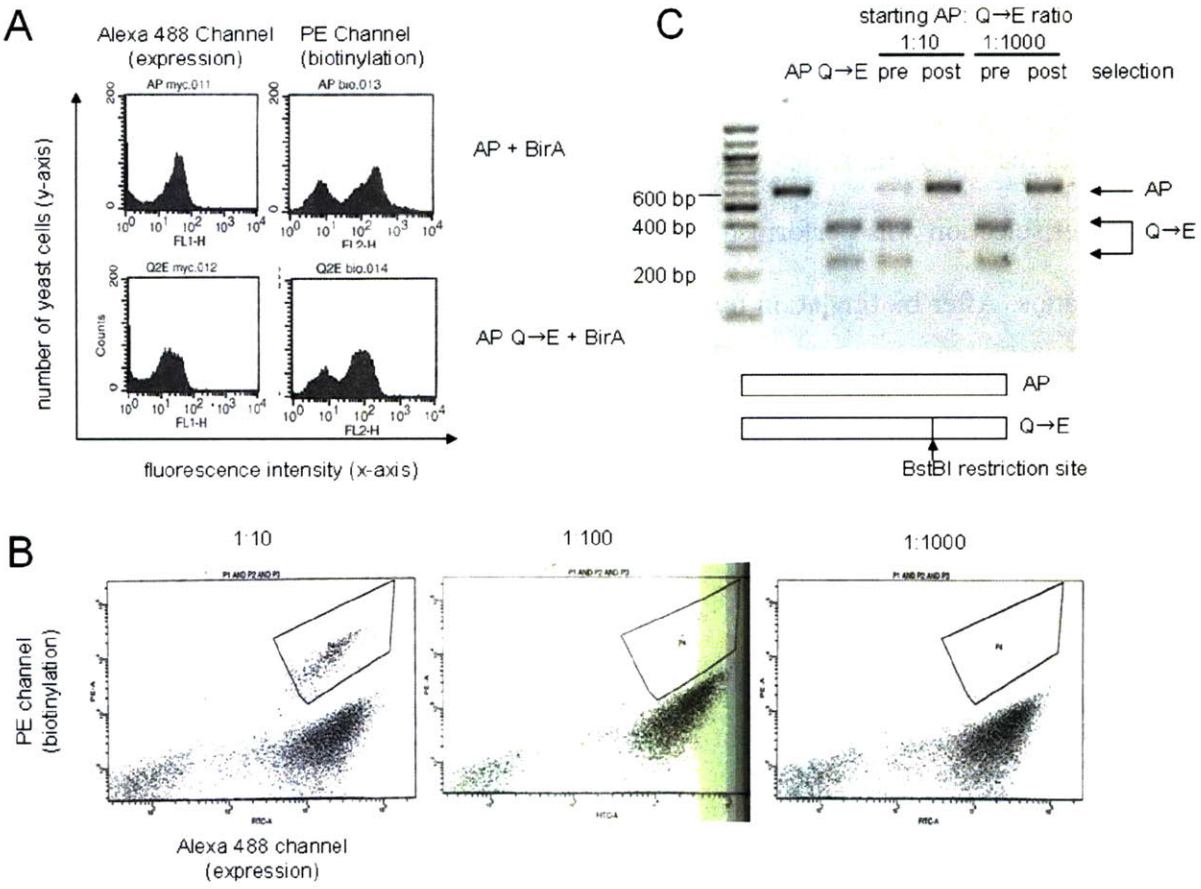


Figure 8 Model selection (2). (A) Yeast cells expressing the AP or AP Q→E (Gln right before the lysine biotinylation site in AP is mutated to Glu) on their surface were biotinylated with BirA. After staining with streptavidin-phycoerythrin (PE) conjugate, yeast cells were analyzed by FACS. To verify similar levels of surface expression, the yeast cells were stained with an anti-myc antibody (chicken), followed by anti-chicken antibody conjugated to Alexa 488. (B) Determination of sorting conditions for model selections. Yeast cells displaying AP or AP Q→E were mixed in 1:10, 1:100 and 1:1000 (AP: AP Q→E) ratios and biotinylated with 0.3 μM BirA at room temperature for 1 h. After staining as in (A), cells were sorted using FACS. The FACS scatter plots show the distribution of yeast cells as functions of PE staining intensity (reflecting extent of biotinylation; y-axis) and *myc* staining intensity (reflecting expression level of the Aga2p AP fusion; x-axis). A cell population on the lower left is present in all three samples, and represents untransfected yeast. Optimized sorting gates, used for the model selections, are shown in rectangles. (C) Results of model selections. Genes from mixtures of yeast cells pre and post-selection were PCR-amplified and digested with BstBI. AP enrichment factor over mutant peptide was ~1000-fold.

extent of biotinylation) for the yeast displaying AP was ~3 times higher than the yeast displaying AP Q→E (Figure 8A). We again created a restriction site (BstBI) on the DNA sequence of the AP Q→E so that we could distinguish AP from the AP Q→E on an agarose gel. The model selection was performed with AP and AP Q→E expressing yeast cells mixed in varying ratios. After biotinylation by BirA and analysis by PCR and restriction digestion, we observed enrichment of AP over AP Q→E by 1000-fold. Thus, yeast display provides enough dynamic range to distinguish excellent substrates from substrates that show less activity, which cannot be achieved by phage display. The present yeast display system, therefore, proved to be a promising methodology to improve the kinetics of phage-evolved yAP. Model selection (2) (AP vs AP Q→E) gave better enrichment than model selection (1) (yAP vs yAla). One explanation for the higher enrichment factor in model selection (2) could be the use of chicken anti-myc antibody (IgY). IgY antibodies generally show higher affinities against the target proteins than IgG antibodies[19]. The model selection (1) (yAP vs yAla) was performed with the mouse anti-myc antibody (IgG). Thus, the antibody might dissociate from the yeast cells much faster, which leads to the loss of cells displaying yAP since sorting gates for model selections were drawn for yeast cells that are positive for both expression and biotinylation. By conserving the expression signal, we could position more double-positive cells in the sorting gate in the model selection (2) (Figure 7B vs Figure 8B). Following selections with the AP2.2 library were performed using chicken anti-myc antibody as well. Another explanation could be the higher fluorescence intensity of cells displaying AP compared to those displaying yAP. The sorting gate for yAP yeast cells was drawn closer to untransfected or sick cells. When yeast cells are not healthy, they are particularly prone to non-specific sticking toward labeling reagents, which might decrease the enrichment factor of model selection (1). The fluorescence

intensity of AP yeast cells is higher than yAP yeast cells, and therefore the sorting gate is drawn in a way that less false positive cells are included.

Library design For our selections, we used the same AP2.2 library that Irwin had designed and constructed for his second-generation phage display selections. After incorporation into the yeast surface expression construct, we checked the library by sequencing five clones at random (Figure 9). All 14 residues were highly mutated and we did not observe any dominance of a certain sequence.

X	X	X	X	V	L	X	A	M	K	M	T	F	X	X
Y	V	T	L	I	F	R	A	T	K	R	M	F	V	H
R	P	W	N	V	P	V	A	M	K	M	P	T	I	T
Y	H	T	H	G	L	C	A	M	K	I	T	F	R	L
E	R	L	V	G	L	Q	A	I	K	M	I	*	N	R
L	I	M	R	A	L	A	A	A	K	M	T	L	K	P

Figure 9 Sequences of the AP2.2 library for yeast display. The design of AP2.2 library is shown in the first row. The lysine biotinylation site is in red. Black positions (X) are completely randomized, while blue positions are partially randomized so that ~34% of clones have the indicated amino acid sequence. Stop codon is represented as a star.

Selections with AP2.2 library The peptide library was transformed into EBY100 yeast cells. Transformation efficiency was ~10^6 cfu/µg, which is usual for yeast display experiments. In order to see the level of surface expression, the transformed yeast cells were labeled with an anti-myc antibody (chicken) followed by anti-chicken secondary antibody conjugated to Alexa 488 fluorophore. Analysis by FACS showed that about 40% of yeast cells were expressing the myc tag (Figure 10).

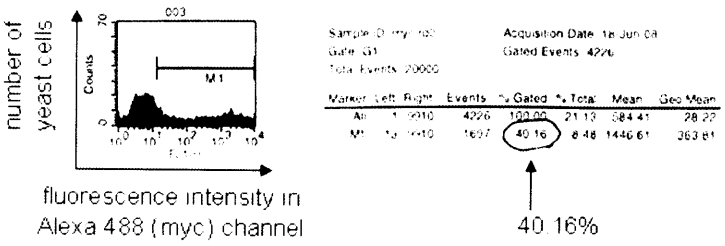


Figure 10 Myc expression of AP2.2 library after transformed to yeast cells. Among all yeast cells analyzed, 40.16% (sub-population in M1 region; arrowed) of cells were expressing the myc tag, which can be detected by anti-myc antibody (chicken) and Alexa 488-conjugated anti-chicken antibody. The M1 region was manually drawn based on the myc expression of untransfected yeast cells.

In the first round of selection, yeast cells were treated with 5 μM yBL at room temperature for 3 h. 0.3% of the yeast cells showed positive signal for both phycoerythrin and Alexa 488, and these cells were collected and amplified for the next round of selection. For the second round of selection, cells were treated with 5 μM yBL for 1.5 h. For the third round, we used 5 μM yBL for 1.5 h.

We tested biotinylation of recovered cells from each round by yBL and BirA. Compared with the cells before selection, the biotinylation extent of yeast cells by yBL was increased after each round of selection (Figure 11), but BirA biotinylation remained as low as before (data not shown).

However, compared to the phage-evolved yAP sequence, we did not observe higher activity in our yeast-evolved pools. This may be a consequence of the conditions used for the yeast analysis, or it may reflect the limitations of our selection or the AP2.2 library. Further characterization was not performed.

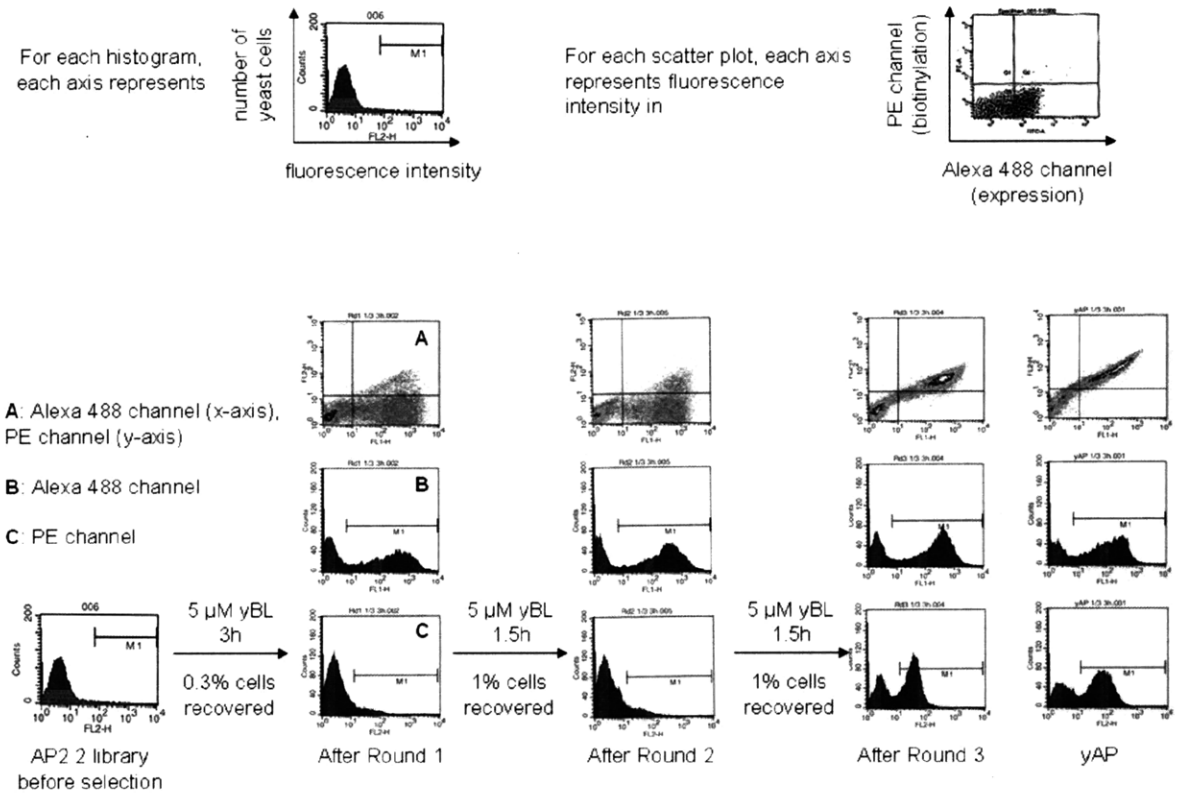


Figure 11 Three rounds of yeast display selection with the AP2.2 library. Selection conditions are given above each arrow. Following each round of selection, collected yeast pools were amplified and uniformly biotinylated with 1.67 μM yBL at room temperature for 3 h. Row A shows the FACS scatter plot for each round (myc intensity on x-axis, biotinylation intensity on y-axis). Row B shows histograms of Alexa 488 channel showing the extent of peptide expression. Row C shows histograms of phycoerythrin channel showing the extent of biotinylation by yeast biotin ligase. The last column shows similar characterization of Irwin Chen's phage-evolved yAP, for comparison.

Conclusions

Here we have developed a yeast display system to engineer the peptide substrates of yeast biotin ligase (yBL). At the outset of the project, yeast display had not previously been used to evolve novel enzyme substrates. By performing two model selections, we demonstrated that the current yeast display method offers a wider dynamic range than the

previously used phage display format. We assume that this is mainly due to the availability of more copies of displayed peptides and the dynamic range of FACS. These features of yeast display enabled discrimination of kinetically superior substrates from inferior ones, with an enrichment factor of ~1000.

Following three rounds of yeast display selection of the peptide library with yBL, the enriched library had activity comparable to Irwin's phage-evolved yAP, although analysis under a wider range of conditions could reveal differences. There was, however, clear enrichment of activity towards yBL, indicating some degree of success in the selection approach.

Also, designing a new library would be another venue for the development of better yBL substrates. Since our knowledge of yBL is still insufficient and the crystal structure of yBL or the yeast biotinyl domain (yPC-104)[20] is not available, it would be harder to design a new library by methods other than using highly conserved sequences from different biotinyl domains. The Met-Lys-Met motif is conserved in almost every biotinyl domains in other organisms, but neither AP nor yAP contains the motif. One possibility would be to utilize the lipoyl domain since biotinyl and lipoyl domains have structural homology as mentioned in Chapter 1. Designing a library with shorter length (~12 amino acids as LAP2 peptide) and different positions of the biotinyl lysine would be helpful to find a better yAP as well.

Preparation of yBL should be improved as well. Purification by anti-BirA antibody cannot completely remove the endogenous BirA coming from bacterial expression, and the residual BirA can drive the selection to enrich substrates used by BirA. It is especially disrupting since the design of library has significant similarity to AP. Another problem is the proteolysis of yBL when expressed in *E. coli*[21]. To remove any BirA contaminants and prevent

the proteolysis, expression of yBL in yeast has been investigated, although we have faced several technical difficulties. Recently, a biotin ligase from *Candida albicans* was expressed in the yeast vector pVT100u[22]. They reported a high yield of protein expression (up to 13 mg/mL) and more importantly did not observe the proteolytic ligase product.

The yAP should also allow for the production of site-specifically biotinylated proteins in live yeast, a tool that should be useful to those who study yeast cell biology or use yeast for heterologous expression of proteins. However, due to technical problems, we could not prove this yet.

Our peptide selection scheme is general and should enable the discovery of more peptide substrates for biotin ligases of other species, conferring the benefits of site-specific biotinylation to proteins expressed in those cells and organisms. The discovery of peptide substrates for biotin ligases that do not recognize yAP and AP would also allow the multiplexing of all applications of site-specific biotinylation, including multicolor imaging. Moreover, the discovery of new orthogonal acceptor peptide/biotin ligase pairs will be necessary to eventually realize the goal of simultaneous labeling of different proteins with different small molecule probes. As mentioned earlier, development of biotin 2/ streptavidin 2 would further expand the AP and yAP methodology to a true orthogonal labeling of proteins.

Experimental

Cloning of yAP-CFP. An insert containing the N-terminal yAP, linker, and the CFP was constructed through two successive PCR reactions. For the first PCR, the CFP-AP gene was used as the template, and the primers 4.2NyAP. F2 (5'- GCT CAG GCT TTC AAG ATG ACG TTT GAT CCG GGC GGG ATG GTG AGC AAG GGC GA; incorporates part of the yAP

sequence TTNWVAQAFKMTFDP) and NyAPCHis. Rv (5'- TAA *GAA TTC* TCC TTG TAC AGC TCG TCC AT; incorporates an *EcoRI* site) were used for the amplification. For the second PCR, the product of the first PCR reaction was used as the template and the primers 4.2NyAP. F1 (5'- TAA *GGA TCC* CAC GAC TAA TTG GGT TGC TCA GGC TTT CAA GAT GAC GTT TG; incorporates part of the yAP sequence and a *BamHI* site) and NyAPCHis. Rv were used for the amplification. The insert was digested with *BamHI* and *EcoRI* and ligated in-frame to similarly digested pET-21b vector (Novagen), which introduces a C-terminal His_6 tag.

Cloning of CFP-yAP. An insert containing the CFP, linker, and the C-terminal yAP was constructed through two successive PCR reactions. For the first PCR, the CFP-AP gene was used as the template, and the primers NHisCyAP. Fw (5'- TAA *GGA TCC* CCA CCA CCA CCA CCA CCA CGG GGG GAT GGT GAG CAA GGG C; incorporates a His_6 tag and a *BamHI* site) and 4.2CyAP. R1 (5'- CGT CAT CTT GAA AGC CTG AGC AAC CCA ATT AGT CGT GCC GCC GGA GGA CTC; incorporates part of the yAP sequence) were used for the amplification. For the second PCR, the product of the first PCR reaction was used as the template and the primers NHisCyAP. Fw and 4.2CyAP. R2 (5'- TAA *GAA TTC* CTT TAC GGA TCA AAC GTC ATC TTG AAA GCC TGA GCA ACC C; incorporates part of the yAP sequence and a *BamHI* site) and NyAPCHis. Rv were used for the amplification. The insert was digested with *BamHI* and *EcoRI* and ligated in-frame to similarly digested pET-21b vector.

Cloning of yAP-HP1. For the PCR reaction, the HP1 with AP fused to the C-terminus (HP1AP) was used for the template, and the primers HP1N4.2. F (5'- TTT TTT *GCT AGC*

ACG ACT AAT TGG GTT GCT CAG GCT TTC AAG ATG ACG TTT GAT CCG GAG GAG GAG TAC GCC GTG; incorporates the yAP sequence and an *NheI* site) and C_HindIII. R (5'- TTT TTT *AAG CTT* GCC GCC ATC CTT GCG GC; incorporates a *HindIII* site) were used for the amplification. The insert was digested with *NheI* and *HindIII* and ligated in-frame to similarly digested pET-21b vector, which introduces a C-terminal His_6 tag.

Cloning of HP1-yAP. An insert containing the HP1, linker, and the C-terminal yAP was constructed through two successive PCR reactions. For the first PCR, the HP1AP was used for the template, and the primers N_Bam_His_HP1. F (5'- TTT CGC *GGA TCC* GCA CCA CCA CCA CCA CCA CGA GGA GGA GTA CGC CGT G; incorporates a His_6 tag and a *BamHI* site) and HP1C4.2. R1 (5'- TGA AAG CCT GAG CAA CCC AAT TAG TCG TGC CGC CAT CCT TGC GGC TCG CCT C; incorporates part of the yAP sequence) were used for the amplification. For the second PCR, the product of the first PCR reaction was used as the template and the primers N_Bam_His_HP1. F and HP1C4.2. R2 (5'- TTT CCG *GAA TTC* ATT ACG GAT CAA ACG TCA TCT TGA AAG CCT GAG CAA CCC A; incorporates part of the yAP sequence and an *EcoRI* site) were used for the amplification. The insert was digested with *BamHI* and *EcoRI* and ligated in-frame to similarly digested pET-21b vector.

Western blot assays for yAP biotinylation. Proteins were biotinylated *in vitro* as follows: 20 μM yAP-CFP, CFP-yAP, yAP-HP1, or HP1-yAP, 5 μM yBL, 4 mM ATP, and 1 mM Biotin in Assay buffer (40 mM Tris, 50 mM KCl, 11 mM $MgCl_2$ pH 8.0) for overnight at 30 °C. The samples were analyzed by 12 or 16% SDS-PAGE and blotting with either Streptavidin-HRP (1:1000 dilution, Pierce) or anti-Histidine-tagged protein antibody (1:1000 dilution,

Calbiochem) followed by HRP-conjugated anti-mouse antibody (1:1000 dilution, Bio-Rad). The HRP signal was visualized with Supersignal West Pico and Supersignal West Femto (Pierce).

Cloning of pCTCon2-Aga2p-AP and pCTCon2-Aga2p-yAP for yeast cell surface labeling (by Irwin Chen). The plasmid expressing Aga2p-AP was constructed by yeast homologous recombination. The insert was made in two steps. First, the oligonucleotides APRev.R (5'- AAA TAA GCT TTT GTT CGG ATC CCT CGT GCC ACT CGA TCT TCT GGG CCT CGA AGA TGT CGT TCA GGC CTC CGC TAG CCG ACC CTC C; incorporates the AP) and Con2For.F (5'- CTA GTG GTG GAG GAG GCT CTG GTG GAG GCG GTA GCG GAG GCG GAG GGT CGG CTA GCG GA; incorporates ~40 bp of overlap with pCTCon2 5′ of the AP sequence) were annealed and the 5' overhangs were made double-stranded with Klenow fragment. The annealed product was then amplified with primers Con2For.F and Con2Rev.R (5'- TAT CAG ATC TCG AGC TAT TAC AAG TCC TCT TCA GAA ATA AGC TTT TGT TCG GAT CC; incorporates ~50 bp of overlap with pCTCon2 3′ of the AP). The insert and pCTcon2 vector (a kind gift from K.D. Wittrup, triply digested with NheI/SalI/BamHI) were transformed together into *S. cerevisiae* EBY100 (Invitrogen) using the Frozen-EZ Yeast Transformation II^{TM} kit (Zymo Research). DNA was isolated from yeast colonies using the ZymoprepTM Yeast Plasmid Minipreparation Kit (Zymo Research) and amplified in *E. coli.*
The insert for Aga2p-yAP was constructed by annealing the oligonucleotides NheyAP.F (5'- GGG TCG GCT AGC GGA GGC ACG ACT AAT TGG GTT GCT CAG GCT TTC; incorporates an NheI site and part of the yAP) and BamyAP.R (5'- TTT TTT GGA TCC CGG ATC AAA CGT CAT CTT GAA AGC CTG AGC AAC CCA ATT; incorporates a BamHI site

and the rest of the yAP), and the 5' overhangs were made double-stranded with Klenow fragment. The insert was digested with NheI and BamHI and ligated in-frame to similarly digested pCTCon2 vector. The Aga2p-yAla insert was similarly constructed using BamyAla.R (5'- TTT TTT GGA TCC CGG ATC AAA CGT CAT CGC GAA AGC CTG AGC AAC CCA ATT).

Yeast cell surface peptide expression, labeling, and FACScanning. Yeast strain EBY100 was used for all experiments[23]. Transformation was performed using Frozen-EZ Yeast Transformation II™ kit (Zymo Research) according to the manufacturer's instructions. After transformation, cells were grown in synthetic minimal defined media (SD-W-U; 0.67% yeast nitrogen base, 2% glucose, 0.01% of adenine, arginine, cysteine, leucine, lysine, threonine and 0.005% of aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, valine) at 30 °C with shaking for 24 h and induced in SG-W-U media (same as SD-W-U except galactose instead of glucose) at room temperature for 24 h with shaking. Induced cells were harvested, washed with PBSB (PBS + 1% BSA), and 2×10^6 cells in 100 μL PBSB were taken out for labeling experiments. For BirA reaction, cells were incubated for 3 h at room temperature with 50 μM biotin, 0.3 μM BirA, 4 mM ATP, and Assay buffer. For yBL reaction, cells were incubated for 3 h at room temperature with 10 μM Biotin, 4.7 μM yBL, 1 mM ATP, and 1.4 mM $Mg(OAc)_2$. The biotinylated cells were then washed once with PBSB, and incubated with Streptavidin-Phycoerythrin (1:100 dilution, Jackson ImmunoResearch) for 1 h at 4 °C. For detection of myc-tag as an expression marker, anti-myc antibody (1:100 dilution, Calbiochem) was added to cells separately at room temperature. After 1 h, cells were washed once with PBSB and incubated with anti-mouse antibody-Phycoerythrin (1:100

dilution, Invitrogen) for 1 h at 4 °C. Cells were rinsed once with PBSB and suspended in 600 μL of PBSB for flow cytometry (FACScan, BD Biosciences).

Cloning of pCTCon2-Aga2p-yAla with BstXI restriction site and pCTCon2-Aga2p-AP Q→E with BstBI for model selection. The plasmid expressing Aga2p-yAla was generated from the pCTCon2-Aga2p-yAP by following the QuikChange protocol (Stratagene) with the oligonucleotides NewyAla(BstXI).F (5′- ACG ACC AAC TGG GTG GCT CAG GCT TTC GCG ATG ACG TTT GAT CCG; incorporates the yAla) and its reverse complement NewyAla(BstXI).R (5′- CGG ATC AAA CGT CAT CGC GAA AGC CTG AGC CAC CCA GTT GGT CGT). The plasmid expressing Aga2p-AP Q→E was similarly generated from the pCTCon2-Aga2p-AP by QuikChange using the oligonucleotides APQ2E(BstBI).F (5'- CGA CAT CTT CGA AGC CGA GAA GAT CGA GTG GC; incorporates 32bp of AP Q→E) and its reverse complement APQ2E(BstBI).R (5'-GCC ACT CGA TCT TCT CGG CTT CGA AGA TGT CG).

Model selection (1). The yeast plasmids for expression of yAP and yAla were transformed to EBY100 cells using the Frozen-EZ Yeast Transformation II^{TM} kit (Zymo Research) according to the manufacturer's instructions. After transformation, cells were grown in synthetic minimal defined media (SD-W-U; 0.67% yeast nitrogen base, 2% glucose, 0.01% of adenine, arginine, cysteine, leucine, lysine, threonine and 0.005% of aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, valine) at 30 °C with shaking for 24 h. Yeast cells expressing yAP and yAla were mixed in 1: 10, 1:100, and 1: 1000 ratios (yAP: yAla) before induced in SG-W-U media (same as SD-W-U except galactose instead of

glucose) for 24 h at room temperature with shaking. Induced cells were harvested, washed with PBSB (PBS + 1% BSA), and 2×10^6 cells in 100 μL PBSB were taken out for labeling experiments. For yBL reaction, cells were incubated for 3 h at room temperature with 10 μM Biotin, 4.7 μM yBL, 1 mM ATP, and 1.4 mM $Mg(OAc)_2$. The biotinylated cells were then washed once with PBSB. For detection of myc-tag as an expression marker, anti-myc antibody (1:100 dilution, Calbiochem) was also added to cells at room temperature. After 1 h, cells were washed once with PBSB and incubated with anti-mouse antibody-Alexa488 (1:100 dilution, Invitrogen) and Streptavidin-Phycoerythrin (1:100 dilution, Jackson ImmunoResearch) for 1 h at 4 °C. Cells were rinsed once with PBSB and suspended in 600 μL of PBSB for flow cytometry (FACScan, BD Biosciences). For the 1:10, 1:100, and 1:1000 mixtures, 0.7%, 0.1%, and 0.2% of the population were collected in eppendorf tubes containing 1 mL of SD-W-U media, respectively.

Collected cells were then grown in 5 mL of SD-W-U with penicillin-streptomycin (1: 100 dilutions) for 24 h at 30 °C. DNA was isolated from yeast cells using the Zymoprep™ Yeast Plasmid Minipreparation Kit (Zymo Research) and PCR amplified using the oligonucleotides flanking.F (5'- CTG TTA TTG CTT CAG TTT TAG CAC) and flanking.R (5'- GTG TAA AGT TGG TAA CGG AAC G). The PCR products were then digested with BstXI for 1 h at 37 °C, and the reactions were analyzed in an agarose gel.

Model selection (2). The yeast plasmids for expression of AP and AP Q→E were transformed to EBY100 cells using the Frozen-EZ Yeast Transformation II™ kit (Zymo Research) according to the manufacturer's instructions. Yeast cells were grown and mixed in 1:10, 1:100 and 1:1000 ratios as model selection (1). After followed the same protocol as above, yeast

cells were biotinylated for 1 h at 30 °C using 0.3 μM of BirA, 4 mM ATP, 50 μM biotin, and Assay buffer. The biotinylated cells were then washed once with PBSB. For detection of myc-tag as an expression marker, anti-myc antibody (1:100 dilution, Invitrogen) was also added to cells at room temperature. After 1 h, cells were washed once with PBSB and incubated with anti-chicken antibody-Alexa488 (1:100 dilution, Invitrogen) and Streptavidin-Phycoerythrin (1:100 dilution, Jackson ImmunoResearch) for 1 h at 4 °C. Cells were rinsed once with PBSB and suspended in 600 μL of PBSB for flow cytometry (FACScan, BD Biosciences). For the 1:10, 1:100, and 1:1000 mixtures, 4.2%, 0.7%, and 0.1% of the population were collected in eppendorf tubes containing 1 mL of SD-W-U media, respectively. Collected cells were then grown as above. DNA was isolated from yeast cells and PCR amplified using the oligonucleotides flanking.F (5'- CTG TTA TTG CTT CAG TTT TAG CAC) and flanking.R (5'- GTG TAA AGT TGG TAA CGG AAC G). The PCR products were then digested with BstBI for 1 h at 37 °C, and the reactions were analyzed in an agarose gel.

Cloning and production of a partially-randomized 15-mer AP2.2 yeast library. The DNA encoding the yeast library was constructed by yeast homologous recombination. The insert was made in two steps. First, the oligonucleotides AP2.2Lib.R (5'- CAT GTT T*CG GCC G*AM NNM NNA AAC GTC ATC TTC ATA GCM NNA AGA ACM NNM NNM NNM NNA GAG TGA GAA TAG AAA GGT ACC CGG G where M=A,C and N= A,G,C,T. A represents 70% A and 10% each of the other bases, G represents 70% G and 10% each of the other bases, and so forth) and newAnneal.F (5'- GGG TCG GCT AGC GGA CCC CGG GTA CCT TTC TAT TCT CAC TCT) were annealed and the 5' overhangs were made double-stranded with Klenow fragment. The annealed product was then amplified with primers fowardsmall.F (5'- CTA

GTG GTG GAG GAG GCT C) and newreverse.R (5'- TCG GCC GAA ACA TGC GGA TCC GAA CAA AAG CTT ATT TCT GAA GAG GAC TTG TAA TAG CTC GAG ATC TGA TA). The insert and pCTcon2 vector (triply digested with NheI/SalI/BamHI) were transformed together into electrocompetent *S. cerevisiae* EBY100 using electroporation. Yeast cells expressing the library were grown in SD-W-U and passaged twice before use to reduce the number of untransformed cells.

Selections on AP2.2 library. Three rounds of selection with yeast biotin ligase (yBL) were performed as follows. Yeast cells (2×10^{10} cells to eliminate dead cells) were biotinylated with 5 μM yBL, 1 mM $Mg(OAc)_2$, 70 μM biotin, 0.7 mM ATP for 3 h (round 1) or 1.5 h (rounds 2 and 3) at room temperature. After washing and labeling with streptavidin-phycoerythrin and anti-myc antibody followed by anti-chicken conjugated to Alexa 488, yeast cells (2×10^7 cells) were sorted using FACS. For round 1, 0.3% of the population was collected, and 1% of the population was collected for rounds 2 and 3. Collected cells in SD-W-U media were grown with penicillin-streptomycin (1: 100 dilutions) for 24 h at 30 °C.

Comparison of three rounds of selection and yAP was performed with 2×10^6 cells using 1.67 μM yBL, 1 mM $Mg(OAc)_2$, 70 μM biotin, 0.7 mM ATP for 3 h.

References

1 Green, N.M., Avidin. *Adv. Protein Chem.* 29, 85-133 (1975).

2 Schatz, P.J., Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. *Biotechnology* 11, 1138-1143 (1993).

[3] Beckett, D., Kovaleva, E., & Schatz, P.J., A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 8, 921-929 (1999).

[4] Athavankar, S. & Peterson, B.R., Control of Gene Expression with Small Molecules: Biotin-Mediated Acylation of Targeted Lysine Residues in Recombinant Yeast. *Chem Biol* 10, 1245-1253 (2003).

[5] de Boer, E. *et al.*, Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. . *Proc Natl Acad Sci USA* 100, 7480-7485 (2003).

[6] 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* 129, 6619-6625 (2007).

[7] Gai, S.A. & Wittrup, K.D., Yeast surface display for protein engineering and characterization. . *Curr. Opin. Struct. Biol.* 17, 467-473 (2007).

[8] Boder, E.T. & Wittrup, K.D., Yeast surface display for screening combinatorial polypeptide libraries. . *Nat. Biotechnol* 15, 553-537 (1997).

[9] Pepper, L.R., Cho, Y.K., Boder, E.T., & Shusta, E.V., A decade of yeast surface display technology: Where are we now? *Comb. Chem. High Throughput Screen.* 11, 127-134 (2008).

[10] Chao, G., Cochran, J.R., & Wittrup, K.D., Fine Epitope Mapping of anti-Epidermal Growth Factor Receptor Antibodies Through Random Mutagenesis and Yeast Surface Display. *J. Mol. Biol.* 342, 539-550 (2004).

[11] Peelle, B.R., Krauland, E.M., Wittrup, K.D., & Belcher, A.M., Probing the interface between biomolecules and inorganic materials using yeast surface display and genetic

engineering. *Acta Biomater* 1, 145-154 (2005).

[12] Furukawa, H., Tanino, T., Fukuda, H., & Kondo, A., Development of Novel Yeast Cell Surface Display System for Homo-oligomeric Protein by Coexpression of Native and Anchored Subunits. *Biotechnol Prog* 22, 994-997 (2006).

[13] Colby, D.W. *et al.*, Engineering antibody affinity by yeast surface display. *Methods Enzymol.* 388, 348-358 (2004).

[14] Chao, G. *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* 1, 755-768 (2006).

[15] Bowley, D.R., Labrijn, A.F., Zwick, M.B., & Burton, D.R., Antigen selection from an HIV-1 immune antibody library displayed on yeast yields many novel antibodies compared to selection from the same library displayed on phage. *Protein Eng Des Sel.* 20, 81-90 (2007).

[16] Paschke, M., Phage display systems and their applications. *Appl. Microbiol. Biotechnol.* 70, 2-11 (2006).

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

[18] 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* 131, 16430-16438 (2009).

[19] Ikemori, Y., Peralta, R.C., Kuroki, M., Yokoyama, H., & Kodama, Y., Research note: avidity of chicken yolk antibodies to enterotoxigenic Escherichia coli fimbriae. *Poult Sci* 72, 2361-2365 (1993).

[20] Val, D.L., Chapman-Smith, A., Walker, M.E., Cronan, J.E., & Wallace, J.C.,

Polymorphism of the yeast pyruvate carboxylase 2 gene and protein: effects on protein biotinylation. *Biochem. J.* 312, 817-825 (1995).

[21] Polyak, S.W., Chapman-Smith, A., Brautigan, P.J., & Wallace, J.C., Biotin Protein Ligase from Saccharomyces cerevisiae. *J. Biol. Chem.* 274, 32847-32854 (1999).

[22] Pendini, N.R. *et al.*, Biotin protein ligase from Candida albicans: Expression, purification and development of a novel assay. *Arch Biochem Biophys* 479, 163-169 (2008).

[23] Boder, E.T. & Wittrup, K.D., Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol.* 328, 430-444 (2000).

Chapter 3: Development of a Biotin Ligase Evolution Platform Based on *In Vitro* Compartmentalization

The chapter describes the development of a novel bead-less *in vitro* compartmentalization (IVC) selection scheme for BirA evolution. My specific contribution is as follows. Based on preliminary results of Bernard Kelly and Andrew Griffiths (Medical Research Council Laboratory of Molecular Biology, UK), Irwin Chen and Dr. Hemanta Baruah developed a model selection with BirA and folA (*E. coli* dihydrofolate reductase = a negative control lacking biotinylation activity) using IVC selection with microbeads displaying AP fusion proteins. I joined the project later, and reproduced their model selections successfully. However, all of us noticed problems related to the use of beads such as non-specific sticking and bead degradation over time. Although we tried to solve those problems, the inconsistency did not improve. Thus we decided to design a new selection scheme without using beads.

Dr. Baruah and I equally contributed to the design and development of bead-less IVC. I developed a protocol for peptide-oligonucleotide conjugation and purification using HPLC. I also characterized the conjugate using MALDI-TOF. The conjugate used in the work was then generated in large scale by Dr. Baruah and Kathy Xie, an undergraduate student in the lab. I independently and successfully performed the model selection with BirA and folA using bead-less IVC. I further performed a model selection with wild type and G115S BirA, and found a lack of enrichment.

Introduction

Many natural enzymes are extremely fast catalysts; they catalyze reactions in milliseconds which otherwise will take millions of years without catalysts[1]. Some enzymes are also highly specific; they only catalyze one type of reaction with the native substrates. Enzymatic reactions are regio-, stereo- and chemoselective[2]. The reaction specificity is achieved by three-dimensional structure of the enzyme. For example, negatively charged enzyme surface attracts binding of the positively charged native protein substrate. Hydrophilicity or hydrophobicity of the side chains in the substrate binding site of enzymes is also important to distinguish native substrates from other biomolecules.

At present, predicting the activity of an enzyme solely based on the structure is a highly difficult task. Furthermore, engineering enzymes to have activities on substrates whose structure are very different from native ones is also challenging[3]. As it will be discussed in Chapter 5, we were able to engineer LplA to take the aryl azide probe by semi-rational design. However, similar rational approaches to modify BirA to accept unnatural biotin analogs were unsuccessful in our early studies. Here directed evolution[4] was performed as an alternative approach for BirA engineering, which harnesses the power of natural selection. Directed *in vitro* evolution consists of three vital components[5]: a method to generate genetic diversity; a way to link genotype and phenotype; and a selection method for desired activity.

Library construction

Creating molecular diversity of the target protein is the first step in any directed evolution experiment. Recent advances in chemical synthesis and molecular biology make it practical to create mutant proteins through targeted and random mutagenesis of the gene

encoding the protein[6]. Although there is theoretically no limit to the number of codons that can be randomized simultaneously, current synthetic protocols only allow 10^{15} different members in a gene library[5], primarily due to the limited amount of DNA that can be handled. Thus, we are sampling only a tiny fraction of sequence space. Saturated mutagenesis of specific residues of the protein has been used to create semi-rational protein mutant libraries. A variety of techniques were also developed to introduce point mutations at random sites of target proteins[7]. The most widely used method is error-prone PCR, which produces random mutations during PCR by reducing DNA polymerase fidelity. This kind of techniques does not require any knowledge on structures or molecular mechanisms of target proteins. Another powerful strategy is gene recombination such as DNA shuffling[8]. By reassembling random fragments of a gene or a pool of related genes (and also mutated corresponding genes) in a self-priming PCR reaction, independent and beneficial mutations can occur to a single gene. Variations and improvement on these basic methods have facilitated creation processes of novel hybrid proteins[9,10,11].

In vitro compartmentalization: linking genotype to phenotype

As reviewed in Chapter 1, IVC is one of the best directed evolution methodologies for enzymes[5]. Compared to cell-based selections, IVC expands the scope of reactions that can be selected (because it is not limited by the survival of the cell) and allows selection on larger library sizes (because it is not limited by cellular transformation efficiencies). Compared to other *in vitro* evolution methods, the great advantage of IVC is that a selection based on enzymatic turnover number is possible because a large number of substrates can be included in one compartment.

In IVC, individual genes are dispersed into separate water droplets suspended in oil (0 or 1 DNA molecule per droplet)[12,13]. Within each droplet, the gene is transcribed and translated into protein by bacterial cell extracts. The result is that genes are co-compartmentalized with the protein products they encode, like an artificial cell. Each microdroplet is ~2 μm in diameter with a volume of ~5 femtoliters, and thus 10^{10} compartments can be generated with a 50 μL reaction, enabling IVC to provide a large library of proteins in a single microtube[14]. In addition, double emulsions (water/oil/water) can be generated when single water/oil emulsions are re-emulsified, and these droplets can be sorted using FACS (fluorescence-activated cell sorting), providing a direct selection without breaking emulsions[15]. Recently, microfluidic devices such as NanoReactors™ were developed to create thousands of droplets per second[16,17], and this facilitates high-throughput screening or selection.

Direct selection for catalytic activity

In vitro selection for catalytic activity is generally based on affinity-purification of the desired product. Although successful examples are reported, indirect selections using transition state analogs or mechanism-based inhibitors are generally limited in scope and in the catalytic efficiency of the resulting enzymes[5]. In direct selection, the selection pressure is for substrate recognition, product formation (and reaction rate acceleration), and turnover. In IVC, the substrate is detached from the enzyme in each compartment, and therefore the substrate should be associated with the enzyme to be transformed to the product. Once the reaction occurs, the product is released due to its lower affinity to enzyme than that of the substrate. Selection for turnover can be devised if a large number of substrate molecules could be attached to each gene (via microbeads) or a lot of soluble substrates are available in the

compartment.

IVC has been successfully employed to evolve faster forms of restriction endonucleases[18], DNA polymerases[19,20], DNA methyltransferases[12,21,22], phosphotriesterase[23], β-galactosidase from Ebg (a protein of unknown function)[15], thiolactonase[24], and ribozymes that catalyze multiple-turnover Diels-Alder cycloadditions[25]. In this chapter, our attempts to evolve BirA mutants with a novel activity using IVC will be discussed. Although biotinylation kinetics of BirA labeling is already sufficient for tagging of many proteins of interest, faster BirA would be more useful especially when the BirA/AP pair is used for studying transient protein-protein interactions[26]. This method employs AP-fused protein I and BirA-fused protein II, where proteins I and II might interact with each other. If the timescale of their interaction is shorter than that of BirA catalytic cycle, it would result false-negatives (no biotinylation signal) even though there is interaction between two proteins. The first goal of the project was to evolve a BirA mutant that has faster biotinylation kinetics than wild type BirA. If this is possible, this method can also be applied for evolution of BirA mutants that uses new interesting probes, such as aryl azide (Chapter 5).

Results and Discussion

Design and construction of a BirA mutant library

Prior to implementing a selection to evolve a faster biotin ligase enzyme, we designed and constructed a library of focused BirA mutants. Based on the crystal structure of BirA in complex with biotin[27], we selected eight positions (T90, Q112, G115, W123, G186, G204, A205, and G206; Figure 1) lining the binding pocket which is close to a bound biotin. Randomization of these eight residues would give a library with theoretical amino acid

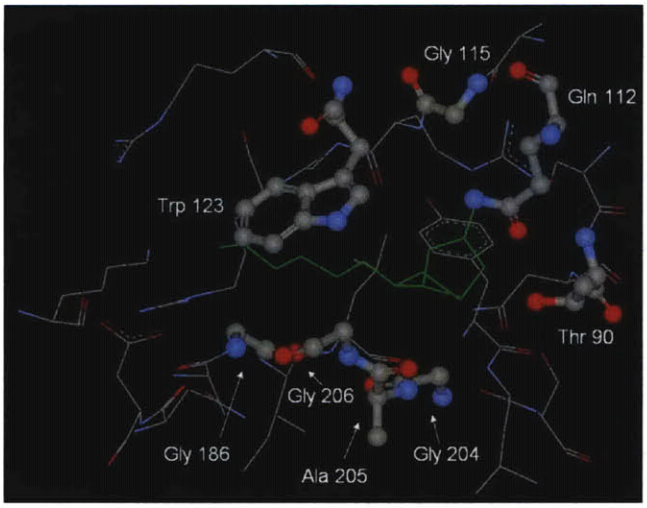


Figure 1 Positions of randomized residues of BirA. Biotin was labeled in green. Residues close to biotin and chosen for randomization were represented with ball and stick models.

	T90	Q112	G115	W123	G186	G204	A205	G206	Additional mutation
L1	Ile	Val	Arg	Phe	Trp	Asn	Met	Ser	881G(A)
L2	Stop	Leu	Phe	Ser	Leu	Asp	Ser	Leu	57C(T), 312T(C)
L3	Arg	Gly	Arg	His	Ile	Lys	Ile	Trp	312T(C), 81G(A)
L4	Pro	Ile	Arg	Asp	Tyr	Pro	Leu	Cys	312T(C)
L5	Stop	Ser	Gly	Glu	His	Thr	Trp	Thr	
L6	Thr	Ala	Arg	Ile	Trp	Stop	Gly	Leu	312T(C)
L7	Ala	Stop	Pro	Trp	Arg	Met	Lys	Arg	
L8	Met	Lys	Stop	Asn	Val	Ser	Gly	Ile	
L9	Val	Ile	Asn	Asn	Met	Lys	Ser	Ser	47G(A)
L10	Asn	Ala	Arg	Gln	Gly	Ile	Tyr	Ser	491C(T)
L11	Lys	Thr	Ala	Thr	Met	Arg	Ile	Gly	
L12	Thr	Val	Ser	Arg	Gly	Tyr	Val	Arg	312T(C),927A(G)
L13	Ala	Phe	Tyr	Gly	Trp	Ser	Phe	Gly	
L14	Ile	Phe	Pro	Met	Thr	Trp	Phe	Arg	312A(C)
L15	Ile	Pro	Gly	Ser	Cys	Cys	Thr	Thr	530A(T)

Table 1 Sequences of selected clones from the BirA mutant library. Additional mutations were generated probably because of the low fidelity of DNA polymerase.

diversity of $20^8 = \sim 10^{10}$. We constructed the DNA library by PCR overlap extension. Synthetic oligonucleotides overlapping with the region of interest contained NNK (N = A, T, G, or C; K

= T or G)[28] codons at the positions we wanted to randomize. The use of the K in the third position excludes two out of three stop codons. After construction of the library, we checked the library by sequencing 15 clones at random (Table 1). All eight amino acids were highly mutated and we did not observe any contamination by the wild type BirA gene.

BirA evolution strategy: bead-based *in vitro* compartmentalization (IVC)

In bead-based IVC, an oil droplet contains single microbead, which provides a solid support linking the gene encoding the enzyme and the product of enzymatic reaction. In particular, streptavidin-coated beads have been widely used, where the bead captures both biotinylated gene and also biotinylated target substrate of the enzyme[23]. In our laboratory, Irwin Chen, Dr. Baruah, and I independently performed an IVC model selection using microbeads to test the validity of the method against BirA evolution, as shown in Figure 2. First, commercially available carboxylate-modifed beads were functionalized with thiols, and maleimide conjugates of streptavidin and AP-fused proteins (generated via succinimidyl 4-[*N*-maleimidomethyl] cyclohexane-1-carboxylate (SMCC) coupling) were then added to the thiol-functionalized beads. The copy number of protein substrates on each bead is expected to be ~10^5 when we assume the coupling efficiency is 100%. One molecule of biotinylated BirA mutant gene was immobilized on each bead via binding to streptavidin (excess beads were added to biotinylated gene). After being emulsified with *in vitro* transcription and translation mixture (IVTM), only active biotin ligases expressed *in vitro* from the bound gene, would ligate biotin onto the AP peptide. The beads were collected after the reaction by breaking the emulsions, and the streptavidin-PE (phycoerythrin fluorophore) conjugate was used to detect biotinylated AP tags on beads by FACS (fluorescence-activated cell sorter). We were able to

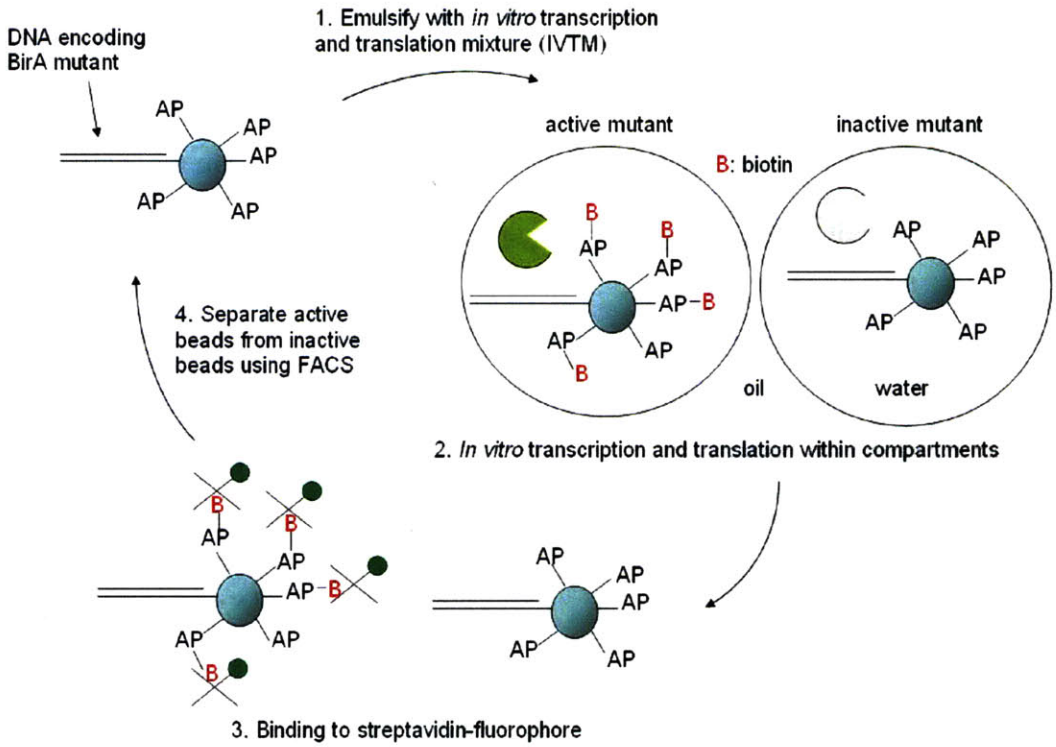


Figure 2 Selection scheme for bead-based IVC. In step 1, BirA mutant genes were attached to beads displaying AP-fused proteins, and the bead-DNA conjugates were mixed with IVTM and segregated in individual aqueous compartments within a water-in-oil emulsion. Within the compartments the BirA mutants are expressed from the gene (step 2). Active mutants (left) biotinylate the AP-fused proteins on beads, whereas inactive mutants (right) do not. After breaking emulsions, the aqueous solution was incubated with streptavidin-fluorophore, which binds to the biotin (represented as "B" in red) on beads (step 3). The beads were washed and sorted using FACS based on the fluorescence intensity. The genes were recovered from the beads through PCR after sorting for input into the next round of selection (step 4).

scan and sort out beads containing genes that express active biotin ligases. A model selection with a mixture of wild type BirA (DNA in the left compartment in Figure 2) and folA gene (DNA in the right compartment in Figure 2) encoding dihydrofolate reductase afforded 1000-fold enrichment of BirA gene over folA gene (data not shown).

Although model selections were successful, we frequently encountered problems

leading to inconsistent results. For efficient product capture, AP-fused proteins should be conjugated to the microbeads. The results of the biotinylation were, however, varied a lot with 1) the protein that is conjugated to beads (e.g. HP1-AP, AP-psBCCP (K→A) (where the lysine biotinylation site of psBCCP (*P. shermanii* transcarboxylase 1.3S subunit) was mutated to alanine), or psBCCP), 2) the batch of commercial beads (which we cannot control), and 3) storage conditions. We also observed that proteins conjugated to the beads were degraded over time. Throughout the protein evolution by IVC, this inconsistency limits the selection process. In addition, although the gene was triply biotinylated to ensure their binding to beads through the process, DNA recovery after selection was sometimes difficult probably due to dissociation of biotinylated gene from streptavidin-coated beads. Unfortunately, our numerous attempts to improve the reproducibility, such as preparing larger batches of microbeads at one time, were not successful. We assumed that microbeads are the reason for inconsistency problems, and decided to design a new IVC platform without microbeads.

BirA evolution strategy: bead-less IVC

In the bead-less IVC strategy, an AP peptide is covalently linked to DNA as shown in Figure 3. The droplet size can be smaller than that of bead-based IVC due to lack of beads, which allows us to input a larger library. Moreover, we can employ an affinity based product capture scheme and avoid a fluorescence-based assay which requires time-consuming flow cytometry.

AP peptide-oligonucleotide conjugates were prepared through aldoxime formation between commercially available 5'-aldehyde modified oligonucleotides and AP peptide with a hydroxylamine group on its N-terminus, followed by reduction with sodium cyanoborohydride. The AP conjugated oligonucleotides are then purified by preparative HPLC (see appendix for

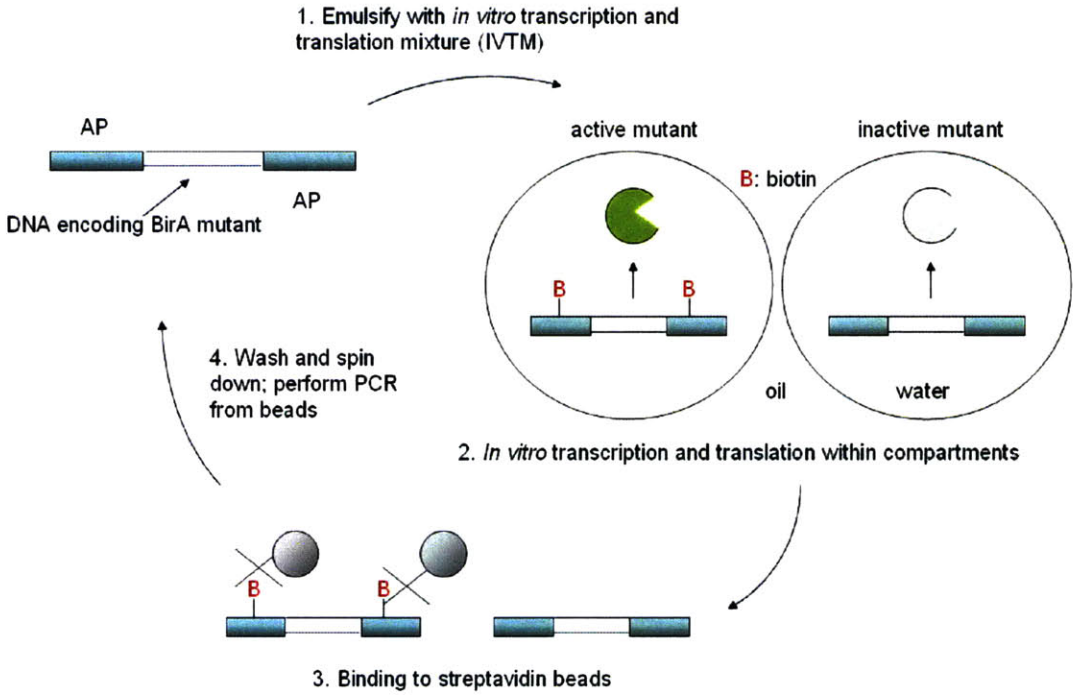


Figure 3 Selection scheme for bead-less IVC. In step 1, AP peptides were covalently attached to both ends of BirA mutant genes, and the peptide-DNA conjugates were mixed with IVTM and segregated in individual aqueous compartments within a water-in-oil emulsion. Within the compartments the BirA mutants are expressed from the gene (step 2). Active mutants (left) biotinylate the AP peptides on the genes, whereas inactive mutants (right) do not. After breaking emulsions, the aqueous solution was incubated with streptavidin beads, which binds to the biotin on the genes (step 3). The beads were washed and spun down. The genes were recovered from the beads through PCR for input into the next round of selection (step 4).

HPLC and MALDI-TOF data of the purified AP conjugated oligonucleotides). We observed efficient biotinylation of the conjugated AP by BirA even in the presence of the oligonucleotide (data not shown). To carry out the model selection with BirA and folA (a negative control enzyme without biotinylation activity), genes for each were amplified by PCR with these AP-modified primers and mixed in a ratio of 1:10 (BirA:folA). This resulting solution was emulsified in mineral oil with IVTM to perform IVC. Biotinylated AP-DNA

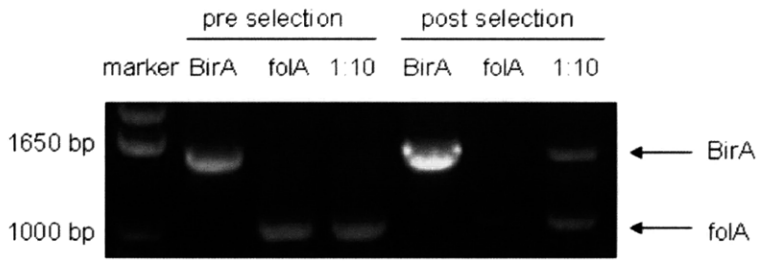


Figure 4 Results of bead-less IVC model selection using BirA and folA. folA is *E. coli* dihydrofolate reductase, a negative control enzyme that lacks biotinylation activity. Genes from before (pre) and after (post) the selection were PCR amplified and analyzed by agarose gel electrophoresis. Selections were performed with BirA only, folA only, and a 1:10 mixture of BirA:folA. It can be seen that the 1:10 mixture becomes enriched in BirA after the selection.

conjugates were captured by streptavidin beads and amplified by PCR reaction (Figures 3 and 4). Consistent enrichment of BirA genes by 25-fold was observed. The enrichment factor is lower than that of bead-based IVC for several reasons. We observed high background biotinylation due to endogenous BirA in *E. coli* lysates of the IVTM. The bead-based platform was less affected by background biotinylation due to the large number of substrates in a compartment. Although it is slightly improved by shorter incubation time, even the small amount of endogenous BirA could biotinylate AP-DNA conjugates under our single-turnover condition, which utilizes only two substrates in one droplet. Also, non-specific binding of DNAs to streptavidin beads decreased the selection efficiency although it could be minimized by pre-blocking of streptavidin beads with Calf thymus DNA.

Model selection with wild type BirA and kinetically impaired G115S BirA.

Bead-less IVC is a two-turnover selection system by its design, which is a disadvantage compared to bead-based IVC, which presents many copies of AP within each droplet. A two-turnover system has less likelihood to enrich highly active catalysts over

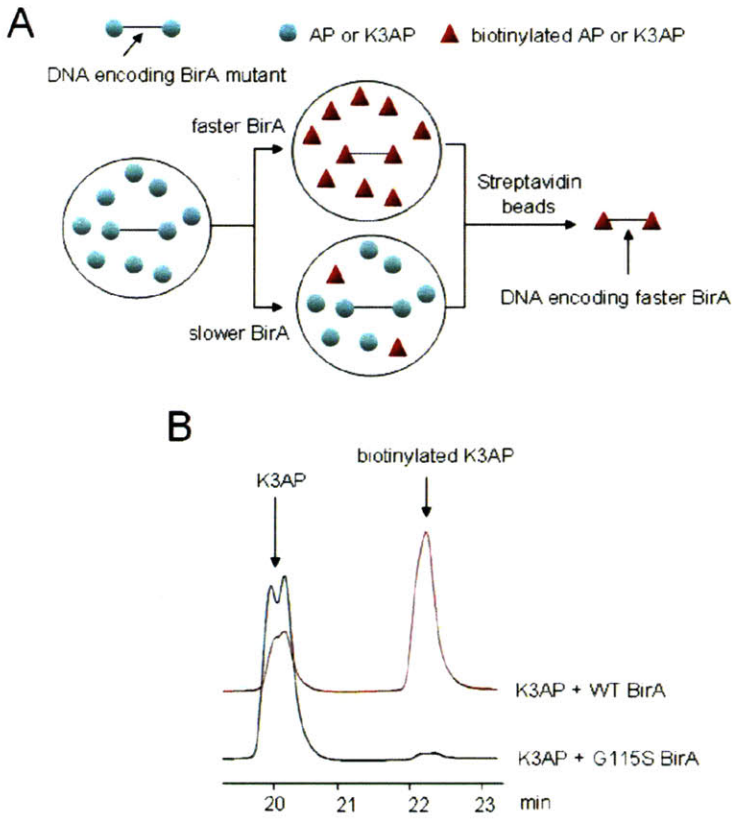
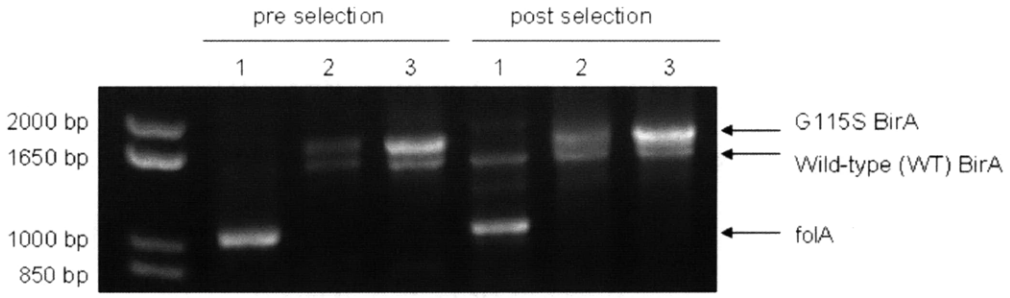


Figure 5 Bead-less IVC model selections using wild type (WT) BirA and kinetically-impaired G115S mutant. G115S BirA shows slower biotinylation compared to WT BirA due to its greater rate of dissociation (k_{off}) for biotin (WT BirA: 0.33 (±0.01) s^{-1}, G115S: 80 (±20) s^{-1})[29]. (A) The horizontal line in the scheme represents DNA. In compartments expressing a fast BirA mutant, such as WT BirA, we expect a high level of biotinylation of both free and bound AP peptide. In compartments expressing a slow BirA mutant, such as G115S, we expect a low degree of biotinylation of free and bound AP peptide. Streptavidin beads are used to selectively pull down genes encoding fast BirA mutants. K3AP is an AP peptide with 3 additional N-terminal lysines to improve solubility. (B) HPLC analysis of K3AP biotinylation by WT BirA and G115S. K3AP peptide was reacted with WT BirA (red) or G115S mutant (blue) and analyzed by HPLC. Reaction condition: 1 μM enzyme (WT or G115S BirA), 1 mM biotin, 100 μM K3AP, 4 mM ATP, 5 mM magnesium acetate, 5 min, 30 °C. Reactions were analyzed by C18 reverse-phase HPLC using a gradient of 10 - 43% acetonitrile in water with 0.1% TFA over 40 min (flow rate 1 mL/min). Chromatograms were recorded at 210 nm.



	Pre selection			Post selection		
lane	1	2	3	1	2	3
Model selection of	BirA vs folA	WT BirA vs G115S BirA	WT BirA vs G115S BirA	BirA vs folA	WT BirA vs G115S BirA	WT BirA vs G115S BirA
IVTM incubation time	6h	3h	6h	6h	3h	6h

Figure 6 Results of model selections with wild type (WT) BirA and G115S using bead-less IVC. Genes from before (pre) and after (post) the selection were PCR amplified and analyzed by agarose gel electrophoresis. Selections were performed with a 1:10 mixture of WT BirA: folA (lane 1), or a 1:10 mixture of WT BirA: G115S (lanes 2 and 3). The difference between lane 2 and 3 is the incubation time with IVTM (3 h for lane 2 and 6 h for lane 3). BirA and folA mixture in lane 1 was incubated with IVTM for 6 h. No enrichment of G115S over WT BirA was observed.

moderately active ones, compared to a multiple-turnover based selection system.

We attempted to introduce multiple-turnover selection to bead-less IVC by adding excess free AP peptide as shown in Figure 5A. In order to improve solubility, K3AP, an AP peptide with three additional N-terminal lysines was added instead of AP. Enzyme expressed in a compartment will more likely use free substrates (~6000 molecules) than immobilized ones on the gene (only two AP's per gene). Concentration of the added K3AP substrates was equal to the K_M value of AP toward BirA (~25 μM[30]). This method was tested by a model selection between wild type BirA and BirA G115S mutant. Gly 115 is located in a disordered loop of BirA that becomes ordered upon biotin binding, and alteration within the loop affects

in enzymatic activity[31,32]. BirA G115S mutant shows a ~200-fold greater rate of dissociation (k_{off}) for biotin than wild type BirA (wild type BirA: 0.33 (±0.01) s^{-1}, G115S: 80 (±20) s^{-1})[29], and thus we observed reduced biotinylation with BirA G115S compared to wild type BirA (68% vs <1% conversion, Figure 5B) using 1 μM enzyme for 5 min at 30 °C. When performing model selections with excess K3AP peptides, however, enrichment of wild type BirA was not observed (Figure 6, lane 3). On the other hand, under identical conditions (with excess free K3AP added), BirA was enriched over folA (Figure 6, lane 1). It is possible that under the conditions of the selection, G115S BirA activity was high enough to biotinylate both free and bound AP, leading to a lack of enrichment over wild type BirA. Addition of more free K3AP peptides than ~6000 molecules / compartment was restricted by the low efficiency of free K3AP peptide removal and by limited binding capacity of streptavidin beads. We also tried to shorten the reaction time (Figure 6, lane 2), but it did not give any enrichment as well.

While our goal is to improve the catalytic efficiency of BirA against the AP peptide, the single-turnover nature of bead-less IVC did not allow us to discriminate kinetically between enzymes. However, our IVC selection could potentially still be a useful tool to find a BirA mutant that accepts unnatural biotin analog probes.

Conclusions

Re-engineering of BirA was attempted by *in vitro* compartmentalization (IVC), which was originally performed using microbeads. We have invented a novel bead-less IVC selection to solve the inconsistency problems of bead-based IVC and observed much more reproducible selections. Several model selections were successfully performed, and the single-turnover characteristic of current selection was analyzed. In parallel, a BirA mutant library for *in vitro* evolution was prepared and characterized.

Although many technical problems of bead-based IVC were solved by inventing bead-less IVC, intrinsic problems of IVC have been clearly noticed. The use of IVTM allows the production of proteins that are toxic to cells. However, the present bead-less IVC, particularly for biotin ligase evolution, was limited by background biotinylation by endogenous BirA in *E. coli* IVTM. This can be resolved by using eukaryotic systems such as wheat germ extracts[33] or rabbit reticulocyte lysates[34]. Unfortunately, these *in vitro* transcription and translation systems are expensive and hard to maintain in house. They are also very sensitive to temperature, and thus when homogenized at high speeds, the generated proteins show loss of their activity[13]. The other drawback of IVTM includes its stickiness. After breaking emulsions, we have purified the genes using size exclusion chromatography. Even though we have repeated washing for three times, the residual IVTM still caused non-specific sticking of DNAs to streptavidin beads, decreasing the enrichment factor of the selection.

Moreover, our knowledge on *in vitro* expression of BirA is still limited. We do not know how long it will take and how many molecules of BirA are expressed in each compartment during the given amount of time. With this information, it would be much easier to design a selection condition that can discriminate faster BirA mutants from others.

Emulsions are created by homogenization, and the size of each compartment varies according to Gaussian-distribution[13]. Therefore, some compartments will have different number of genes or substrates, and this introduces a random bias in the selection. Recently, Weitz and coworkers employed microfluidic devices to create droplets with a uniform size[35]. Although they did not apply the technology to evolve an enzyme yet, it will allow them to perform selections under almost identical conditions while preserving the quality of IVTM.

The use of PCR to amplify the resulting clones from each round also limits the IVC method. We had to perform at least 52 cycles of PCR to get enough DNA for next rounds in bead-based or bead-less IVC. Such high numbers of PCR cycles sometimes generate chimeric fragments by homologous recombination, and resulting shorter DNA fragments are more readily amplified by PCR than original templates, producing false DNA product[36]. In order to solve the problem in part, Griffiths and coworkers developed emulsion PCR, where each gene is compartmentalized with PCR reaction mixture[37]. They observed narrow distribution of the PCR product size when PCR reactions were emulsified.

In order to evolve faster BirA using IVC, the selection should be under multiple-turnover condition. Although bead-based IVC offers multiple turnover selection, microbeads caused most of inconsistency problems of the method. Replacing microbeads with other solid support such as yeast cells might improve the consistency of IVC. Yeast cells displaying AP peptides were proven as efficient substrates for BirA (Chapter 2). By emulsifying each gene encoding a mutant of BirA and yeast cell(s) displaying multiple AP peptides, we will be able to achieve multiple-turnover selection. These emulsions can then be re-emulsified to generate water/oil/water double emulsions, which can be sorted using FACS. Genes in the sorted compartments will be amplified by emulsion PCR.

Experimental

Cloning of a partially-randomized BirA library by overlap extension PCR. The library was synthesized from five randomized BirA gene fragments 1 - 5 using overlap extension PCR (Figure 7). In a typical overlap extension PCR reaction, 1.7 nM of each fragment were mixed in the presence of Roche Expand High Fidelity^{PLUS} PCR System and the reactions were cycled

10 times (94 °C 0.5 min, 50 °C 0.5 min, 72 °C 3 min) without primers. Then a pair of flanking primers was added, and the reactions were cycled 15 more times. Fragment 1 was amplified from the pIVEX BirA template using the primers LMB2-1 (5'- CAG GCG CCA TTC GCC ATT) and the rvT90 (5'- GTC AAT CAC TGG CAG CAC G), using 30 reaction cycles as described above. Fragment 2 was similarly amplified using the primers fwT90 (TGC TGC CAG TGA TTG ACT CCN NKA ATC AGT ACC TTC TT, where N= A, G, C, or T and K=G or T; randomizes Thr 90) and rvQ112 (5'- CTG CTG GTA TTC TGC AAT GC). Fragment 3 was similarly amplified with the primers fwQ112 (5'- GAT GCT TGC ATT GCA GAA TAC NNK CAG GCT NNK CGT GGT CGC CGG GGT CGG AAA NNK TTT TCG CCT TTT GG; randomizes Gln 112, Gly 115, and Trp 123) and rvG186 (5'- TGC CAG CTT GCG ATC CTG). Fragment 4 was similarly amplified with the primers fwG186 (5'- CAG GAT CGC AAG CTG GCA NNK ATT CTG GTG GAG C, where N= A, G, C, or T and K=G or T; randomizes Gly 186) and rvG204 (5'-AAT GAC TAT TTG CGC CGC AT). Fragment 5 was similarly amplified with the primers fwG204toG206 (5'-ATG CGG CGC AAA TAG TCA TTN NKN NKN NKA TCA ACA TGG CAA TG; randomizes Gly 204, Ala 205, and Gly 206) and PIVB1 (5'- GCG TTG ATG CAA TTT CT). Finally, the entire fragment (1+2+3+4+5) was PCR amplified with triply biotinylated primers LMB2-1 and PIVB1.

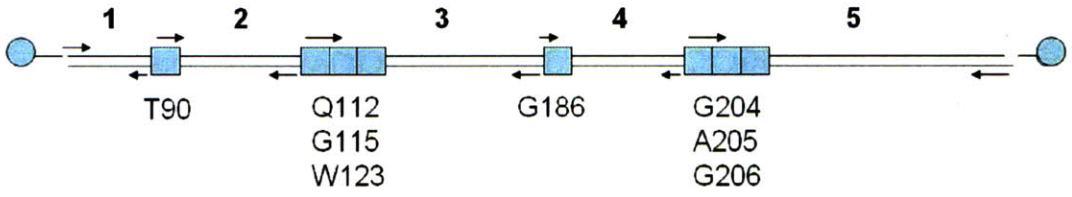


Figure 7 The BirA library was created using overlap extension PCR. Five fragments 1 - 5 of gene were shown. Each end of the gene was biotinylated for immobilization on streptavidin-coated beads.

Synthesis of microbeads functionalized with HP1-AP, psBCCP, or AP-psBCCP (K→A)

Cloning and bacterial expression of HP1-AP, psBCCP, and AP-psBCCP (K→A). The pET-23d plasmid for bacterial expression of *P. shermanii* transcarboxylase 1.3S subunit (psBCCP) was a kind gift from A.D. Griffiths and B. T. Kelly (MRC Laboratory of Molecular Biology, UK). Others in the lab constructed the pET21 plasmid for bacterial expression of HP1-AP by standard PCR subcloning techniques.

In order to clone AP-psBCCP (K→A), site-directed mutagenesis on the lysine biotinylation site of psBCCP was performed with psBCCP (K/A).F (5'-GCT CGT TCT CGA GGC CAT GGC GAT GGA GAC CGA GAT CCA AC) and psBCCP (K/A).R (5'-GTT GGA TCT CGG TCT CCA TCG CCA TGG CCT CGA GAA CGA GC) using QuikChange system (Stratagene). Two successive PCR reactions were performed to construct an insert containing a HA tag, AP, and psBCCP. For the first PCR, the psBCCP (K→A) gene was used as the template, and the primers HAAPfw1 (5'-GCC TCA ATG ACA TCT TCG AAG CCC AGA AAA TCG AAT GGC ACG AAA TGA AAC TGA AGG TAA CAG TCA ACG GCA C; incorporates AP sequence) and HAAPrv1 (5'-TCG ATG *AGC TC*G AGA TCC CCG ATC TTG ATG AGA CCC TGA C; incorporates a *SacI* site) were used for the amplification. For the second PCR, the product of the first PCR reaction was used as the template and the primers HAAPfw2 (5'- TTC ATG *GGA TCC* TAC CCC TAC GAC GTC CCC GAC TAC GCC GGC GGC TCC GGC GGC CTC AAT GAC ATC TTC GAA GCC; incorporates the HA sequence and a *BamHI* site) and HAAPrv1 were used for the amplification. The insert was digested with *BamHI* and *SacI* and ligated in-frame to similarly digested pET-23d vector (Novagen), which introduces a C-terminal His_6 tag.

The pET-23d plasmid containing the psBCCP or AP-psBCCP gene and the pET-21b

plasmid containing the HP1-AP gene were introduced into *E. coli* strain BL21 (DE3) by heat-shock transformation. The cells were cultured in LB supplemented with ampicillin (100 μg/mL) at 37 °C until OD_{600} 0.5. Protein expression was induced by the addition of IPTG to a final concentration of 0.4 mM. After incubation at 30 °C for 3 h, the cells were harvested by centrifugation. The cells were lysed by sonication at 4 °C (three 10-second pulses at half-maximal power with 1 min between each pulse) in lysis buffer (50 mM Tris pH 7.8, 300 mM NaCl) containing 0.5 mM phenylmethylsulfonyl fluoride (PMSF), protease inhibitor cocktail (Calbiochem), and DNase I (Roche). After incubation at room temperature for 20 min, the lysate was cleared by centrifugation, and the supernatant was loaded onto a Ni-NTA agarose column (Qiagen). Fractions containing the desired proteins (psBCCP, AP-psBCCP (K→A) or HP1-AP) were consolidated and dialyzed against PBS pH 7.4 overnight. The concentration of protein was determined by BCA assay. Removal of biotinylated proteins was performed with streptavidin-agarose beads as described elsewhere. The cysteine residue on HP1-AP was capped by incubating with 4 mM N-ethylmaleimide for 30 min at room temperature, and excess N-ethylmaleimide was removed by dialysis.

Synthesis of thiol-modified beads (modified from Irwin Chen and Dr. Baruah's protocol)

Microbeads were vortexed before use. A total of 25 μL of carboxylate-modified microbeads (10% suspension, ~4 x 10^9 beads, Bangs laboratories) were washed (resuspended in buffer and pelleted by centrifugation (13,200 rpm, 2 min, 20 °C)) with PBS-T (phosphate-buffered saline +0.1% Tween) once. The beads were then pelleted and washed once with 500 μL of MES buffer (50 mM 2-(*N*-morpholino)ethanesulfonic acid pH 6.2) and resuspended in 400 μL MES buffer. To the bead suspension was added sulfo-*N*-hydroxysuccinimide to a final concentration of 100 mM and then 1-ethyl-3-(3-dimenthylaminopropyl) carbodiimide hydrochloride (EDC)

to a final concentration of 100 mM in MES buffer. The pH of the reaction was maintained at 6.2. The beads were agitated for 20 min at room temperature and then washed twice with 500 μL MES buffer. The beads were resuspended in 200 μL of a solution of 1 mM cystamine in bicarbonate buffer (0.1 M $NaHCO_3$ pH 8.2, prepared every time before use) and agitated for 2.5 h. The beads were washed twice with 500 μL PBS-T, resuspended in 500 μL of PBS containing 20 mM dithiothretiol (DTT), and then agitated at room temperature for 1.5 h. The beads were washed three times with 500 μL PBS-T prior to conjugation to protein.

Synthesis of protein-SMCC conjugates (modified from Irwin Chen and Dr. Baruah's protocol) In the meantime, both streptavidin (either prepared in house or purchased from New England Biolabs) and the protein substrate were reacted with succinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate (SMCC). To make psBCCP-functionalized beads, SMCC was added to 50 μL of psBCCP solution (20 mg/mL in PBS) and 50 μL of streptavidin solution (1.5 mg/mL in PBS) to final concentrations of 4.8 mg/mL and 0.7 mg/mL, respectively. The reactions were incubated at room temperature for 45 min and the protein-SMCC conjugates were purified from excess SMCC using size exclusion chromatography on a NAP-5 (SephadexTM G-25 resin, GE Healthcare) column. The proteins were eluted from the column in 10x100 μL fractions in PBS supplemented with 1 mM EDTA. The four most concentrated fractions of psBCCP-SMCC conjugate were combined with the two most concentrated fractions of streptavidin-SMCC.

To make HP1-AP-functionalized beads, SMCC was added to 120 μL of HP1-AP solution (1.0 mg/mL in PBS) and 50 μL of streptavidin (1.5 mg/mL in PBS) to final concentrations of 0.2 mg/mL and 0.3 mg/mL, respectively. Following 45 min, the reactions were purified on a NAP-5 column as described above for psBCCP. The HP1-AP and

streptavidin-SMCC conjugates were mixed as described above for psBCCP.

Conjugation of protein-SMCC to thiol-functionalized beads The thiol-functionalized beads synthesized above were resuspended (by sonication) in the streptavidin-protein substrate of BirA-SMCC mixtures and agitated at room temperature overnight. The beads were washed three times with 500 μL of PBS-T and resuspended in 225 μL of PBS-T to make a 1% suspension, assuming a 90% recovery of the beads. This corresponds to 1.6 x 10^7 beads/ μL, or a concentration of 27 pM.

Model selection of BirA and folA using bead-based IVC (modified from B. Kelly's protocol for use in Ting Lab).

Emulsification of microbeads. A total of 1.28 x 10^8 beads were pre-blocked in 40 μL PBS-T containing 8 mg/mL heparin for 1 h. The beads were washed twice with PBS-T, resuspended in a solution of biotinylated BirA DNA in 40 μL PBS-T, and agitated for 30 min. In typical experiments, the DNA was bound to the beads in a 1:3 ratio (DNA/bead). For model selections, BirA gene and folA gene (dihydrofolate reductase (DHFR) gene) were amplified from pIVEX plasmids harboring BirA and folA genes (a kind gift from A.D. Griffiths and B. T. Kelly) with 5'-biotinylated primers. The biotinylated BirA and folA genes were mixed in 1:10, 1:100, and 1: 1000 ratios (BirA/folA). The beads were washed twice with TA buffer (5 mM Tris-acetate pH 8.0), resuspended by sonication in 20 μL TA, and chilled on ice. To the bead suspension was added 33 μL of chilled 1x IVTM from the RTS 100 E. coli HY Kit (Roche), additionally supplemented with 23 μM of biotin. 50 μL of the resulting well-mixed bead suspension was then added to a 2 mL Nunc CryoTube vial containing 500 μL of chilled mineral oil (Sigma) with 4.5% w/w Span-80 and 0.5% w/w Triton-X-100. The water-in-oil emulsions were formed

by stirring on ice at 1100 rpm for 3 min and homogenizing further on ice at 11,000 rpm for 3 min using an IKA Labortechnik Ultra-Turrax T8 homogenizer. The emulsions were then incubated at 30 °C for at least 8 h.

Breaking the emulsion. The emulsion was transferred to a 1.5 mL eppendorf tube and centrifuged (13,200 rpm, 15 min, 15 °C). The less dense oil phase was removed and the remaining emulsion was broken by three extractions with 1 mL mineral oil. One round of extraction involves resuspension of the emulsion in oil, centrifugation (13,200 rpm, 10 min, 15 °C), and removal of the oil phase. After the emulsion was broken to reveal two distinct phases, an oil phase on top and an aqueous phase containing the bead pellet, both layers were removed carefully, and the pellet was resuspended in 100 μL of PEET buffer (PBS-T supplemented with 5 mM ethylenediamine tetraacetic acid (EDTA) and 5 mM ethylene glycol tetraacetic acid (EGTA)). One mL of mineral oil was added; the mixture was vortexed; the beads were pelleted by centrifugation (13,200 rpm, 3 min, 15 °C); and both phases were removed from the pellet. The beads were resuspended in 100 μL of PEET, and residual oil was removed from the aqueous layer by three extractions with 1 mL hexane. The remaining hexane in the aqueous layer was evaporated by heating at 30 °C for 15 min. The beads were pelleted and washed twice with wash buffer (10 mM Tris pH 7.4 with 1 mM EDTA and 2 M NaCl).

Streptavidin-PE staining for biotinylated beads. The beads were stained with streptavidin-PE (1:100 dilution in 0.5x wash buffer, Jackson ImmunoResearch) for 1 h. The beads were washed three times with PBS-T, and resuspended in 500 μL PBS-T for flow cytometry (FACScan, BD Biosciences) or bead sorting.

DNA recovery from the beads. Following a sort, the collected beads were pelleted by centrifugation (13,200 rpm, 10 min, 20 °C) and washed twice in 1x PCR buffer and

resuspended in 25 μL of PCR buffer. The beads were input into 100 μL PCR reactions with the primers LMB2-7 (5'-AAG TTG GGT AAC GCC AGG) and PIVB-7 (5'-TAC GCG ATC ATG GCG AC), and the reactions were cycled 22 times (94 °C 0.5 min, 50 °C 0.5 min, 72 °C 2 min) before holding at 72 °C for 7 min. Then, 1 μL of this PCR reaction served as the template for the next PCR reaction with nesting primers (30 cycles like above).

Bead-less IVC

Conjugation of AP peptide and primers. Two aldehyde modified custom primers ALD-LMB2-3 (GGA AGG GCG ATC GGT GCG) and ALD-PIVB3 (GAG CAC TGT CCG ACC GC) were purchased from Invitrogen. We made 100 μM stock solutions of these primers in water. The conjugation reaction with 50 μM of primers, 10 mM NaOAc pH 4.5, and 1 mM AP-hydroxylamine (hydroxylamine on N-terminus, custom synthesized from Tufts University Core Facility, used as 10 mM stock solution in water and minimal sodium phosphate pH 8) was incubated at room temperature for 6 h. Sodium cyanoborohydride was then added to 50 mM of final concentration, and the reaction was further incubated at room temperature for 3 h. The reaction was purified by C18 reverse-phase HPLC using a gradient of 3 - 60% acetonitrile in 50 mM ammonium acetate pH 7 over 30 min (flow rate 1 mL/min). Chromatogram was recorded at 260 nm. Intact oligonucleotides have retention times of 8.5 min, but after conjugation to AP peptide, the retention time should be increased to ~13.5 min. After purification, each primer was run on HPLC to estimate the purity of the conjugate (Figure 1 in Appendix). The identity of conjugation product was confirmed by MALDI-TOF measurement (Figure 2 in Appendix). For some reason, we have failed to get mass peaks corresponding to AP conjugated PIVB3, even with desalted and concentrated samples. However, the retention

time of AP conjugated PIVB3 (~13.5 min) was different from starting materials (AP-hydroxylamine peptide: 16 min, ALD-PIVB3: 8.5 min). The mock conjugation control without AP peptide also showed that the retention time of ALD-PIVB3 does not vary much after reduction. Therefore, we assumed that the purified material from the conjugation reaction is actually AP conjugated PIVB3.

AP conjugated genes (wild type BirA, folA, and G115S BirA) were prepared by PCR reactions using these AP conjugated primers and the corresponding templates.

Preparation for making emulsions. 500 μL of mineral oil mix were put in a Nunc CyroTube vial with an 8x2 mm stir bar, and the vial was placed on ice to chill the oil mixture. AP-conjugated genes were prepared as a 3.3 nM solution in 10 mM Tris-acetate pH 8. This will give a final concentration of DNA in the IVTM of 275 pM. Two falcon tubes containing water and absolute ethanol separately were prepared to rinse the homogeneizer.

Making DNA/IVTM mixtures. In a 650 μL eppendorf tube, 22 μL of sterile water and 5 μL of 3.3 nM stock solution of AP conjugated gene construct were mixed. For model selections, DNA mixture was combined to achieve a final concentration of 275 pM in 60 μL. To make 1:10, 1:100, and 1: 1000 dilutions (BirA: folA or WT: G115S BirA), serial dilutions of original stock solution were made. For model selections with wild type BirA and G115S mutant, K3AP was added to DNA mixture to a final concentration of 25 μM. The tube was then placed on ice. IVTM mixture was generated on ice by combining 11.5 μL of lysate, 9.6 μL of reaction mixture, 11.5 μL of amino acids, 0.98 μL of methionine, 5 μL of reconstitution buffer, and 0.9 μL of 1mM biotin. Immediately prior to forming emulsions, 33 μL of IVTM was added to the 27 μL of gene construct made as above. The reaction mixture was remained on ice before homogenizing.

Forming Emulsions. The tube containing the mineral oil mixture was placed on ice and put on the stirrer at 1,150 rpm. 50 μL of IVTM-DNA mixture were added to the stirring oil and stirred further for 3 min. The oil became cloudy, and the tube was removed from the stirrer, and the stir bar was also removed from the tube. The tip of the homogenizer was placed into the tube and the mixture was homogenized at 25,300 rpm for 3 min. The homogenizer was then put out, and the tube was capped and incubated at 30 °C for overnight. The homogenizer was cleaned with one thorough wash with ethanol, followed by a thorough wash with water. All traces of water in the final wash should be removed before next use.

Break Emulsions. 200 μL of PBS-T was added to emulsions and the mixture was transferred to a clear 1.5 mL eppendorf tube. The mixture was then spun down at 15 °C with 13,000 rpm for 10 min. We usually saw a pellet of emulsions at the bottom of the tube, then an aqueous layer of quench buffer, and a top layer of mineral oil. The layer of mineral oil was removed. 1 mL of pure mineral oil (Sigma) was added to the tube and the pellet was re-suspended by flicking with fingers. The tube was again spun down for 5 min at 15 °C with 13,000 rpm. The top layer was removed again. Until no more pellet is observed and the solution is clear, mineral oil was added and removed as above for three to four more times. Any white interphase was removed. In order to remove residual mineral oil, 1 mL of hexane was added. The mixture was resuspended and spun down for 1 min at 13,000 rpm. The hexane layer was removed, and this hexane washing was repeated for two more times. Residual hexane was evaporated off by air-drying for 10 min. The aqueous phase is then purified with size exclusion column (G-25, Bio-Rad) to remove residual IVTM or excess K3AP peptide.

Binding to streptavidin beads. M-280 streptavidin-coated dynabeads (50 μL) were washed twice with 2X binding and washing buffer plus Triton X-100 (BWT) buffer (10 mM Tris-HCl,

pH 7.5/ 1 mM EDTA/ 2.0 M NaCl/ 0.1% vol/vol Triton X-100). The beads were then resuspended in 2X BWT buffer. 180 μL of IVTM reaction was added to the beads and the rest 20 μL was saved for future PCR. The beads were then rotated at room temperature for 15 min and washed with 200 μL of 2X BWT; 200 μL of GHCl Buffer (3 M guanidinium hydrochloride/ 25 mM sodium phosphate, pH 7.2/ 1mM EDTA); 200 μL of 2X BWT; 200 μL of water. The beads were finally resuspended in 25 μL of water.

DNA recovery by PCR from beads. 10 μL of the beads above was added to a PCR reaction solution on ice containing 10 μL of 10X Pfu buffer, 4 μL of mixture of dNTPs, 1 μL of 100 μM primer 1 (PIVB-9; 5'- TAT CCG GAT ATA GTT CC), 1 μL of 100uM primer 2 (LMB2-9; 5'- GTA AAA CGA CGG CCA GT), 2 μL of Pfu-Turbo DNA polymerase, and 72 μL of sterile water. PCR was performed for the following cycles: a) 94 °C 2 min, b) 94 °C 10 sec, c) 50 °C 30 sec, d) 72 °C 54 sec, e) repeat b) to d) 19 times, f) 72 °C for 7 min, g) 4 °C forever. 2 μL of this PCR product was taken to the nested PCR reaction solution containing 10 μL of 10X Pfu buffer, 4 μL of mixture of dNTPs, 1 μL of 100 μM primer 1 (sacI-outer-R; 5'- GCT TTG TTA CCG GAT CCC GG), 1 μL of 100 μM primer 2 (PVX-rbs-ncoI-F; 5'- TTA AGA AGG AGA TAT ACC ATG G), 2 μL of Pfu-Turbo DNA polymerase, and 72 μL of sterile water. PCR were cycled under the same condition as above. PCR products were analyzed on an agarose gel.

References

1 Radzicka, A. & Wolfenden, R., A proficient enzyme. *Science* 267, 90-93 (1995).

2 Jaeger, K.E. & Eggert, T., Enantioselective biocatalysis optimized by directed evolution. *Curr. Opin. Biotechnol.* 15, 305-313 (2004).

3 O'Brien, P.J. & Herschlag, D., Catalytic promiscuity and the evolution of new

enzymatic activities *Chem Biol* 6, R91-R105 (1999).

4 Moore, J.C. & Arnold, F.H., Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents *Nat. Biotechnol* 14, 458-467 (1996).

5 Griffiths, A.D. & Tawfik, D.S., Man-made enzymes — from design to in vitro compartmentalisation. *Curr. Opin. Biotechnol.* 11, 338-353 (2000).

6 Besenmatter, W., Kast, P., & Hilvert, D., New enzymes from combinatorial library modules. . *Methods Enzymol.* 388, 91-102 (2004).

7 Yuan, L., Kurek, I., English, J., & Keenan, R., Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 69, 373-392 (2005).

8 Stemmer, W.P., Rapid evolution of a protein in vitro by DNA shuffling. *Nature* 370, 389-391 (1994).

9 Ostermeier, M., Shim, J.H., & Benkovic, S.J., A combinatorial approach to hybrid enzymes independent of DNA homology. *Nat. Biotechnol* 17, 1205-1209 (1999).

10 Sieber, V., Martinez, C.A., & Arnold, F.H., Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol* 19, 456-460 (2001).

11 Voigt, C.A., Martinez, C., Wang, Z.-G., Mayo, S.L., & Arnold, F.H., Protein building blocks preserved by recombination. *Nat. Struct. Biol.* 9, 553-558 (2002).

12 Tawfik, D.S. & Griffiths, A.D., Man-made cell-like compartments for molecular evolution *Nat. Biotechnol* 16, 652-656 (1998).

13 Miller, O.J. *et al.*, Directed evolution by in vitro compartmentalization *Nat. Methods* 3, 561-570 (2006).

14 Rothe, A., Surjadi, R.N., & Power, B.E., Novel proteins in emulsions using in vitro compartmentalization. *Trends. Biotechnol.* 24, 587-592 (2006).

[15] 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* 12, 1291-1300 (2005).

[16] Link, D.R. *et al.*, Electric control of droplets in microfluidic device. *Angew Chem Int Ed Engl* 45, 2556-2560 (2006).

[17] Leamon, J.H., Link, D.R., Egholm, M., & Rothberg, J.M., Overview: methods and applications for droplet compartmentalization of biology. *Nat. Methods* 3, 541-543 (2006).

[18] Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N., & Yanagawa, H., In vitro selection of restriction endonucleases by in vitro compartmentalization. *Nucleic Acids Res.* 32, e95 (2004).

[19] Ghadessy, F.J., Ong, J.L., & Holliger, P., Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci USA* 98, 4552-4557 (2001).

[20] Ghadessy, F.J. *et al.*, Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat. Biotechnol* 22, 755-759 (2004).

[21] 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.* 17, 3-11 (2004).

[22] Lee, Y.F., Tawfik, D.S., & Griffiths, A.D., Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. *Nucleic Acids Res.* 30 (4937-4944) (2002).

[23] Griffiths, A.D. & Tawfik, D.S., Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. *EMBO J.* 22, 24-35 (2003).

[24] Aharoni, A., Griffith, A.D., & Tawfik, D.S., High-throughput screens and selections of enzyme-encoding genes. *Curr. Opin. Chem. Biol.* 9, 210-216 (2005).

[25] Agresti, J.J., Kelly, B.T., Jaschke, A., & Griffiths, A.D., Selection of ribozymes that catalyse multiple-turnover Diels–Alder cycloadditions by using in vitro compartmentalization. *Proc. Natl. Acad. Sci. USA* 102, 16170-16175 (2005).

[26] Fernandez-Suarez, M., Chen, T.S., & Ting, A.Y., Protein− Protein Interaction Detection in Vitro and in Cells by Proximity Biotinylation. *J. Am. Chem. Soc* 130, 9251-9253 (2008).

[27] Wilson, K.P., Shewchuk, L.M., Brennan, R.G., Otsuka, A.J., & Matthews, B.W., Escherichia coli biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc Natl Acad Sci USA* 89, 9257-9261 (1992).

[28] Hughes, M.D., Nagel, D.A., Santos, A.F., Sutherland, A.J., & Hine, A.V., Removing the redundancy from randomised gene libraries. *J. Mol. Biol.* 331, 973-979 (2003).

[29] Kwon, K. & Beckett, D., Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci.* 9, 1530-1539 (2000).

[30] Beckett, D., Kovaleva, E., & Schatz, P.J., A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* 8, 921-929 (1999).

[31] Choi-Rhee, E., Howard, S., & Cronan, J.E., Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein Sci.* 13, 3043-3050 (2004).

[32] Weaver, L.H., Kwon, K., Beckett, D., & Matthews, B.W., Corepressor-induced organization and assembly of the biotin repressor: A model for allosteric activation of a transcriptional regulator. . *Proc Natl Acad Sci USA* 98, 6045-6050 (2001).

[33] Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., & Yanagawa, H., DNA

display for in vitro selection of diverse peptide libraries *Nucleic Acids Res.* 31, e118 (2003).

[34] Ghadessy, F.J. & Holliger, P., A novel emulsion mixture for in vitro compartmentalization of transcription and translation in the rabbit reticulocyte system. *Protein Eng. Des. Sel.* 17, 201-204 (2004).

[35] Clausell-Tormos, J. *et al.*, Droplet-Based Microfluidic Platforms for the Encapsulation and Screening of Mammalian Cells and Multicellular Organisms. *Chem Biol* 15, 427-437 (2008).

[36] Meyerhans, A., Vartanian, J.P., & Wain, H.S., DNA recombination during PCR. *Nucleic Acids Res.* 18, 1687-1691 (1990).

[37] Williams, R. *et al.*, Amplification of complex gene libraries by emulsion PCR. *Nat. Methods* 3, 545-550 (2006).

Chapter 4: Development of a Streptavidin Evolution Platform Based on *In Vitro* Compartmentalization

This chapter describes our efforts to evolve streptavidin variants that bind a ketone analog of biotin, using *in vitro* compartmentalization (IVC). My specific contributions were as follows. The selection scheme and model selections with wild type and S45A streptavidin (SA) were developed by Dr. Dan Chinnapen and Dr. Guobin Luo, post-doctoral fellows in the Ting laboratory. I reproduced their model selection. I then prepared two streptavidin mutant libraries and performed three rounds of selection with ketone biotin.

Introduction

Biotin binding to streptavidin is one of the strongest non-covalent interactions (K_d = 10^{-13} M) found in Nature[1]. By virtue of its unusual stability, the streptavidin-biotin pair has been used in many fields from biochemistry to nanotechnology[2,3]. The protein is a homo-tetramer, and each subunit (~15 kDa) binds to one molecule of biotin. The crystal structure of biotin-bound streptavidin showed that various types of interactions are simultaneously involved to achieve the extremely high affinity of streptavidin to biotin. Four tryptophan residues (Trp 79, Trp 92, Trp 108, and Trp 120) of the protein contribute the van der Waals type interaction with biotin. Seven hydrogen bonding interactions further increase the affinity, and an ureido-oxyanion resonance form of biotin is stabilized by the extended hydrogen bonding network including Ser 45 and Asp 128 (Figure 1). These two residues particularly play a crucial role in the affinity of the protein to biotin. When one of the two residues was mutated to alanine, the protein affinity to biotin was decreased 1000-fold. Residues 45 to 52 also form an open binding loop in the apo form, which is closed after biotin binds.

Altering the extraordinary binding specificity of streptavidin toward biotin has been of high interest since the development of a new streptavidin (streptavidin 2)-biotin analog (biotin 2) pair will extend the range of current streptavidin-biotin technology (Figure 2). Moreover, if the newly developed streptavidin 2-biotin 2 pair is orthogonal to streptavidin-biotin pair, meaning the streptavidin 2 does not bind to biotin with high affinity while still exhibiting great affinity toward biotin 2, the pair could be used simultaneously with the streptavidin-biotin pair. Since the binding site of streptavidin is highly optimized for biotin, any kind of mutations in the area will cause a reduced affinity toward biotin[4]. Cantor and coworkers have generated streptavidin mutants to reduce the binding affinity to biotin while maintaining affinities to

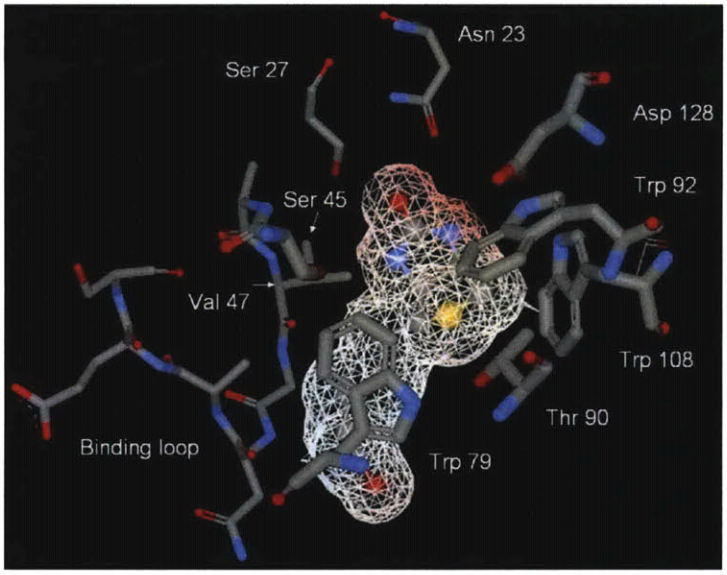


Figure 1 Crystal structure of the biotin binding site of streptavidin. The surface of biotin is shown with the wire mesh. Residues close to biotin and the binding loop (residues 45 to 52) are labeled.

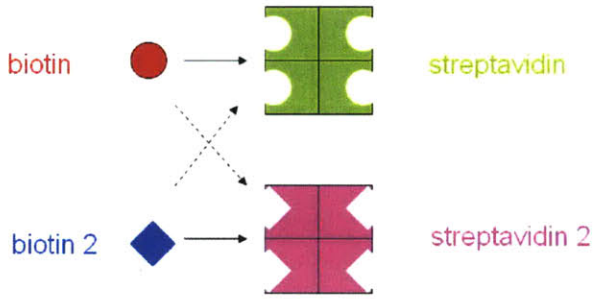


Figure 2 A proposed orthogonal pair of streptavidin and biotin analogs. "Streptavidin 2" does not bind to biotin, and wild type streptavidin does not bind to "biotin 2" (dotted arrows). However, streptavidin 2 binds to biotin 2 with high affinity (solid arrows), like the wild type streptavidin-biotin interaction.

2-iminobiotin and diaminobiotin (Figure 3)[5]. In order to change the local electrostatic charge distribution of the binding pocket, they created a N23A and S27D double mutant based on molecular modeling and free energy calculations. They found that this mutant shows higher affinity to 2-iminobiotin (K_d = 1 x 10^{-6} M) than to biotin (1 x 10^{-4} M), although the affinity to diaminobiotin (3.7 x 10^{-5} M) is lower than that to biotin (3.4 x 10^{-6} M). Although the affinity

achieved is too weak for labeling applications, they showed that the ligand specificity of streptavidin can be modified using *de novo* design. Dixon and coworkers also employed a computational approach to find a biotin analog that binds streptavidin tighter than biotin. 9-R-methylbiotin (Figure 3), which has a methyl group immediately before the carboxylic acid group of the aliphatic tail, was observed to have 10-fold higher K_d value (10-fold weaker interaction) than biotin to streptavidin[6].

Combinatorial approaches have been employed to change the ligand specificity as well. Cantor and coworkers accidentally found a single-chain dimeric (SCD) streptavidin mutant that exhibits a K_d of 10^{-10} M for biotin-4-fluorescein (B4F, Figure 3) whereas its K_d for biotin is as low as 10^{-5} M[7]. Originally, their goal was to create a SCD streptavidin with as high affinity to biotin as wild type. To achieve this, a phagemid library of SCD was created by a

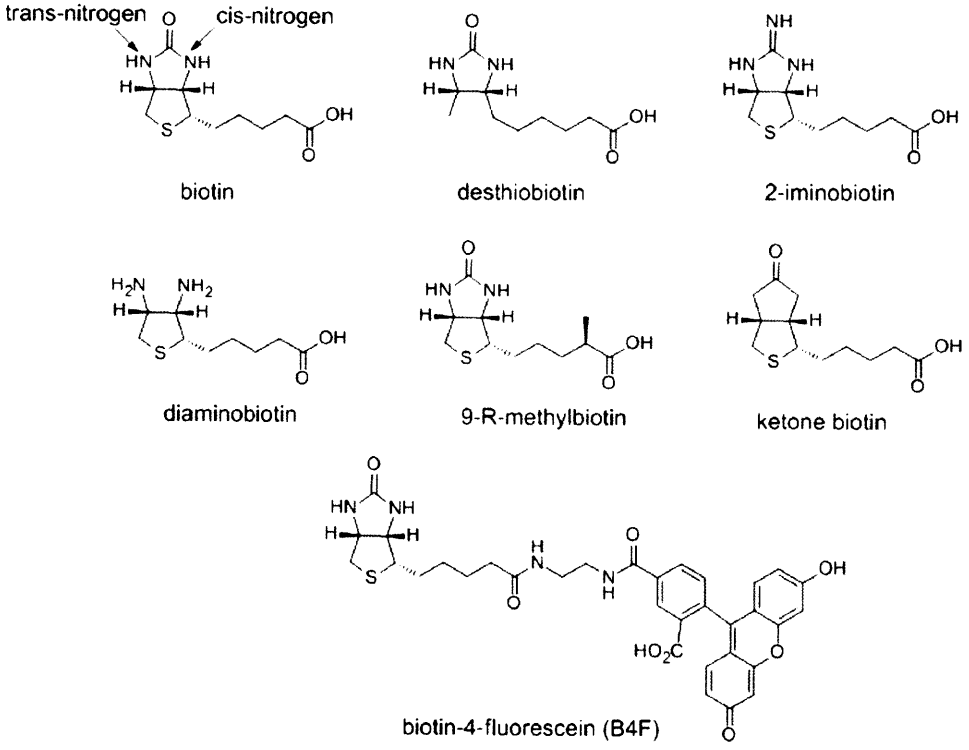


Figure 3 Structure of biotin analogs that have been investigated for streptavidin engineering.

circular permutation of streptavidin gene followed by error-prone PCR. They explain that the unexpected tight binding of B4F to the SCDs is due to the fluorescein moiety, complementing Trp 120 of wild type streptavidin. In addition to searching for streptavidin mutants that bind small molecules other than biotin, streptavidin has also been engineered to have a higher affinity toward the Strep-tag II peptide, a small streptavidin binding peptide[8]. Functional streptavidin mutants from *E. coli* colonies were captured by anti-streptavidin antibody, and their binding to Strep-tag II was assayed using alkaline phosphatase-fused Strep-tag II. After incubated with chromogenic substrates, streptavidin mutants that bind the Strep-tag II could be screened. At the outset of the project, however, directed evolution of streptavidin with systematic binding assays was not reported.

As reviewed in Chapter 1 and Chapter 3, *in vitro* compartmentalization (IVC) has been successfully used for directed evolution of enzymes. Selections for binding, however, were mostly performed via panning in phage display. Usually, proteins were displayed on phage, RNA, or DNA rather than on beads since beads are too heavy for the relatively weak binding forces[9]. Omitting the transformation step, IVC can be superior to other *in vivo* selections (e.g. phage display) because a larger library can be produced. Various protein binders were successfully engineered using bead-less IVC. For example, zinc finger proteins binding specific DNA sequence with high affinity have been evolved from a background of a related zinc finger protein mutants[10]. Variants from the SH3 domain of the Fyn kinase were also generated using this technique to have nanomolar affinities to mouse serum albumin (MSA)[11]. More interestingly, novel p53 variants with altered response elements specificities were evolved using IVC[12]. In the study, the P75L mutant of p53 which is transactivation-

deficient in yeast was selected using IVC, highlighting the advantage of *in vitro* selection which is independent from cell-viability.

In the present chapter, our attempt to evolve an orthogonal biotin 2/ streptavidin 2 pair using IVC will be discussed. We envisioned that ketone biotin (Figure 3)[13] would be a good candidate for biotin 2 since ketone biotin has a low binding affinity to streptavidin, most likely due to the absence of the ureido nitrogens. Directed evolution of a streptavidin mutant which binds to ketone biotin with very high affinity and specificity, while maintaining a low affinity toward biotin, would be the ultimate goal of the project. Dr. Chinnapen and Dr. Luo in our lab have developed and optimized an IVC protocol for streptavidin evolution against other biotin 2 candidates. Design and creation of streptavidin mutant libraries, selection with these libraries for ketone biotin binding, and the analysis of affinities were performed independently.

Results and Discussion

In order to select streptavidin mutants that bind to a ketone analog of biotin, ketone biotin was covalently linked to DNA encoding streptavidin mutants as shown in Figure 4. In the microdroplet, streptavidin mutants are transcribed and translated *in vitro* as in IVC for BirA. If a streptavidin mutant binds to ketone biotin with a high affinity, the binding will be maintained through the process of breaking emulsions and washing steps. Since streptavidin mutants have His_6 tags by design, pulling down with Ni-NTA beads will separate the genes of mutants with high affinities from those of mutants with low affinities. Thus, PCR amplification of the DNA from Ni-NTA beads will provide the genes for high affinity mutants selectively.

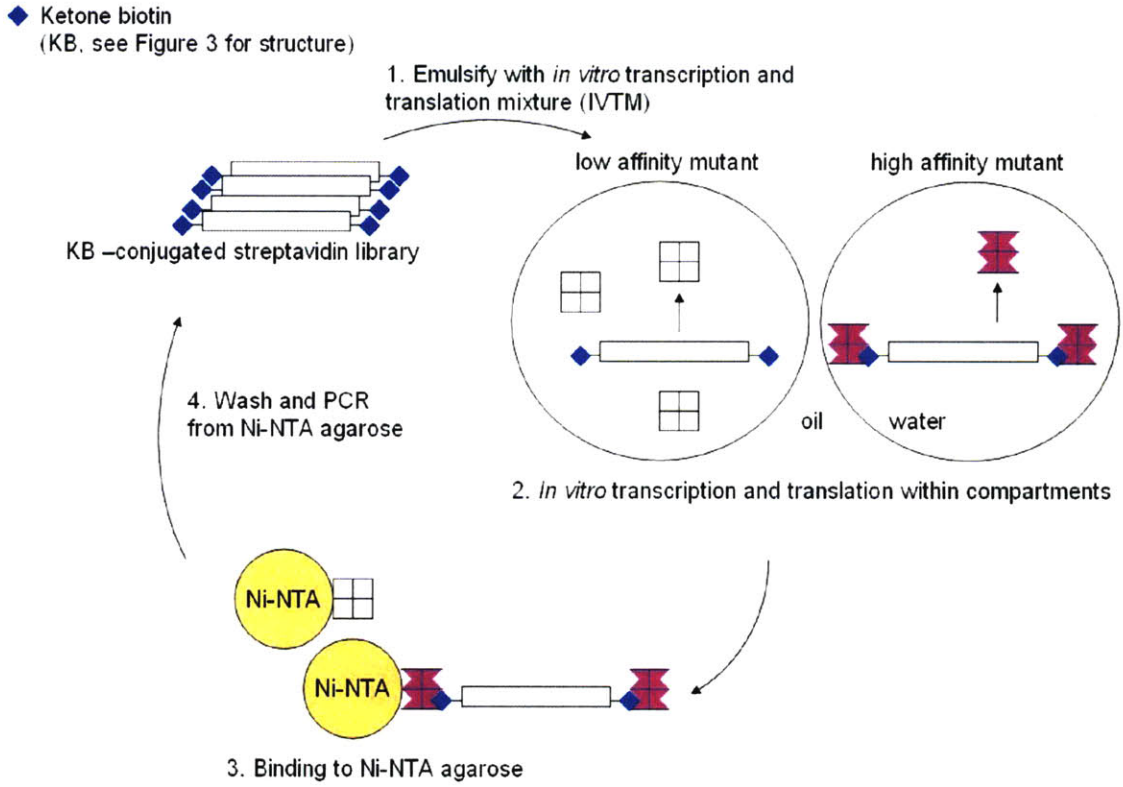


Figure 4 Proposed bead-less IVC selection scheme for streptavidin mutants that bind to ketone biotin[13]. In step 1, a ketone biotin (blue)-conjugated streptavidin gene library is mixed with IVTM and segregated in individual aqueous compartments within a water-in-oil emulsion. Within the compartments the streptavidin mutants are expressed from the gene (step 2). Mutants with high affinities to ketone biotin (right, pink) bind to the ketone biotin at the ends of DNA, whereas mutants with low affinities to ketone biotin (left, grey) do not bind to the DNA. After breaking emulsions, the aqueous solution is incubated with Ni-NTA agarose, which binds to the His_6 tag of each streptavidin molecule (step 3). The Ni-NTA beads are washed and the genes are recovered from the beads through PCR for input into the next round of selection (step 4).

After we started the project, Levy and Ellington reported that streptavidin mutants were evolved to have high affinities toward desthiobiotin (K_d = ~10^{-13} M) using the exactly same selection scheme as our IVC selection[14]. They used real time PCR reactions (RT-PCR) to estimate the amount of bound streptavidin mutant genes from each round. Here, however,

desthiobiotin cannot be used as an orthogonal biotin 2. Compared to ketone biotin which binds to streptavidin about 10^7-fold weaker than biotin, desthiobiotin already strongly interacts with streptavidin (only 100-fold weaker than biotin[1]).

Model selection with wild type (WT) and S45A streptavidin (developed by Dr. Chinnapen and Dr. Luo; I reproduced their selection as shown in Figure 6)

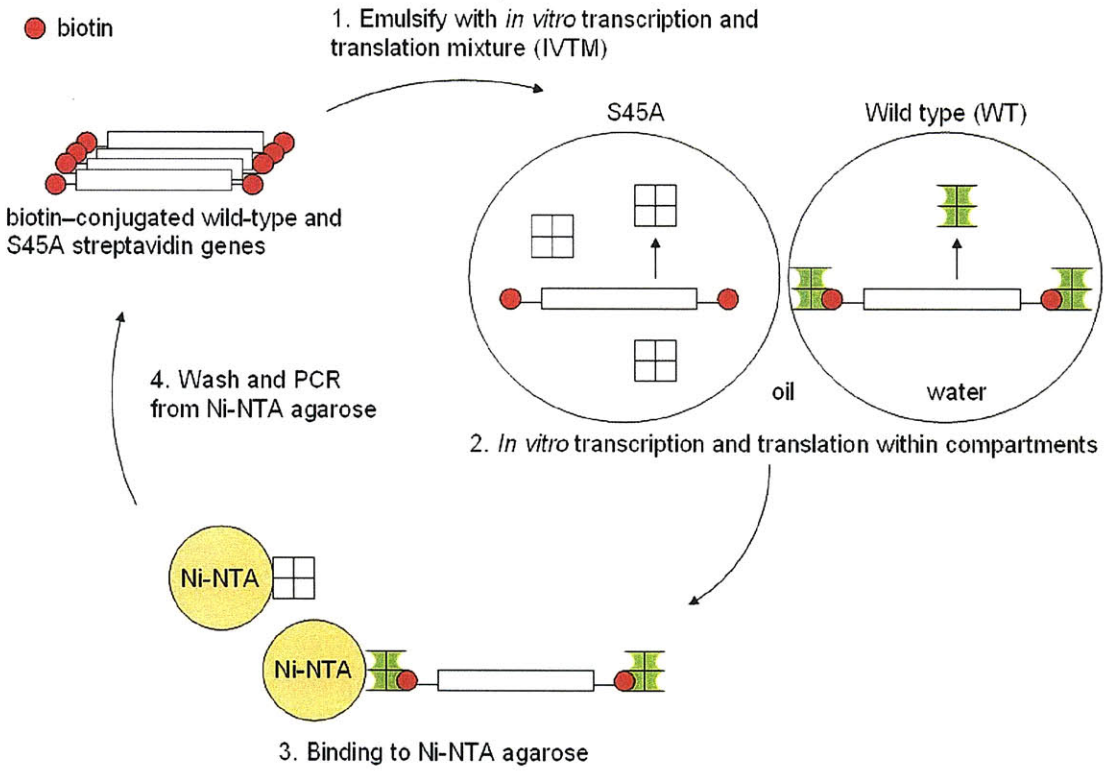


Figure 5 IVC model selection scheme with wild type (WT) and S45A streptavidin. In step 1, a biotin (red)-conjugated WT and S45A streptavidin (50 bp longer than WT by design) genes are mixed with IVTM and segregated in individual aqueous compartments within a water-in-oil emulsion. Within the compartments the streptavidin mutants are expressed from the gene (step 2). WT streptavidin (right, green) binds to the biotin at the ends of DNA very tightly, whereas S45A streptavidin (left, white) binds to the DNA more weakly. After breaking emulsions, the aqueous solution is incubated with Ni-NTA agarose, which binds to the His_6 tag of WT and S45A mutant (step 3). The Ni-NTA beads are washed and the genes are recovered from the beads through PCR and analyzed on an agarose gel (step 4).

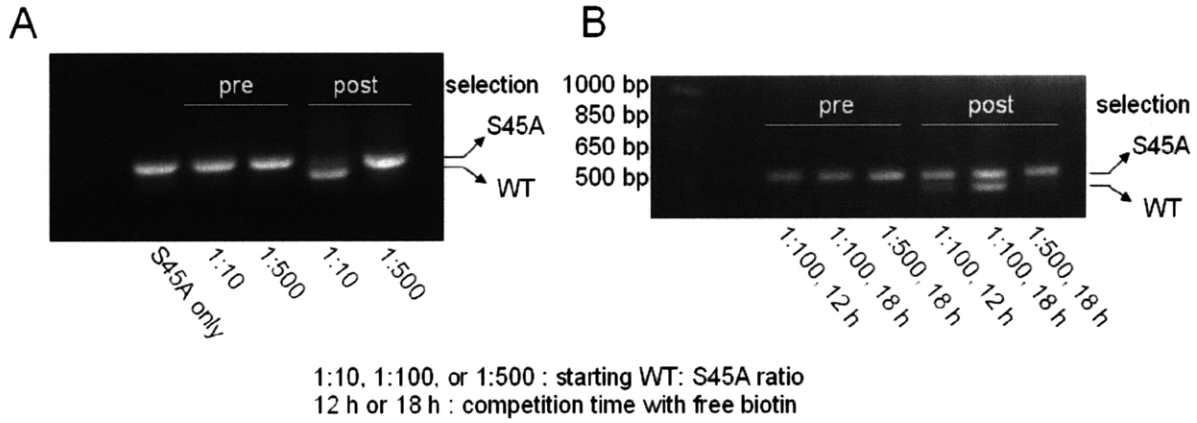


Figure 6 Results of model selections with wild type (WT) and S45A streptavidin. (A) The genes were mixed in 1:10 and 1:500 ratios (WT: S45A). IVTM reactions were incubated at room temperature for 3 h, followed by competition with excess biotin at 37 °C for 11 h. pre: before selection, post: after selection. When WT and S45A were mixed in a 1:10 ratio, WT was enriched over S45A after selection (~30-fold). (B) Model selections with different lengths of competition time were performed. The enrichment factor was higher (~100-fold) when the competition with free biotin was performed longer (18 h). pre: before selection, post: after selection. S45A mutant gene is 50 bp longer than that of WT by design.

Model selection of WT genes from S45A mutant genes of streptavidin was performed to test the validity of the evolution method. As reviewed above, S45A mutant of streptavidin exhibits a 1000-fold higher K_d (1000-fold weaker binding) than that of WT. Extra 50 DNA base pairs were introduced to S45A mutant gene so that we can distinguish WT and S45A after PCR amplification. The genes were then amplified by PCR with biotinylated primers and mixed in a ratio of 1:10, 1:100, or 1: 500 (WT: S45A). This resulting solution was emulsified in mineral oil with IVTM to perform IVC. Streptavidin mutants were then captured by Ni-NTA beads, and the attached mutant genes were amplified by PCR reaction (Figure 5). Since S45A mutant also binds to biotin with picomolar K_d, a competitive elution of the genes from

the beads was performed by incubation with excess biotin at 37 °C. This exploits the faster off-rate of the mutant, which has a half-life of 14 s at 37 °C, compared to that of WT streptavidin which is much longer (6.6 h)[15]. We observed higher enrichment of WT over S45A mutant (up to 100-fold) when competition got longer (Figure 6). However, the enrichment was low when the competition went too long such as for 24 h, probably due to the dissociation of WT. Obtained enrichment factor in this model selection is significant considering that the difference in K_d is 1000-fold, which is only 10-fold higher than the enrichment factor.

Design and construction of a streptavidin mutant library 1

Prior to implementing a selection to evolve a streptavidin mutant that binds ketone biotin with a high affinity, a library of focused streptavidin mutants must be carefully designed and constructed. Based on the crystal structure of streptavidin in complex with biotin, we selected nine positions (Q24, S45, V47, W79, T90, W92, W108, L110, and D128; Figure 7) lining the binding pocket and close to a bound biotin. Two different libraries were generated, namely cis and trans libraries, and mixed in a 1:1 ratio for selection. For the cis library, Ser 45 and Val 47 were chosen due to their hydrogen bond interaction with the cis nitrogen of biotin (Figure 3). For the trans library, Gln 24 and Asp 128 were chosen due to their involvement in the hydrogen bonding to the trans nitrogen of biotin. The rest of the randomized residues were common for the two libraries and responsible for binding the lower thiolane ring of biotin. Thr 90 interacts with the sulfur group, and Trp 79, Trp 92, Trp 108 and Leu 110 play crucial roles in forming the hydrophobic binding pocket. We have chosen these residues in the lower binding pocket to compensate for any changes in the binding mode of ketone biotin compared with that of biotin. Theoretical amino acid diversity of the library is $20^7 = \sim 10^9$.

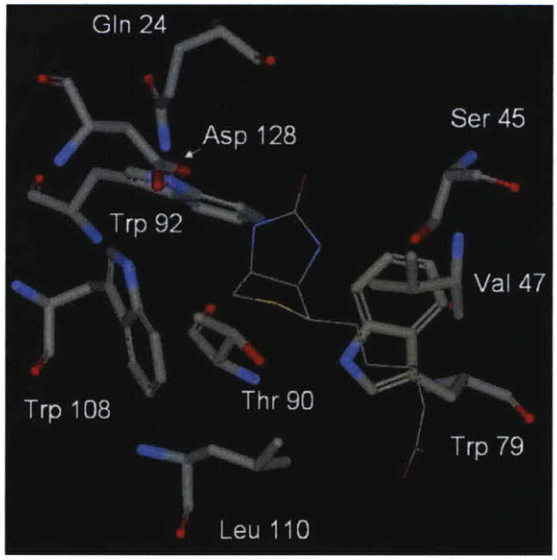


Figure 7 Randomized positions in streptavidin mutant library 1

	Q24	**S45**	**V47**	**W79**	**T90**	**W92**	**W108**	**L110**	**D128**
Rd 0-1	Asn	Ala	Glu	Val	Trp	Thr	Pro	Cys	Asp
Rd 0-2	Asn	Val	Arg	Gly	Arg	Glu	Thr	Thr	Asn
Rd 0-3	Lys	Ser	Val	Leu	Ser	Ser	Leu	Tyr	His
Rd 0-4	Thr	Ser	Val	Glu	Ala	Asn	Ala	Glu	Lys
Rd 0-5	Gly	Ser	Val	Val	Ile	Trp	Val	Tyr	Asp

Table 1 Sequences of random clones from streptavidin mutant library 1

We constructed the DNA library by PCR overlap extension. Synthetic oligonucleotides, overlapping with the region of interest, contained NNK (N = A, T, G, or C; K = T or G)[16] codons at the positions of saturated mutagenesis. The use of the K in the third position excludes two out of three stop codons. After construction of the library, we checked the library by sequencing five clones at random (Table 1). All nine amino acids were highly mutated and we did not observe any contamination by the wild type streptavidin gene.

Selection with the library (1)

The dissociation constant of streptavidin-ketone biotin binding was measured through collaboration with Kulomaa's lab (University of Tampere, Finland). Isothermal titration calorimetry was used to get a K_d value of 8.85 x 10^{-7} M for ketone biotin-streptavidin interaction, which is 10^7 times weaker than biotin binding to streptavidin. The difference in the binding affinity is the same as for S45A and D128A double mutants (K_d =1.22 x 10^{-6} M)[17], suggesting that the library should be focused on the residues involved in the hydrogen bonding with two ureido nitrogens.

An *N*-hydroxysuccinimide ester of ketone biotin was synthesized *in situ* and conjugated to DNA oligonucleotides with amine-modified 5' ends. The conjugate was purified by reverse-phase HPLC, and its mass was confirmed by MALDI-TOF measurement (See the experimental section and appendix for characterization data). Finally, the streptavidin library was PCR amplified with these primers.

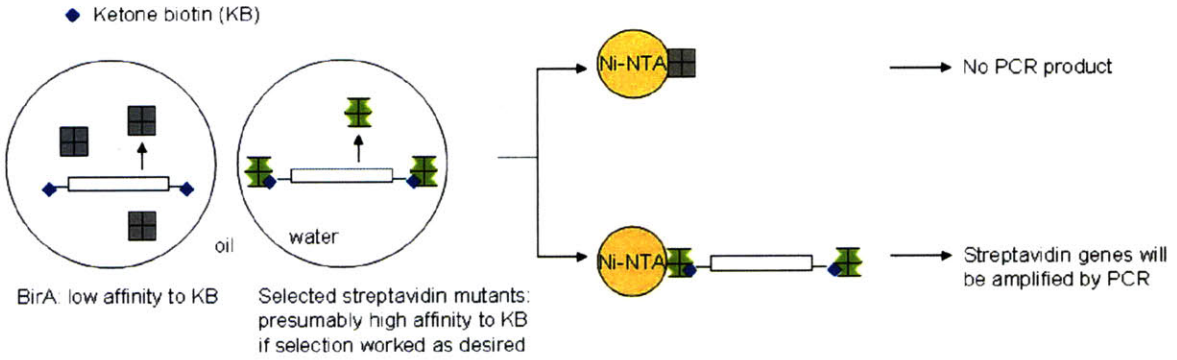
For the first three rounds of selection, the *in vitro* transcription and translation mixture (IVTM) was incubated for 3 h. No competition was performed to include the mutants with very weak affinities to ketone biotin. Sequence analysis of clones from the rounds showed that the trans library was slightly more dominant than the cis library, but it should be confirmed if the selection is directed toward ketone biotin binding mutants as desired.

Functional assay: model selection with BirA

In early rounds of selection, the affinities of selected mutants toward ketone biotin will be very low and therefore hard to determine. Currently available methods for measuring the binding affinity provide reliable results only for those interactions with K_d in a micromolar range or stronger, and in most cases a large amount of protein samples must be prepared.

Therefore, it was hard to see if the binding affinity was increasing over the rounds in the early stages of the selection by these traditional binding assays.

Instead, we performed model selections with the genes from each round against the BirA gene because BirA does not bind ketone biotin with a high affinity (K_M of ketone biotin will be lower than that of biotin toward BirA, which is in micromolar range). The BirA gene was PCR amplified also with ketone biotin conjugated primers. The selected streptavidin mutant genes from each round (round 0 which is before selection, round 1 and 2) and BirA genes were mixed in a 1: 10 ratio (streptavidin mutants from each round: BirA). IVC selection was performed with these gene mixtures (as depicted in Figure 4) to examine whether screened streptavidin mutants have higher ketone biotin binding affinities than BirA, thereby streptavidin mutant genes are enriched from the BirA gene. After pulling down with Ni-NTA beads, the genes were PCR amplified and analyzed on an agarose gel. As shown in Figure 8, we observed slight enrichment of the streptavidin mutant genes from round 1 over the BirA gene. This data indirectly indicates that selected streptavidin mutants from round 1 selection might possess higher binding affinities to ketone biotin than BirA. However, the genes amplified from the round 2 were shorter in length, probably due to the PCR-related artifacts. As briefly discussed in Chapter 3, high numbers of PCR cycles (50 cycles here) with small amount of template DNA sometimes generate chimeric fragments by homologous recombination, and resulting shorter DNA fragments are more readily amplified by PCR than original templates. In general, this kind of failure in IVC selection was major problem for the current protein evolution. It is, therefore, still hard to determine whether the selection is working as desired with this result.



lane	PCR was performed with the gene mixture	Streptavidin mutants from
1	before selection	Round 0
2	before selection	Round 1
3	before selection	Round 2
4	after selection	Round 0
5	after selection	Round 1
6	after selection	Round 2

Figure 8 Functional assays of selected streptavidin mutants from each round, via model selection with BirA. Streptavidin genes and BirA genes were PCR amplified with ketone biotin primers and mixed in a 1:10 ratio. IVC selection was performed with these gene mixtures (as depicted in Figure 4). After pulling down with Ni-NTA beads, the genes were again PCR amplified and analyzed on an agarose gel. After one round of selection, enrichment of streptavidin mutants was observed. After the second round of selection, the results were inconclusive due to contamination in the PCR reaction.

We also analyzed sequences of several random clones from each round of selection, but the sequence information would be significant only after we have a better assay to analyze the binding affinity of each clone. Characterization of individual clones at each round of selection can be performed by transferring current streptavidin gene fragments to bacterial expression plasmid and expressing the streptavidin mutants in large scale. This, however,

would not provide an entire view of each round of selection, and many clones should be analyzed to get significant information.

We also noticed a problem with library design. The library used for these three rounds of selection has a theoretical diversity of 10^9, which is the same number of compartments we can practically generate. Assuming perfectly even distribution of genes to compartments (one gene per droplet), we will only have a single copy of each streptavidin mutant gene. Practically, some fraction of the library is always lost while we perform the entire process of one round of selection. Since there is only a single copy of each mutant gene, it is highly probable that we lose mutant genes with various binding properties at an early stage of the selection. In order to circumvent any permanent loss of mutants, we should design a library with lower diversity to ensure that we have several copies of each mutant.

Although the current library exhibits high sequence diversity, a significant portion will produce inactive streptavidin mutants due to improper protein folding. Thus, the mutation rate at the active site should be reduced to increase the number of properly folded and therefore functional streptavidin mutants in the library. It should be also noted that a library with an extremely low mutation rate will not give a novel mutant with the desired function, and thus we need to optimize the mutation rate as well.

Design and construction of a streptavidin mutant library 2

A new streptavidin library for ketone biotin included seven positions (N23, L25, S27, Y43, S45, V47 and D128; Figure 9) that are close to the upper ring of biotin. Theoretical amino acid diversity of the library is same as before, $20^7 = \sim 10^9$. We constructed the DNA library by PCR overlap extension as described earlier. Instead of saturation mutagenesis at each position, this time each residue was randomized so that 34% of the sequences in the

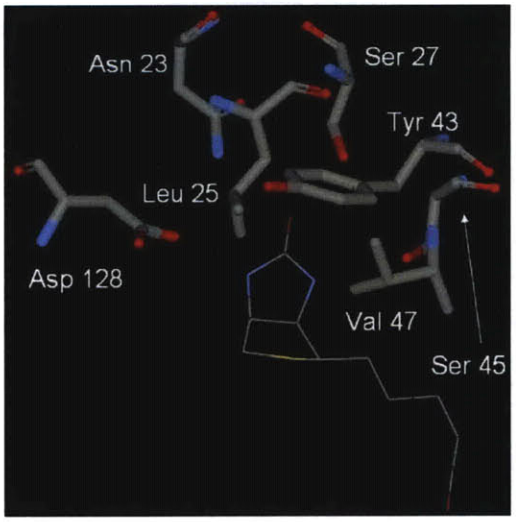


Figure 9 Randomized positions in streptavidin mutant library 2. Residues close to the upper ring of biotin and chosen for randomization were labeled.

	Asn 23	**Leu 25**	**Ser 27**	**Tyr 43**	**Ser 45**	**Val 47**	**Asp 128**
Rd 0-1	Ile	Leu	Thr	Phe	Ser	Ser	Asp
Rd 0-2	Asn	Leu	Pro	Asn	Ser	Gly	Glu
Rd 0-3	Ile	Leu	Ser	Stop	Ser	Val	Asp
Rd 0-4	Asn	Met	Ser	Tyr	Ser	Val	Asp

Table 2 Sequences of random clones from streptavidin mutant library 2

library would represent the wild type residue at a particular position. The new library 2 is therefore more conservative than the previous library 1. Moreover, mutants in the library 2 have His_{10} tags instead of His_6 to ensure their tight binding to Ni-NTA beads. After construction of the library, we checked the library by sequencing four clones at random (Table 2). All positions were mutated as expected from the design and we did not observe any contamination by the wild type streptavidin gene.

Selection with the library (2)

The genes from the new library were prepared as before. For the first three rounds of

selection, the *in vitro* transcription and translation mixture (IVTM) was incubated for 3 h. Competition by incubation with 10 μM ketone biotin was performed for 25 min at round 1 and for 30 min at round 2 and 3. We expected that more mutants in the library 2 will bind to ketone biotin with higher affinities compared to those in the library 1, and thus competition will provide finer discrimination of mutants with high affinities from those with moderate or low affinities.

Functional assay: gel shift assays

Model selections with the BirA gene were used previously to test if there is a steady increase in the binding affinity over the rounds of selection. As discussed above, however, additional IVC selection for functional assay caused consistency problems. Thus, we developed a new gel shift assay to estimate the affinities of newly evolved mutants. One molar equivalent (eq) of biotinylated DNA fragments was incubated with WT streptavidin (biotin: biotin binding site = 1: 4), and we observed shifted bands of streptavidin bound DNA on an agarose gel (Figure 10, lane 1). For comparison, WT streptavidin was treated with 4 eq of free biotin before it is incubated with biotinylated DNAs. In this case, all biotin binding sites should be occupied by free biotin, but we observed a shifted band in addition to non-bound DNA (Figure 10, lane 2). This kind of sub-stoichiometric binding of biotin to streptavidin has been observed when relatively low excess (less than 10 eq) of biotin was used for streptavidin binding[18]. When WT streptavidin was incubated with ketone biotin conjugated DNA fragments, however, no shifted bands were observed due to weak interaction between WT streptavidin and ketone biotin. We expected that, if the selection is working as desired (so that streptavidin mutants, which bind ketone biotin tightly, are selected), we might see some shifted ketone biotin-conjugated DNA fragments over the rounds of selection. However, we

Lane	Streptavidin (SA)	DNA used for gel shift assay	Pre-treatment of free ligand to streptavidin?
1	Wild-type	Biotinylated DNA	No
2	Wild-type	Biotinylated DNA	Yes, biotin
3	Wild-type	Ketone biotin-conjugated DNA	No
4	Wild-type	Ketone biotin-conjugated DNA	Yes, ketone biotin
5	Round 0	Ketone biotin-conjugated DNA	No
6	Round 0	Ketone biotin-conjugated DNA	Yes, ketone biotin
7	Round 1	Ketone biotin-conjugated DNA	No
8	Round 1	Ketone biotin-conjugated DNA	Yes, ketone biotin
9	Round 2	Ketone biotin-conjugated DNA	No
10	Round 2	Ketone biotin-conjugated DNA	Yes, ketone biotin

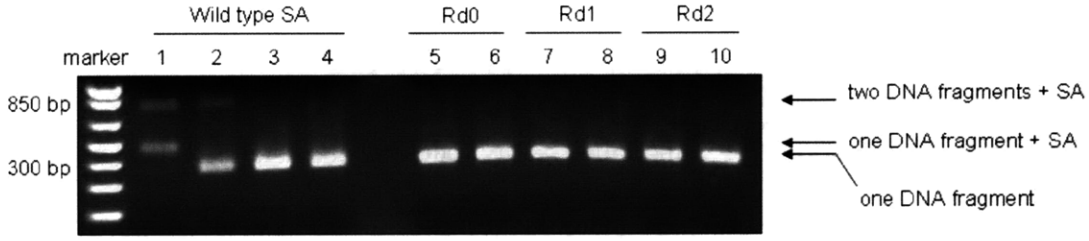


Figure 10 Functional assays on selected streptavidin (SA) mutants from each round of selection, via gel shift. One molar equivalent (eq) of biotinylated DNAs was incubated with wild type SA (biotin: biotin binding site = 1: 4, lane 1). In lane 2, wild type SA was pre-treated with 4 eq of free biotin and then incubated with 1 eq of biotinylated DNA. If the selected SA mutants bind to ketone biotin, we expect to see shifted DNAs as in lane 1. Ketone biotin-conjugated DNAs (KB-DNA, 1 eq) were incubated with wild type SA with (lane 4) or without (lane 3) prior treatment of free ketone biotin (4 eq). In lanes 5, 7, and 9, KB-DNAs (0.1 nmole) were incubated with selected SA mutants from each round *without* prior treatment of free ketone biotin. In lanes 6, 8, and 10, KB-DNAs (0.1 nmole) were incubated with selected SA mutants from each round *with* prior treatment of ketone biotin (0.4 nmole). No evidence of binding to ketone biotin is seen in any of the lanes.

did not observe any shifted ketone biotin-conjugated DNA bands upon the incubation with the streptavidin mutants, which were expressed *in vitro* from each round (Figure 10). The sensitivity of ethidium bromide DNA detection might not be high enough to identify small

amounts of shifted DNAs. Therefore, we also prepared DNA fragments with fluorophores on one end and performed the same experiments in the hope that the fluorescence will offer a higher sensitivity. Unfortunately, none of these attempts gave reliable data (data not shown).

Although we have successfully performed model IVC selection for streptavidin evolution, two IVC selections to evolve a novel streptavidin that binds ketone biotin with newly designed libraries were largely inconclusive. With still improving IVC technology and tools to measure weak protein-small molecule interactions in hand, however, streptavidin evolution against ketone biotin based on our developed assay may be realized in near future.

Conclusions

Here we have attempted to evolve novel streptavidin variants which show high affinities toward ketone biotin using *in vitro* compartmentalization. In a model selection, the 100-fold enrichment of wild type streptavidin was obtained over S45A mutant, which binds biotin 1000-fold weaker. Two different DNA libraries were designed and created. Selections with these two libraries were performed using IVC. However, the protein evolution process could not be confirmed due to the lack of a sensitive method to measure very weak bindings between ketone biotin and streptavidin mutants selected during earlier rounds of selection. To solve this problem of our selection, real time PCR could be used as an alternative approach. Compared to our direct (measuring binding events) but rather qualitative and insensitive functional assays, RT-PCR provides quantitative data on the varying levels of bound mutant genes, therefore offering indirect but clear evidence regarding the selection process.

In our selection scheme, we only have two binding sites (at two ends of the gene) for a streptavidin mutant in each droplet. Other directed evolution platforms usually include more binding sites from five to a million. Multiple binding sites might be beneficial to avoid loss of

weak binders from initial naïve libraries[9], such as the library 1 used in earlier selections.

Due to its tetrameric property, directed evolution of streptavidin by other selection methods was technically challenging. Cell surface displays inevitably include the anchor protein, and the activity of proteins to be selected will be reduced by the presence of the anchor protein. Monomeric streptavidin was reported to have a reversible binding to biotin (K_d $=10^{-7}$ M) because of its incomplete binding pocket[19]. Others recently showed that homotetrameric streptavidin can be displayed on yeast cell surface by coexpression of the native and anchored subunits of streptavidin[20]. Yeast surface display of streptavidin mutants for affinity maturation toward ketone biotin would be worthy of investigation since the large copy numbers of displayed proteins might capture weak binders by the avidity effect.

The IVC technology is in an early stage and should be improved further. Few successful examples have been reported, and most are model selections, not providing any information on novel proteins. However, once technical difficulties are resolved, IVC will be a powerful tool for directed evolution of proteins, enabling high throughput investigation of large libraries.

Experimental

Model selection of wild type and S45A streptavidin mutants (modified from Dr. Chinnapen and Dr. Luo's protcol).

Preparation for making emulsions. 500 μL of mineral oil mix was put in a Nunc CyroTube vial with an 8x2 mm stir bar, and the vial was placed on ice to chill the oil mixture. Biotin-conjugated genes were prepared as a 2.4 nM solution in 10 mM Tris-acetate pH 8. This will give a final concentration of DNA in the IVTM of 200 pM. Two falcon tubes containing water and absolute ethanol separately were prepared to rinse the homogeneizer.

Making DNA/IVTM mixtures. In a 650 μL eppendorf tube, 22 μL of sterile water and 5 μL of 2.4 nM stock solution of singly biotinylated gene construct were mixed. For model selections, DNA mixture was combined to achieve a final concentration of 200 pM in 60 μL. To make 1:10, 1:100, and 1: 500 dilutions, serial dilutions of original stock solution were made. The tube was then placed on ice. IVTM mixture was generated on ice by combining 11.5 μL of lysate, 9.6 μL of reaction mixture, 11.5 μL of amino acids, 0.98 μL of methionine, 5 μL of reconstitution buffer, and 0.9 μL of 1 mM biotin. Immediately prior to forming emulsions, 33 μL of IVTM was added to the 27 μL of gene construct made as above. The reaction mixture was remained on ice before homogenizing.

Forming Emulsions. The tube containing the mineral oil mixture was placed on ice and put on the stirrer at 1,150 rpm. 50 μL of IVTM-DNA mixture were added to the stirring oil and stirred further for 3 min. The oil became cloudy, and the tube was removed from the stirrer, and the stir bar was also removed from the tube. The tip of the homogenizer was placed into the tube and the mixture was homogenized at 11,000 rpm for 3 min. The homogenizer was then removed, and the tube was capped and incubated at 30 °C for overnight. The homogenizer was cleaned with one thorough wash with ethanol, followed by a thorough wash with water. All traces of water in the final wash should be removed before next use.

Break Emulsions. 200 μL of quench buffer was added to emulsions and the mixture was transferred to a clear 1.5 mL eppendorf tube. The mixture was then spun down at 15 °C with 13,000 rpm for 10 min. We usually saw a pellet of emulsions at the bottom of the tube, then an aqueous layer of quench buffer, and a top layer of mineral oil. The layer of mineral oil was removed. 1 mL of pure mineral oil (Sigma) was added to the tube and the pellet was re-suspended by flicking with fingers. The tube was again spun down for 5 min at 15 °C with

13,000 rpm. The top layer was removed again. Until no more pellet is observed and the solution is clear, mineral oil was added and removed as above for three to four more times. Any white interphase was removed. In order to remove residual mineral oil, 1 mL of hexane was added. The mixture was resuspended and spun down for 1 min at 13,000 rpm. The hexane layer was removed, and this hexane washing was repeated for two more times. Residual hexane was evaporated off by air-drying for 10 min.

Binding to Ni-NTA beads. Seventy microliter of 50% slurry of Ni-NTA beads pre-equilibrated in binding buffer was added to a 1.5 mL eppendorf tube. 180 μL of IVTM reaction were added to the Ni-NTA beads, and the rest 20 μL of IVTM reaction were saved for PCR amplification. The beads were than rotated at 4 °C for 2 h and washed with 1 mL of binding buffer. The washing step was repeated once.

Competition of weak binders from beads. 200 μL of competition buffer (100 μM biotin in Ni-NTA binding buffer) were added to the beads. The beads were then incubated at 37 °C on a rotator for desired time (1-20 h). The beads were washed twice with 800 μL binding buffer. 50 μL double distilled H_2O were added to the beads and the mixture were heated >90 °C for 15 min to denature streptavidin mutants.

DNA recovery by PCR from beads. 20 μL of the above slurry containing beads and supernatant were added to a PCR reaction solution on ice containing 10 μL of 10X Pfu buffer, 4 μL of mixture of dNTPs (10 mM), 1 μL of 100 μM primer 1 (PIVB-9; 5'- TAT CCG GAT ATA GTT CC), 1 μL of 100 μM primer 2 (LMB2-9; 5'- GTA AAA CGA CGG CCA GT), 2 μL of Pfu-Turbo DNA polymerase, and 72 μL of sterile water. PCR was performed for the following cycles: a) 94 °C 2 min, b) 94 °C 10 sec, c) 50 °C 30 sec, d) 72 °C 54 sec, e) repeat b) to d) 19 times, f) 72 °C for 7 min, g) 4 °C forever. 2 μL of this PCR product was taken to

the nested PCR reaction solution containing 10 μL of 10X Pfu buffer, 4 μL of mixture of dNTPs (10 mM), 1 μL of 100 μM primer 1 (sacI-outer-R; 5'- GCT TTG TTA CCG GAT CCC GG), 1 μL of 100 μM primer 2 (PVX-rbs-ncoI-F; 5'- TTA AGA AGG AGA TAT ACC ATG G), 2 μL of Pfu-Turbo DNA polymerase, and 72 μL of sterile water. PCR were cycled under the same condition as above. PCR products were analyzed on an agarose gel.

Coupling of ketone biotin to 5'-amine modified oligonucleotides. The ketone biotin-conjugated primers were synthesized as follows: 5 μL of 100 mM ketone biotin solution in dry DMSO, 1 μL of 500 mM *N*-hydroxysuccinimide in dry DMSO, and 1 μL of 500 mM *N, N'*-dicyclohexylcarbodiimide were combined and incubated at room temperature for 12 h. One microliter of this reaction was added to 5 μL of 200 μM 5'-amine modified oligonucleotides (NH-LMB2-1: 5'- CAG GCG CCA TTC GCC ATT, NH-PIVB-1: 5'- GCG TTG ATG CAA TTT CT) and 5 μL of 100 mM $NaHCO_3$ pH 8.3. The reaction was then incubated at room temperature with rotation for 4 h. The crude reaction was run on reverse phase HPLC using a gradient of 6 - 30% acetonitrile in 50 mM ammonium acetate over 30 min with a 1 mL/min flow rate. Chromatogram was recorded at 260 nm. Intact oligonucleotides have retention times of ~7 min (LMB2-1) and ~9 min (PIVB-1), but after conjugation to ketone biotin, the retention time should be increased to ~10 min (ketone biotin conjugated LMB2-1) or ~11 min (ketone biotin conjugated PIVB-1). Approximately 85% yield was observed for both coupling reactions (Figure 11). The conjugated product was collected, lyophilized, and reconstituted in water before use. MALDI-TOF analysis confirmed that the mass of conjugates match the calculated number (Figures 3 and 4 in Appendix): ketone biotin conjugated PIVB-1= 5596.02 (observed), 5595.7 (calculated), ketone biotin conjugated LMB2-1 = 5841.15 (observed),

5839.9 (calculated).

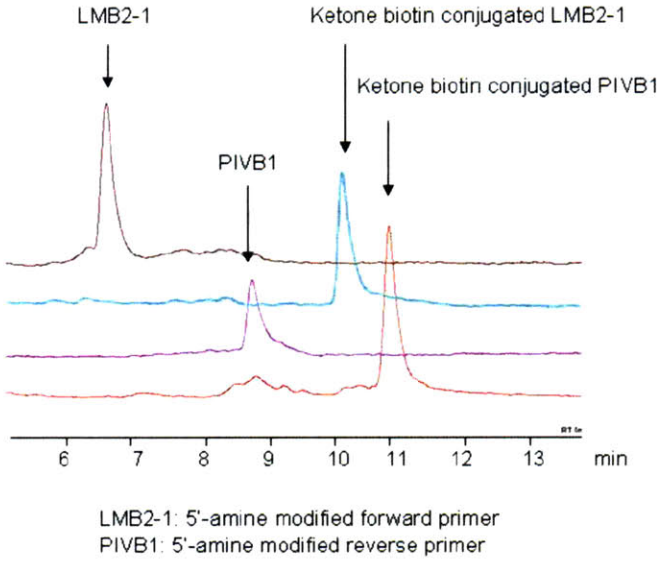


Figure 11 HPLC traces of ketone biotin modified oligonucleotides. The conjugation reaction mixture was analyzed by C18 reverse-phase HPLC using a gradient of 6 - 30% acetonitrile in 50 mM ammonium acetate over 30 min with a 1 mL/min flow rate. Chromatogram was recorded at 260 nm. Intact primers (LMB2-1: forward primer for PCR, PIVB1: reverse primer for PCR) were run on HPLC for comparison. The yield of conjugation was ~85% for both primers. The product peaks were collected, lyophilized, and reconstituted in water before use.

Cloning of a partially-randomized streptavidin library by overlap extension PCR. The first library was synthesized from a 1:1 mixture of the cis and trans libraries generated by Dr. Chinnapen. The second library was synthesized from four fragments by overlap extension PCR as described in Chapter 3. Fragment 1 was amplified from the pIVEX streptavidin template using the primers LMB2-1 (5'- CAG GCG CCA TTC GCC ATT) and the rv23_25_27 (5'- GTA CCA GGT GCC GGT GAT), using 30 reaction cycles as described above. Fragment 2 was similarly amplified using the primers fw23_25_27 (5'- ATC ACC GGC ACC TGG TAC **AAC** CAG **CTG** GGA **TCC** ACC TTC ATC GTT AC, where **A**

represents 70% A and 10% each of the other bases, C represents 70% C and 10% each of the other bases, and so forth; randomizes Asn 23, Leu 25, and Ser 27) and rv43_45_47 (5'- GGT ACC GGT CAG AGC ACC). Fragment 3 was similarly amplified with the primers fw43_45_47 (5'- GGT GCT CTG ACC GGT ACC TAC GAA TCC GCT GTT GGT AAC GCT GAA TC; randomizes Tyr 43, Ser 45, and Val 47) and rv128 (5'- GTG ACC AAC CAG GGT GGA T). Fragment 4 was similarly amplified with the primers fw128 (5'- ATC CAC CCT GGT TGG TCA CGA CAC CTT CAC CAA AGC; randomizes Asp 128) and PIVB1 (5'- GCG TTG ATG CAA TTT CT). The final library was PCR amplified with the ketone biotin-conjugated primers so that every member of the library has two ketone biotin at the ends.

Gel shift assays. DNA fragments (~300 bp) with biotin or ketone biotin at only one end were prepared by PCR. In an eppendorf tube, 0.1 nmole of DNA fragments, 5.55 μL of 18 μM wild type streptavidin, and 3.4 μL of 1x PBS were added. Some wild type streptavidin was treated with 0.4 nmole of free ligands (biotin or ketone biotin) before incubation with DNAs. In order to analyze the binding of streptavidin mutants from each round, IVTM reactions with genes from round 0, 1, and 2 were incubated at room temperature for 24 h to get large amount of proteins. After breaking emulsions, addition of excess ketone biotin was omitted, and the aqueous solution was purified with 70 μL of Ni-NTA agarose. The reaction mixture was then incubated with 0.1 nmole of DNA fragments and analyzed immediately on an agarose gel.

References

1 Green, N.M., Avidin. *Adv Protein Chem* 29, 85-133 (1975).

2 Wilchek, M. & Bayer, E.A., The avidin-biotin complex in bioanalytical applications.

Anal. Biochem. 171, 1-32 (1988).

[3] Wilchek, M. & Bayer, E.A., Introduction to avidin-biotin technology. *Methods Enzymol.* 184, 5-13 (1990).

[4] Laitinen, O.H., Hytonen, V.P., Nordlund, H.R., & Kulomaa, M.S., Genetically engineered avidins and streptavidins *Cell. Mol. Life. Sci.* 63, 2992-3017 (2006).

[5] Reznik, G.O., Vajda, S., Sano, T., & Cantor, C.R., A streptavidin mutant with altered ligand-binding specificity. *Proc Natl Acad Sci USA* 95, 13525-13530 (1998).

[6] Dixon, R.W. *et al.*, Theoretical and Experimental Studies of Biotin Analogues That Bind Almost as Tightly to Streptavidin as Biotin. *J. Org. Chem.* 67, 1827-1837 (2002).

[7] Aslan, F.M., Yu, Y., Mohr, S.C., & Cantor, C.R., Engineered single-chain dimeric streptavidins with an unexpected strong preference for biotin-4-fluorescein. *Proc Natl Acad Sci USA* 102 (8507-8512) (2005).

[8] Voss, S. & Skerra, A., Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng.* 10, 975-982 (1997).

[9] Leemhuis, H., Stein, V., Griffiths, A.D., & Hollfelder, F., New genotype–phenotype linkages for directed evolution of functional proteins. *Curr. Opin. Struct. Biol.* 15 (472-478) (2005).

[10] Sepp, A. & Choo, Y., Cell-free selection of zinc finger DNA-binding proteins using in vitro compartmentalization. *J. Mol. Biol.* 354, 212-219 (2005).

[11] Bertschinger, J., Grabulovski, D., & Neri, D., Selection of single domain binding proteins by covalent DNA display. *Protein Eng. Des. Sel.* 20, 57-68 (2007).

[12] 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.* 371, 1238-1248 (2007).

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

[14] Levy, M. & Ellington, A.D., Directed Evolution of Streptavidin Variants Using In Vitro Compartmentalization. *Chem Biol* 15, 979-989 (2008).

[15] Hyre, D.E., Trong, I.L., Freitag, S., Stenkamp, R., & Stayton, P.S., Ser45 plays an important role in managing both the equilibrium and transition state energetics of the streptavidin-biotin system. *Protein Sci.* 9, 878-885 (2000).

[16] Hughes, M.D., Nagel, D.A., Santos, A.F., Sutherland, A.J., & Hine, A.V., Removing the redundancy from randomised gene libraries *J. Mol. Biol.* 331, 973-979 (2003).

[17] Hyre, D.E. *et al.*, Cooperative hydrogen bond interactions in the streptavidin biotin system. *Protein Sci.* 15, 459-467 (2006).

[18] Freitag, S., Trong, I.L., Klumb, L., Stayton, P.S., & Stenkamp, R., Structural studies of the streptavidin binding loop. *Protein Sci.* 6, 1157-1166 (1997).

[19] Wu, S.-C. & Wong, S.-L., Engineering Soluble Monomeric Streptavidin with Reversible Biotin Binding Capability. *J. Biol. Chem.* 280, 23225-23231 (2005).

[20] Furukawa, H., Tanino, T., Fukuda, H., & Kondo, A., Development of Novel Yeast Cell Surface Display System for Homo-oligomeric Protein by Coexpression of Native and Anchored Subunits. *Biotechnol. Prog.* 22, 994-997 (2006).

Chapter 5. An Engineered Aryl Azide Ligase for Site-Specific Mapping of Protein-Protein Interactions via Photocrosslinking

The majority of this chapter has been published in: Baruah, H. Puthenveetil, S., Choi, Y. A., Shah, S., and Ting, A. Y. An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photocrosslinking. *Angew. Chem. Int. Ed.* 47, 7018-7021 (2008). My specific contribution was as follows: I synthesized and characterized the aryl azide probe used in this study.

Introduction

Protein-protein interactions (PPI) govern a large number of cellular events[1]. Among thousands of protein-protein interactions in a living organism, currently available techniques have elucidated primarily the strong interactions with $K_d < \sim 10^{-6}$ M[2]. Understanding weak or transient protein-protein interactions ($K_d > 10^{-4}$ M) is also extremely important to fully elucidate the details of biological pathways[2,3]. Non-invasive detection of these weak interactions in cells has been a challenging task, due to the limitations of detection technologies. For example, many protein-fragment complementation-based assays (PCAs), including the yeast two-hybrid system[4], split-Green Fluorescent Protein[5], split-β-lactamase[6], and split-luciferase[7,8], and FRET (Fluorescence Resonance Energy Transfer[9]), are limited by low sensitivity, narrow dynamic range, or the time for chromophore formation, which is longer than the lifetime of weak protein-protein interactions[10,11]. Recently, Regan and coworkers used a split GFP system to detect weak interaction with a K_d of ~1 mM[11]. However, it was irreversible and did not allow any temporal control. Detection methods which require cell lysis, such as immunoprecipitation, cannot detect weak or transient interactions as well because the protein interaction must survive lysis, dilution, and several incubation and washing steps in order to be detected.

Photoaffinity (PA) probes, containing photoactivatable moieties such as aryl azide and benzophenone groups, have been successfully used to detect transient protein-protein interactions, largely *in vitro*[12]. The small size of PA probes introduces minimal steric bulk, which reduces the rate of false negatives. Proteins labeled with PA probes usually retain the ability to bind to their interaction partners, due to the small size of the PA probes (in contrast to protein reporter based methodologies, which frequently give false negatives due to steric

hindrance). Activated probes form covalent bonds with interacting protein partners in as little as nanoseconds[13], enabling detection of transient interactions. Purification of the crosslinked complexes and analysis by mass-spectrometry, Western blotting, or other techniques reveal the identities of the interaction partners of the protein of interest. Moreover, only one protein partner needs to be labeled (the other can be endogenous unmodified proteins) and thus false negatives and positives are both minimized. Finally, photocrosslinking is compatible with all cell types and all subcellular compartments, in contrast to yeast two-hybrid, for instance[14].

Despite the advantages offered by PA probes, their use in the live cell context, where most relevant protein-protein interactions take place, has been limited by lack of suitable ways to label cellular proteins with PA probes[15]. The best current method for photocrosslinker introduction is unnatural amino acid mutagenesis[16-22]. The non-position specific form of this technique[21, 22], introduces high background, reduced signals, and potential interference with protein functions. Position-specific unnatural amino acid mutagenesis[16-20] is extremely powerful, but its mammalian cell applications are often restricted by the prevalence of natural amber codons[23], low suppression efficiencies[24], and the generation of prematurely truncated protein products, which can sometimes produce dominant negative effects[25]. Methods that enable the targeting of PA probes to specific proteins in the cellular context would greatly expand our working tools for elucidating signal transduction pathways.

As a first step towards developing intracellular PA probe labeling methodologies as proposed in Figure 1, the development and characterization of an aryl azide photoaffinity probe ligase will be discussed in this chapter. This chapter will focus on how we rationally designed the protein ligase that specifically introduces our aryl azide photoaffinity probe to a substrate polypeptide. Dr. Hemanta Baruah found the first mutant of lipoic acid ligase (LplA)

that takes an aryl azide probe and subsequently obtained a further modified ligase with a better performance. Dr. Sujiet Puthenveetil performed photocrosslinking assays based on my preliminary crosslinking experiments. Dr. Samit Shah provided mammalian expression plasmids for this project. A significant portion of this chapter has been published on Angewandte Chemie International Edition[26].

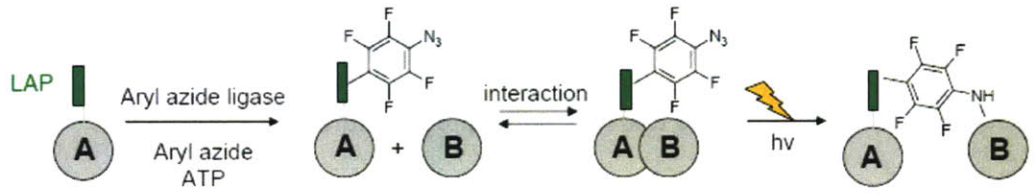


Figure 1 Scheme illustrating site-specific aryl azide ligation to LAP fusion protein, followed by photocrosslinking to interacting protein. LAP is the LplA acceptor peptide.

Results and discussion

Design, synthesis, and incorporation of aryl azide by wild type and mutants of LplA

We first investigated the substrate selectivity of LplA with several structural analogs of lipoic acid. Lipoic acid is a carboxylic acid featuring a hydrophobic dithiolane ring. This dithiolane ring is bigger than cyclopentane because of the size and electron repulsion of sulfur atoms. Thus, we rationalized that LplA may accept a phenyl ring instead of the dithiolane ring in its substrate pocket. Although lipoic acid has a stereogenic center, the active site of LplA looks wide enough to tolerate the substrate diversity. Based on preliminary screening results (data not shown), we synthesized a fluorinated aryl azide substrate **1** by simple coupling of 5-aminopentanoic acid and *N*-hydroxysuccinimide ester of 4-azido-2, 3, 5, 6-tetrafluorobenzoic acid (Figure 2). The design of the aryl azide analog as a substrate of LplA was rather ambitious at this point since we installed an azide substitute on the phenyl group. The ability

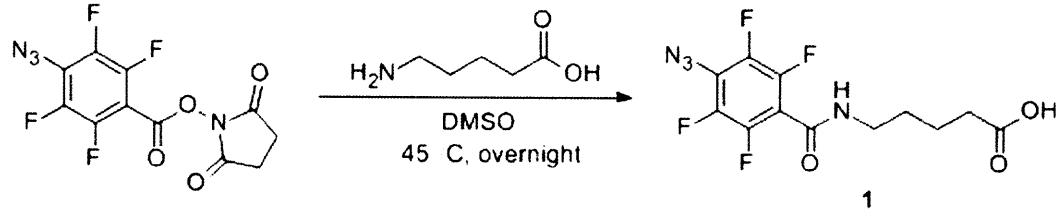


Figure 2 Synthetic scheme for aryl azide probe **1**

of wild type LplA to utilize probe **1** as a substrate was subsequently tested. No product conjugate was detected in the reaction with the probe. Further engineering of the active site was necessary to make LplA to take this new substrate structure.

Engineering of LplA mutants to use aryl azide probe 1 (Dr. Baruah's work). Dr. Baruah planned to mutate a series of amino acids that locate in the LplA active site to increase the volume of binding pocket of the enzyme. The co-crystal structure of *E. coli* LplA with lipoic acid has been reported[27], but structural information, providing more insights for lipoylation by LplA, was obtained from lipoyl-AMP complex of LplA of *Thermoplasma acidophilum* (Ta)[28] as we mentioned in Chapter 1. After examining the crystal structure, he chose nine amino acid side chains (L18, D21, Y39, R72, Y73, T74, H81, L159, and H161) which are located within 7.5 Å of the dithiolane ring of lipoic acid. He then generated individual alanine mutants of *E. coli* LplA (L17, E20, W37, R70, S71, S72, H79, F147, and H149) to decrease side chain sizes of these residues (Figure 3). HPLC analysis of incorporation reactions onto E2p by these LplA mutants showed that W37A, E20A, and S71A mutants could accept probe **1** to a small degree. Under identical reaction conditions, W37A was the best ligase for probe **1** among them.

Since W37 was the most important residue for recognizing probe **1**, he produced more W37 LplA mutants, and found that W37V LplA exhibited the higher probe **1** ligation activity than W37A (Figure 4). This result indicates that the size of active site is not the only factor

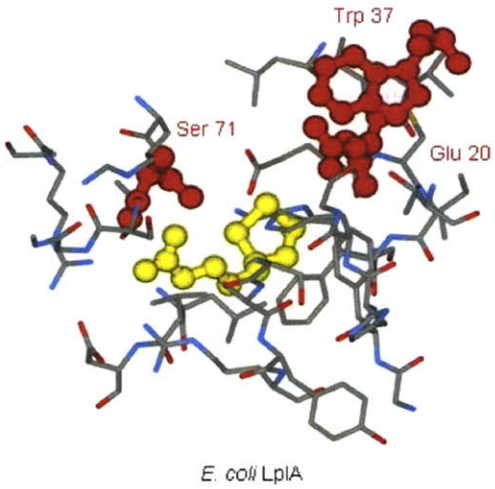


Figure 3 The active site of *E. coli* LplA. Lipoic acid is shown in yellow. With single alanine mutation on Glu 20, Trp 37, or Ser 71 (shown in red), LplA could ligate aryl azide **1** to E2p to a small degree, and W37A showed the highest ligation efficiency. Further engineering on Trp 37 showed that W37V LplA exhibits the higher ligation activity than any other mutants generated for aryl azide **1**.

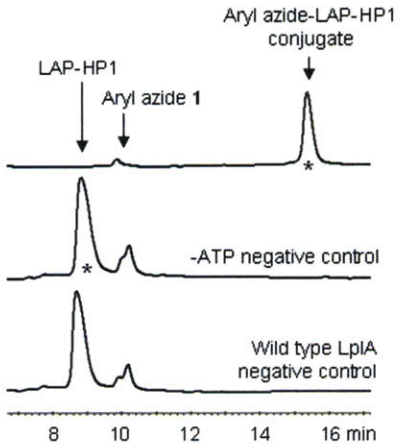


Figure 4 HPLC assay of W37V-catalyzed ligation of aryl azide **1** onto LAP fused HP1 protein (LAP-HP1). When ATP is omitted or wild type LplA is used instead of W37V, no product was observed. The two starred peaks were collected for mass-spectrometry analysis. The aryl azide-LAP-HP1 conjugate has the calculated mass for one molecule of aryl azide **1** plus one LAP-HP1 protein, minus water. This figure was modified from the reference 26.

that affects the catalytic efficiency of LplA. He determined a k_{cat} of 0.31 ± 0.04 s^{-1} for the ligation (the k_{cat} of lipoic acid incorporation onto E2p by wild type LplA is 0.25 s^{-1})[29], which is greater than that of any other LplA-catalyzed ligation of unnatural probes developed so far. However, K_M was 80 μM or less (data not shown), which is considerably greater than the K_M of wild type LplA for lipoic acid (1.7 μM[30] or 4.5 μM[27]). Finally, he showed that aryl azide **1**-ligated LAP-HP1 protein has the calculated mass for one molecule of aryl azide **1** plus one LAP-HP1 protein, minus water. Both the aryl azide moiety and the protein substrate are intact after the reaction by W37V LplA.

Specificity of aryl azide ligation by W37V LplA (Dr. Baruah's work). He then tested if the W37V-catalyzed ligation of aryl azide shows the same specificity against the substrate protein as the lipoylation. By virtue of the azide moiety, aryl azide-ligated proteins can be detected by either Staudinger reaction[31] or [3+2] cycloaddition[32]. After labeling the total lysate from HEK cells expressing a LAP fusion to cyan fluorescent protein (LAP-CFP[29]) with W37V LplA and aryl azide **1**, only LAP-CFP was subsequently modified with phosphine FLAG (via Staudinger reaction, Figure 5A). In addition, aryl azide **1** could be modified with cyclooctyne-Cy3 conjugate[29] (via [3+2] cycloaddition) and therefore used for cell surface labeling in HeLa cells expressing extracellular LAP fusion (Figure 5B). He also found that non-LAP proteins such as FRB-AP, FKBP-AP, and BCCP-87 are not labeled with aryl azide **1** by performing ligation reactions followed by mass measurement.

Application of aryl azide ligase for photocrosslinking proteins (Dr. Puthenveetil's work). As a proof of principle experiment, we attempted to detect a known protein-protein interaction (PPI) within a complex mixture. Model PPI complexes such as CFP dimer, HP1 protein with

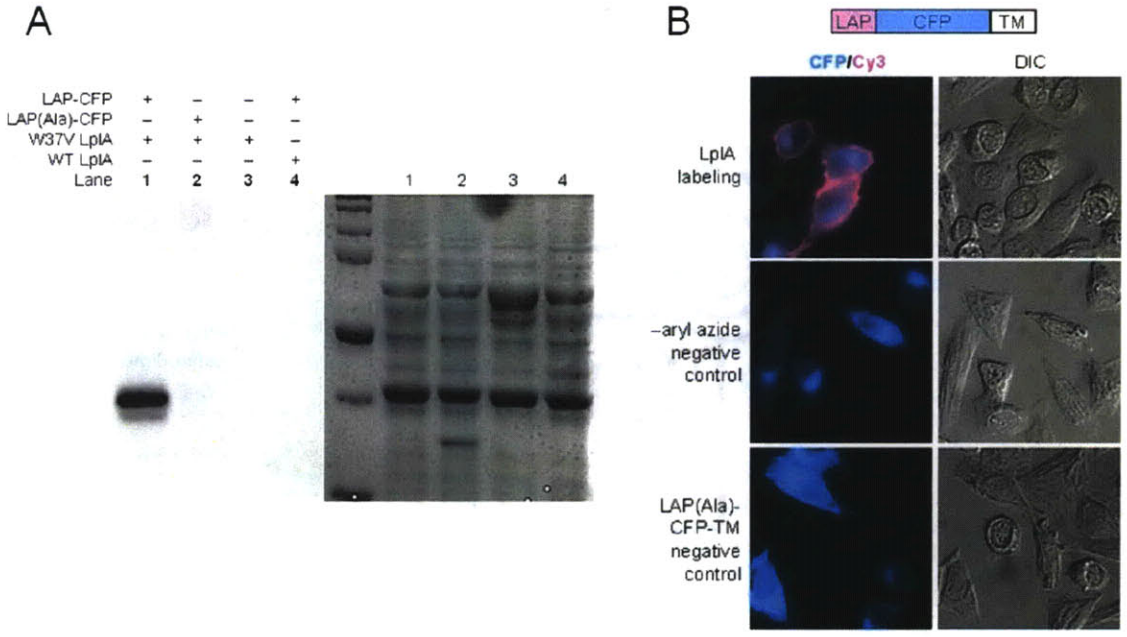


Figure 5 Specificity of aryl azide ligation. (A) HEK cells expressing LAP-CFP were lysed, and the lysate was labeled with aryl azide **1** *in vitro*. The ligated aryl azide was detected by Staudinger reaction with phosphine FLAG. Blotting with anti-FLAG antibody is shown to the left of the Coomassie-stained gel. Negative controls are shown with an alanine mutation in LAP (lane 2), wild type LplA instead of W37V (lane 4), and untransfected cells (lane 3). (B) HeLa cells expressing LAP-CFP-TM were labeled with W37V LplA and aryl azide **1**, followed by cyclooctyne-Cy3 conjugate, to selectively derivatize the aryl azide. Imaging shows pink Cy3 staining on the membranes of transfected CFP-positive cells. No labeling was observed when aryl azide **1** was omitted (middle row) or when LAP-CFP-TM was replaced by an alanine mutant (bottom row). This figure was modified from the reference 26.

His_6 tag and anti-His_6 antibody, and a FRB-FKBP-rapamycin complex were tested for demonstration of the use of the developed aryl azide ligase (W37V LplA). Photocrosslinking of CFP dimer was highly inefficient and only possible using high concentration of CFP. The interaction between His_6 tag and anti-His_6 antibody was too weak to give a strong signal of crosslinking by western blot assays. FKBP is known to interact with FRB only in the presence

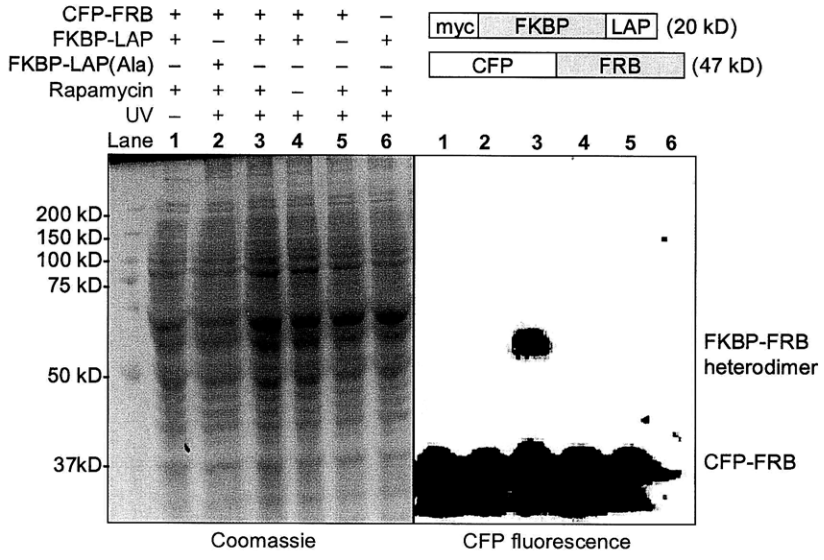


Figure 6 Aryl azide-mediated photocrosslinking of FKBP and FRB in mammalian cell lysate. The aryl azide **1** ligation was performed *in vitro* using recombinant W37V LplA after lysis of HEK cells separately expressing FKBP-LAP and CFP-FRB. The crosslinked heterodimer was detected by CFP imaging only in the presence of rapamycin (lane 3, right gel). This figure was modified from the reference 26.

of the small molecule rapamycin[33,34]. We finally chose this complex because 1) its interaction is strong (FKBP·rapamycin + FRB ↔ FKBP·rapamycin·FRB: K_d = 12 nM, FRB·rapamycin + FKBP↔ FKBP·rapamycin·FRB: K_d = ~100 fM)[33] and 2) the crystal structure of the complex is reported, which will allow us to successfully locate the LAP sequence on the target protein. We prepared FKBP-LAP (and Ala mutant of the lysine lipoylation site as a negative control) and FRB-LAP constructs (and also Ala mutant), and native FKBP and FRB proteins without any targeting sequence. FKBP-LAP and FRB-LAP proteins were separately incubated with W37V LplA, ATP, probe **1**, and Mg^{2+}, and the aryl azide incorporation was confirmed by the HPLC and mass spectrometry readout (data not shown).

In order to optimize photocrosslinking reactions, various irradiation conditions have

been applied. Major problem here was a background signal, probably caused by irradiation-mediated random crosslinking of a highly stable FKBP/FRB complex. Even when the protein was not labeled with probe **1**, the band corresponding to heterodimer of FKBP and FRB was seen under the mildest irradiation (data not shown). Another problem was very small difference in molecular weights between homodimers and heterodimers of FKBP and FRB. We could solve the problem by fusing either FKBP or FRB with other proteins such as BirA (Here BirA was added to make a big difference in size. It does not play any role as an enzyme.) or more simply CFP. Preliminary experiments with FRB-BirA and FKBP-LAP looked promising (data not shown), and Dr. Puthenveetil further purified these proteins to get clearer heterodimer bands. Finally, Dr. Puthenveetil performed the same experiments in HEK cells separately expressing a FKBP-LAP and a CFP-FRB fusion. CFP fusion to FRB not only made the signal more distinguishable on western blots, but also improved its expression in mammalian cells. The lysates were combined and labeled with aryl azide **1**, ATP, and W37V LplA, prior to photocrosslinking with a 300-360 nm UV light for 4 min in the presence or absence of rapamycin. After separation on SDS-PAGE gels, CFP-FRB proteins were detected by in-gel CFP fluorescence (Figure 6). Only in the presence of rapamycin, with UV light applied, and LAP tag intact, they observed a higher molecular weight band corresponding to the covalent FKBP-FRB heterodimer.

Conclusions

We have demonstrated that *E. coli* LplA can be engineered to ligate unnatural aryl azide to its peptide substrate LAP while not losing the high sequence-specificity of the enzyme. LplA mutants could accept aryl azide with only single residue mutation. This feature of LplA

is unique; although homologous to LplA, BirA, by contrast, did not show the same malleability. It is probably due to the structure of lipoic acid, which does not have any hydrogen bonding donor or acceptor and therefore only relies on the van der waals type interaction for its binding to the enzyme. The ease of engineering as well as the hydrophobic nature of the active site makes LplA a great enzyme for small molecule labeling onto target proteins since most fluorophores and photoaffinity labels are hydrophobic. One drawback of the methodology is that it is hard to pinpoint the exact interaction site due to the size of LAP. In this case, unnatural amino acid incorporation methodology is superior to our method. Also, if LAP is not located properly, the interaction partner could not be captured (false-negative) even though it is interacting with the LAP-fusion protein.

It is also intuitively interesting to design a small molecule probe that can be used by LplA. Investigating the small molecule specificity will give us additional insights on the enzyme reaction, which might not be obtainable from the static crystal structure since one small molecule should go through many conformational changes to be ligated. Compared to biotin analogs, lipoic acid analogs can be relatively easily synthesized. Simple synthesis of small molecule probes is always favorable, allowing others to apply the methodology. For example, ketone biotin is not widely used due to the difficulty of synthesis although the methodology is extremely powerful.

Next goal of the project would be aryl azide labeling of proteins inside living cells. Although it is not demonstrated with the aryl azide ligase yet, others in the lab have shown intracellular proteins could be labeled with unnatural probes using co-expressed LplA. Currently, we do not know if aryl azide **1** is membrane permeable. Membrane permeability might be improved by esterification of the probe, increasing its hydrophobicity. If achieved, it

will allow identifying many potential endogenous interaction partners of a protein of interest.

Experimental

General synthetic methods

Reagents were purchased from the described vendors and used without further purification. Analytical thin layer chromatography was performed using 0.25 mm silica gel 60 F_{254} plates. Flash column chromatography was carried out using silica gel (ICN SiliTech 32-63D). Mass spectra were recorded on an Applied Biosystems 200 QTRAP Mass Spectrometer using electrospray ionization. High resolution Mass Spectra (HR-MS) were obtained at the Mass Spectrometry Facility at MIT (Cambridge, MA). HPLC was performed on a Varian Prostar Instrument equipped with an autosampler, using a reverse-phase 250 x 4.6 mm Microsorb-MV 100 C18 column. Chromatograms were recorded at 210 nm unless otherwise noted. ^{1}H NMR and ^{19}F NMR spectra were recorded on a Varian Mercury 300 MHz. Chemical shifts are reported in delta (δ) units, parts per million (ppm), and referenced to the residual solvent peak. Coupling constants (J) are reported in hertz (Hz). The following abbreviations for multiplets are used: s, singlet; dt, doublet of triplets; t, triplet; m, multiplet. Probes were stored as 100 mM stock solutions at -20 °C in DMSO.

Synthesis of aryl azide analog 1

To a solution of 5-aminovaleric acid (Alfa Aesar, 12 mg, 100 μmole, 1.3 eq) in dry DMSO (250 μL), N-Succinimidyl 4-azido-2, 3, 5, 6-tetrafulorobenzoate (as ATFB, SE from Invitrogen, 25 mg, 75 μmole, 1 eq) was added. The reaction was allowed to proceed overnight at 45 °C in dark. The mixture was then acidified with 250 μL of 0.5 N hydrochloric acid and

extracted with 400 μL of ethyl acetate three times. The combined extracts were evaporated *in vacuo* and purified by silica chromatography using 2:1 hexane-ethyl acetate to yield 5-(4-azido-2,3,5,6-tetrafluorobenzamido) pentanoic acid **1** as a white powder. Further recrystallization in hexane-ethyl acetate provided **1** as a white needle crystal (3 mg, 12%). TLC: Rf = 0.3 (1: 1 hexane-ethyl acetate) ^{1}H-NMR (300MHz, $(CD_3)_2SO$): δ 12.03 (s, 1H), 8.89 (t, 1H, J = 5.4), 3.23 (q, 2H, J = 6.3), 2.22 (t, 2H, J = 7.2), 1.497 (m, 4H). ^{19}F-NMR (300MHz, $(CD_3)_2SO$): δ -143.69 (m), δ -151.95 (m). HR-MS m/z: (neg) $[M-H]^-$ calculated: 333.0616, observed: 333.0622.

Tests of various analog ligations to E2p with wild type LplA

To see if a structural analog is incorporated to E2p by wild type LplA, reactions were assembled as follows: 1 μM LplA, 100 μM E2p, 500 μM probe **1**, 1 mM ATP, and 4 mM magnesium acetate in 25 mM sodium phosphate, pH 7.0. For the negative controls, ATP was omitted from the reaction. LplA and E2p proteins were expressed and purified as previously described.[29] Ligation reactions were incubated at 30 °C for overnight, and then quenched with 100 mM EDTA (final concentration). Reactions were analyzed by C18 reverse-phase HPLC using a gradient of 30 - 45% acetonitrile in water with 0.1% trifluoroacetic acid over 20 minutes with a 1 mL/minute flow rate. E2p had a retention time of 12 min, but after conjugation to a probe, the retention time should be increased ~16 min.

Aryl azide ligation to FRB-LAP with W37V LplA

Reactions were assembled as follows: 0.5 μM W37V LplA, 100 μM FRB-LAP, 250 μM aryl azide **1**, 1 mM ATP, and 4 mM magnesium acetate in 25 mM sodium phosphate, pH

7.0. For the negative controls, ATP was omitted from the reaction. Ligation reactions were incubated at 30 °C for overnight, and then quenched with 100 mM EDTA (final concentration). Reactions were analyzed by C18 reverse-phase HPLC using a gradient of 30 - 65% acetonitrile in water with 0.1% trifluoroacetic acid over 20 minutes with a 1 mL/minute flow rate. FRB-LAP had a retention time of 14 min, but after conjugation to a probe, the retention time increased to 14.4 min.

Cloning of FRB-LAP and FKBP-LAP for bacterial expression

To clone FRB-LAP in pET-21b, we first amplified the FRB gene using the primers FRBfw (5'- GGA TCC GTA TCC GTA CGA CGT AC) and LAPrv (5'- GAA TTC CGG TAC TTC CAG AAC TAC TTT GTC GGT TTC GAT TTC AAC CAG TAC TTC GTC TTT GCC CGA GCC CGA GGT C) to introduce an N-terminal BamHI site and a C-terminal 10-amino acid linker (GSGSTSGSGK) followed by LAP (DEVLVEIETDKVVLEVPASADG) followed by an EcoRI site. The PCR product was digested with BamHI and EcoRI enzymes and ligated into similarly-digested pET-21b vector to obtain the FRB-LAP-pET-21b plasmid.

To clone FKBP-LAP in pET-21b, we first amplified the FRB gene using the primers FKBPfw (5'- GGA TCC GGA ACA AAA ACT TAT TTC TGA AGA AG) and LAPrv to introduce an N-terminal BamHI site and a C-terminal 10-amino acid linker (GSGSTSGSGK) followed by LAP (DEVLVEIETDKVVLEVPASADG) followed by an EcoRI site. The PCR product was digested with BamHI and EcoRI enzymes and ligated into similarly-digested pET-21b vector to obtain the FKBP-LAP-pET-21b plasmid.

Bacterial expression and purification of FRB-LAP and FKBP-LAP in pET-21b

The pET-21b plasmid containing the FRB-LAP (or FKBP-LAP) gene was introduced into *E. coli* strain BL21 (DE3) by heat-shock transformation. The cells were cultured in LB supplemented with ampicillin (100 μg/mL) at 37 °C until OD_{600} 0.5. Protein expression was induced by the addition of IPTG to a final concentration of 0.4 mM. After incubation at 30 °C for 3 h, the cells were harvested by centrifugation. The cells were lysed by resuspension in bacterial protein extraction reagent (B-PER, Pierce) supplemented with 0.5 mM PMSF, protease inhibitor cocktail (Calbiochem), and DNase I (Roche). After incubation at room temperature for 20 min, the lysate was cleared by centrifugation, and the supernatant was loaded onto a Ni-NTA agarose column (Qiagen). Fractions containing FRB-LAP (or FKBP-LAP) were consolidated and dialyzed against PBS pH 7.4 for overnight. The concentration of protein was determined by BCA assay.

Photocrosslinking of FRB-LAP and FKBP-LAP in presence or absence of rapamycin

FRB-LAP (2 μL of 10 μM stock, C_f=1 μM) and FKBP-LAP (2 μL of 10 μM stock, C_f=1 μM) were added to a small eppendorf tube. To some reactions, 2 μL of 100 μM rapamycin was added (C_f=10 μM). Photocrosslinking was performed on 6 μL samples, placed 1.5 inches from a 800 W Hanovia UV lamp separated by Pyrex glass to filter out light <300 nm. After irradiation for 1, 3, 5 min on ice, we analyzed the samples by Western blotting using anti-HA and anti-myc antibodies.

References

[1] Nooren, I.M.A. & Thornton, J.M., Diversity of protein-protein interactions. *EMBO J.* 22, 3486-3492 (2003).

[2] Vaynberg, J. & Qin, J., Weak protein-protein interactions as probed by NMR spectroscopy. *Trends. Biotechnol.* 24, 22-27 (2006).

[3] van der Merwe, P.A. & Barclay, A.N., Transient intercellular adhesion: the importance of weak protein-protein interactions. *Trends. Biochem. Sci.* 19, 354-358 (1994).

[4] Fields, S. & Song, O., A novel genetic system to detect protein-protein interactions. *Nature* 340, 245-246 (1989).

[5] Hu, C.D. & Kerppola, T.K., Simultaneous visualization of interactions between multiple proteins in living cells using multicolor biomolecular fluorescence complementation analysis. *Nat. Biotechnol.* 21, 539-545 (2003).

[6] Galarneau, A., Primeau, M., Trudeau, L.E., & Michnick, S.W., β-Lactamase protein fragment complementaion assays as *in vivo* and *in vitro* sensors of protein-protein interactions. *Nat. Biotechnol.* 20, 619-622 (2002).

[7] Luker, K.E. *et al.*, Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A* 101, 12288-12293 (2004).

[8] Stefan, E. *et al.*, Quantification of dynamic protein complexes using Renilla luciferase fragment complementation applied to protein kinase A activities in vivo. *Proc. Natl. Acad. Sci. U. S. A* 104, 16916-16921 (2007).

[9] Truong, K. & Ikura, M., The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Curr. Opin. Struct. Biol.* 11, 573-578 (2001).

[10] Michnick, S.W., Protein fragment complementation strategies for biochemical network mapping. *Curr. Opin. Biotech.* 14, 610-617 (2003).

[11] Magliery, T.J. *et al.*, Detecting Protein–Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc* 127, 146-157 (2005).

[12] Choi, S.K., Kalivretenos, A.G., Usherwood, P.N.R., & Nakanishi, K., Labeling studies of photolabile philanthotoxins with nicotinic acetylcholine receptors: mode of interaction between toxin and receptor. *Chem. Biol.* 2, 23-32 (1995).

[13] Grutter, T., Goeldner, M., & Kotzyba-Hibert, F., Nicotinic Acetylcholine Receptor Probed with a Photoactivatable Agonist: Improved Labeling Specificity by Addition of Ce^{IV}/ Glutathione. Extension to Laser Flash Photolabeling. *Biochemistry* 38, 7476-7484 (1999).

[14] van Criekinge, W. & Beyaert, R., Yeast Two-Hybrid: State of the Art. *Biol. Proced. Online.* 2 (1-38) (1999).

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

[16] Chin, J.W. *et al.*, Addtion of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* 124, 9026-9027 (2002).

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

[18] Trippmann, E.M., Liu, W., Summerer, D., Mack, A.V., & Schultz, P.G., A Genetically Encoded Diazirine Photocrosslinker in Escherichia coli. *Chembiochem* 8, 2210-2214 (2007).

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

[20] Ye, S. *et al.*, Site-specific Incorporation of Keto Amino Acids into Functional G Protein-coupled Receptors Using Unnatural Amino Acid Mutagenesis. *J. Biol. Chem.* 283, 1525-1533 (2008).

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

[22] Kirshenbaum, K., Carrico, I.S., & Tirrell, D.A., Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine analogues. *Chembiochem.* 3, 235-237 (2002).

[23] Liu, W., Brock, S., Chen, S., & Schultz, P.G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* 4, 239-244 (2007).

[24] Wang, L., Xie, J., & Schultz, P.G., Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225-249 (2006).

[25] Lu, T. *et al.*, Probing ion permeation and gating in a K^+ channel with backgone mutations in the selectivity filter. *Nat. Neuroscience* 4, 239-246 (2001).

[26] Baruah, H., Puthenveetil, S., Choi, Y.A., Shah, S., & Ting, A.Y., An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photo-cross-linking. *Angew Chem Int Ed Engl* 47 (37), 7018-7021 (2008).

[27] Fujiwara, K. *et al.*, Crystal Structure of Lipoate-Protein Ligase A from Escherichia coli: Determination of the lipoic acid-binding site. *J.Biol. Chem.* 280, 33645-33651 (2005).

[28] Kim, D.J. *et al.*, Crystal Structure of Lipoate-Protein Ligase Bound with the Activated

Intermediate: Insights into interaction with lipoyl domains. *J. Biol. Chem.* 280 (38081-38089) (2005).

[29] Fernandez-Suarez, M. *et al.*, Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* 25 (12), 1483-1487 (2007).

[30] Green, D.E., Morris, T.W., Green, J., Cronan, J.J.E., & Guest, J.R., Purification and properties of the lipoate protein ligase of Escherichia coli. *Biochem J.* 309, 853-862 (1995).

[31] Luchansky, S.J., Goon, S., & Bertozzi, C.R., Expanding the diversity of unnatural cell-surface sialic acids. *Chembiochem* 5, 371-374 (2004).

[32] Agard, N.J., Baskin, J.M., Prescher, A., Lo, A., & Bertozzi, C.R., A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* 1, 644-648 (2006).

[33] Banaszynski, L.A., Liu, C.W., & Wandless, T.J., Characterization of the FKBP•rapamycin•FRB Ternary Complex. *J. Am. Chem. Soc.* 127, 4715-4721 (2005).

[34] Choi, J., Chen, J., Schreiber, S.L., & Clardy, J., Structure of the FKBP12-Rapamycin Complex Interacting with Binding Domain of Human FRAP. *Science* 273, 239-242 (1996).

Appendix

Characterization of AP conjugated oligonucleotides :

HPLC and MALDI-TOF (Chapter 3)

Characterization of ketone biotin conjugated oligonucleotides :

MALDI-TOF (Chapter 4)

Spectral characterization of aryl azide 1 (Chapter 5)

Characterization of AP conjugated oligonucleotides :

HPLC and MALDI-TOF (Chapter 3)

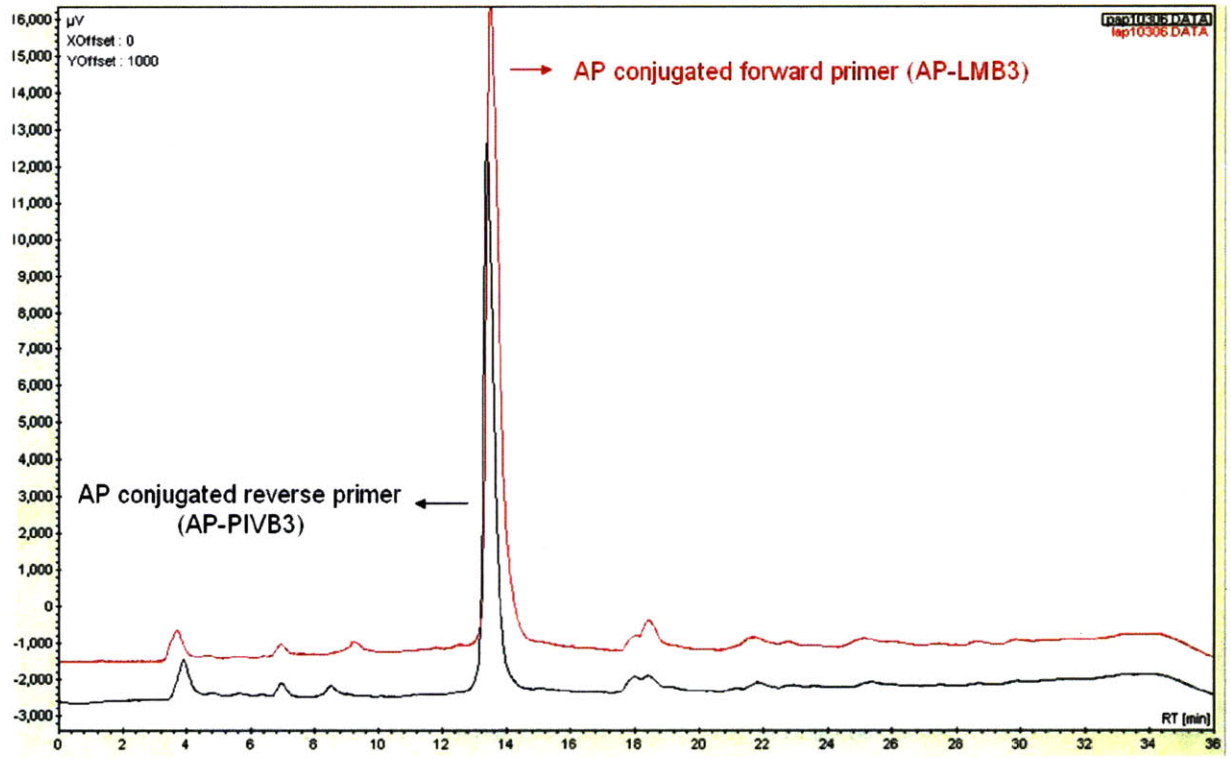


Figure 1 HPLC analysis of the purified AP conjugated oligonucleotides (Chapter 3). AP conjugated LMB3 (a forward primer for PCR; oligonucleotide sequence: GGA AGG GCG ATC GGT GCG) is shown in red, and AP conjugated PIVB3 (a reverse primer for PCR; oligonucleotide sequence: GAG CAC TGT CCG ACC GC) is shown in black. The purity of both samples are >90%.

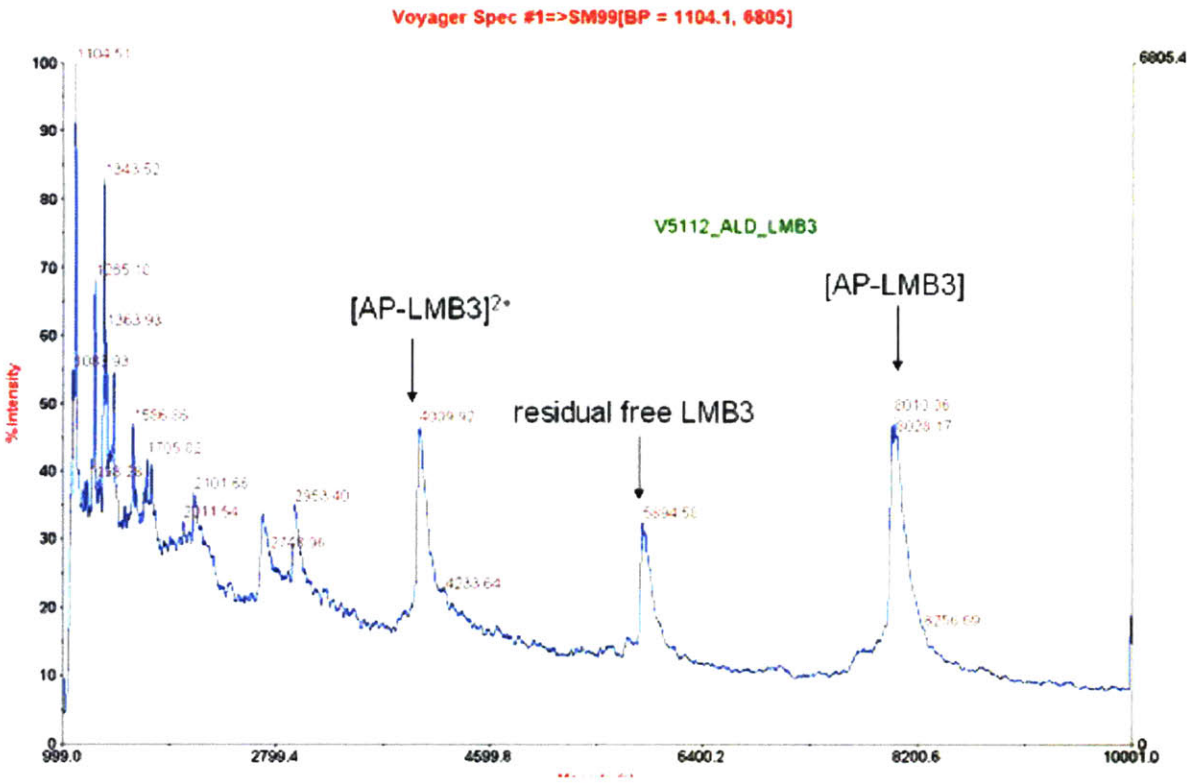


Figure 2 MALDI-TOF analysis of the purified AP conjugated oligonucleotide. AP conjugated LMB3 (a forward primer for PCR; oligonucleotide sequence: GGA AGG GCG ATC GGT GCG) is shown here, and the observed mass (8010.06) matched with the calculated value (8005.7) within error range. For some reason, we have failed to get mass peaks corresponding to AP conjugated PIVB3 (see Chapter 3 for further discussion)

Characterization of ketone biotin conjugated oligonucleotides :

MALDI-TOF (Chapter 4)

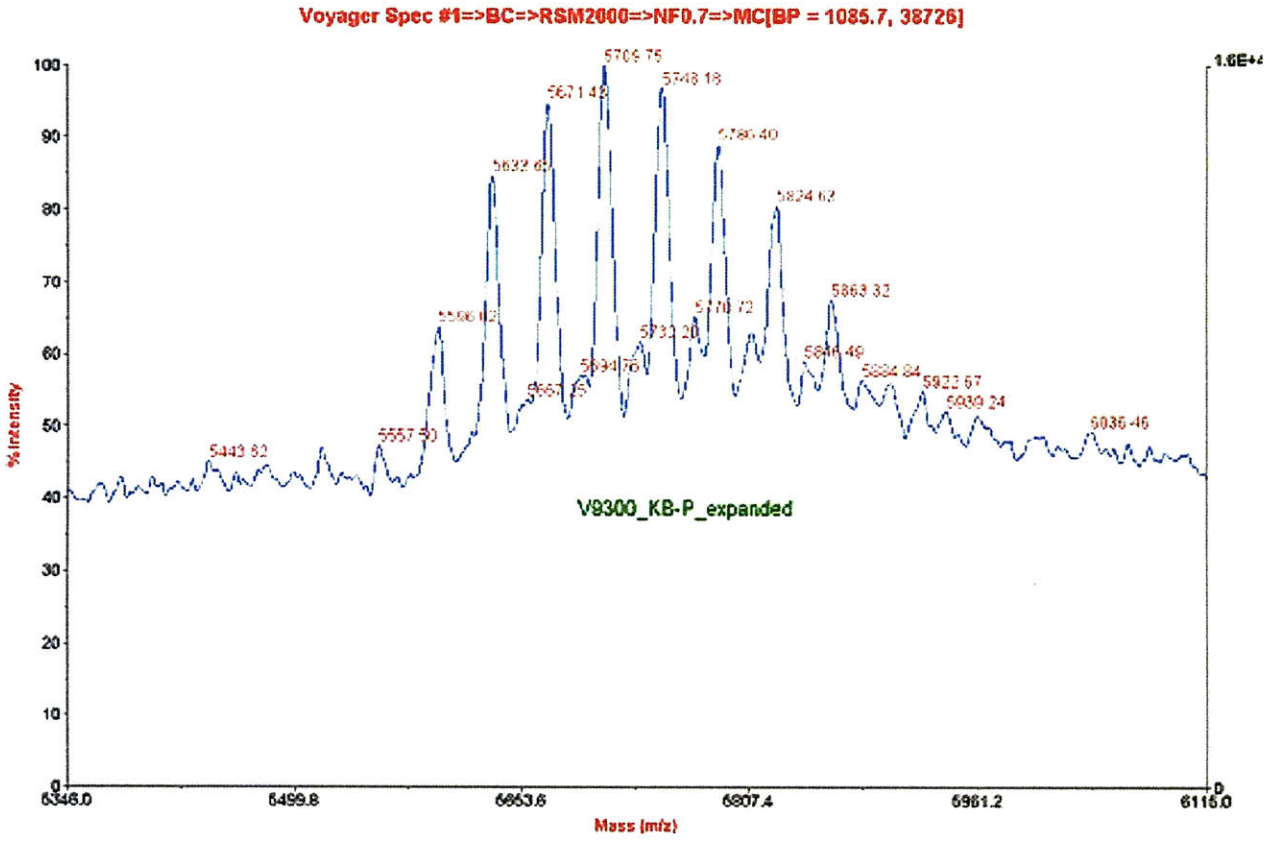


Figure 3 MALDI-TOF analysis of the purified ketone biotin conjugated oligonucleotide (PIVB-1). PIVB-1 is a reverse primer for PCR (oligonucleotide sequence: 5'- GCG TTG ATG CAA TTT CT). The observed mass (5596.02, the most left peak) matched with the calculated value (5595.7) within error range. Multiple peaks are responsible for Na^+ or K^+ increments, which are usually observed for oligonucleotides.

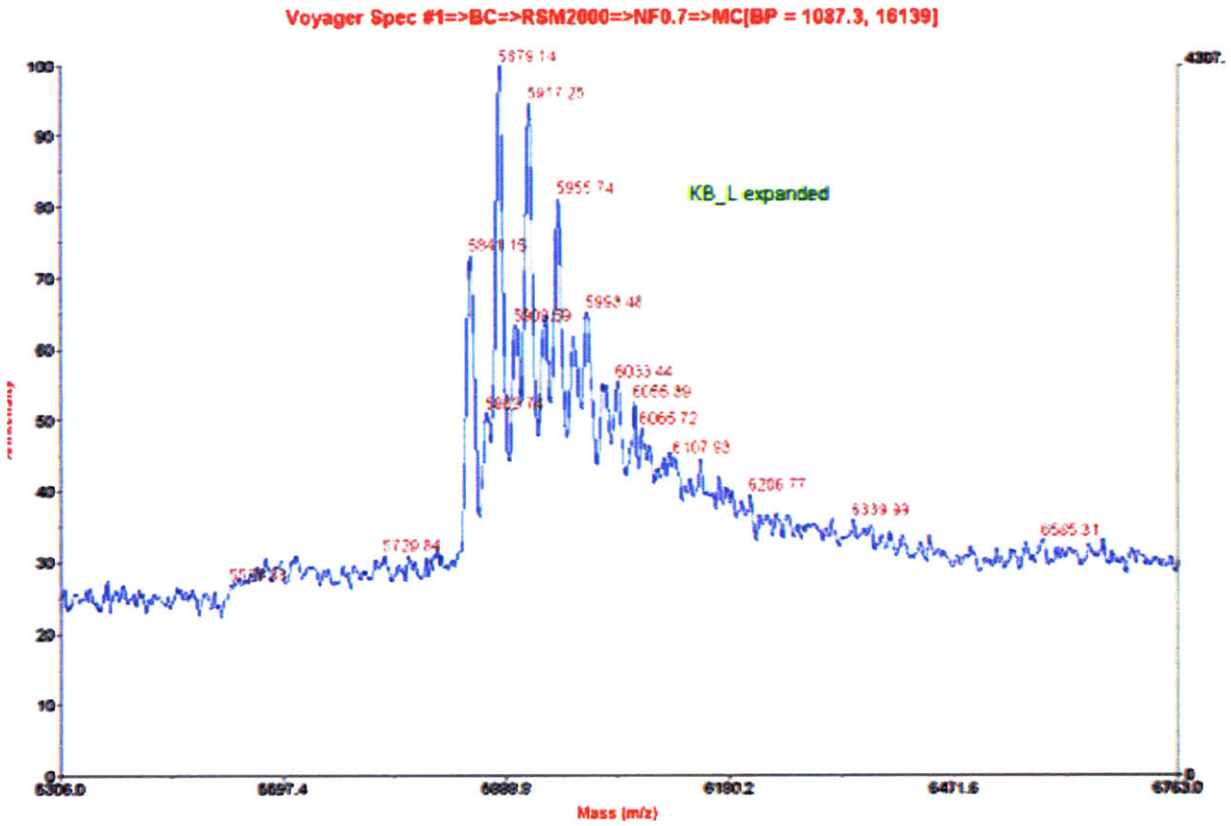


Figure 4 MALDI-TOF analysis of the purified ketone biotin conjugated oligonucleotide (LMB2-1). LMB2-1 is a forward primer for PCR (oligonucleotide sequence: 5'- CAG GCG CCA TTC GCC ATT). The observed mass (5841.15, the most left peak) matched with the calculated value (5839.9) within error range. Multiple peaks are responsible for Na^+ or K^+ increments, which are usually observed for oligonucleotides.

Spectral characterization of aryl azide 1 (Chapter 5)

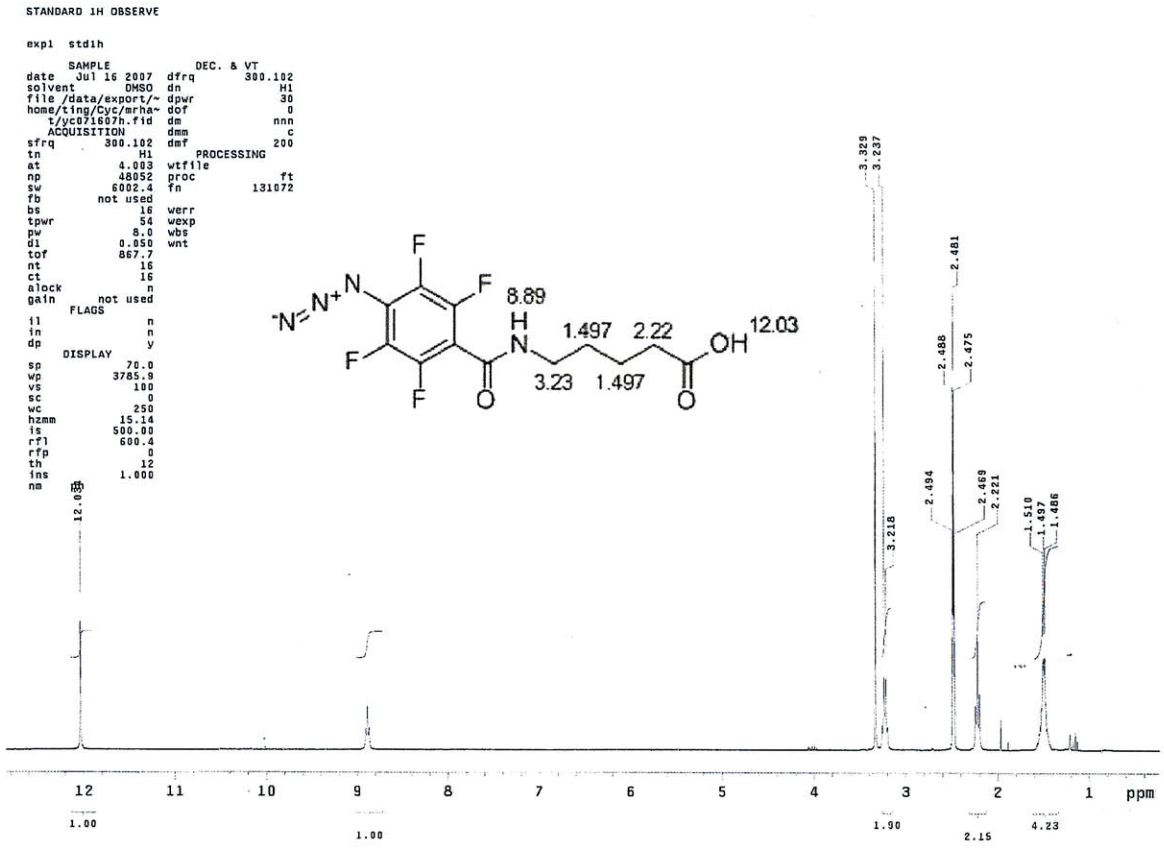


Figure 5 ^{1}H-NMR spectra of aryl azide **1**

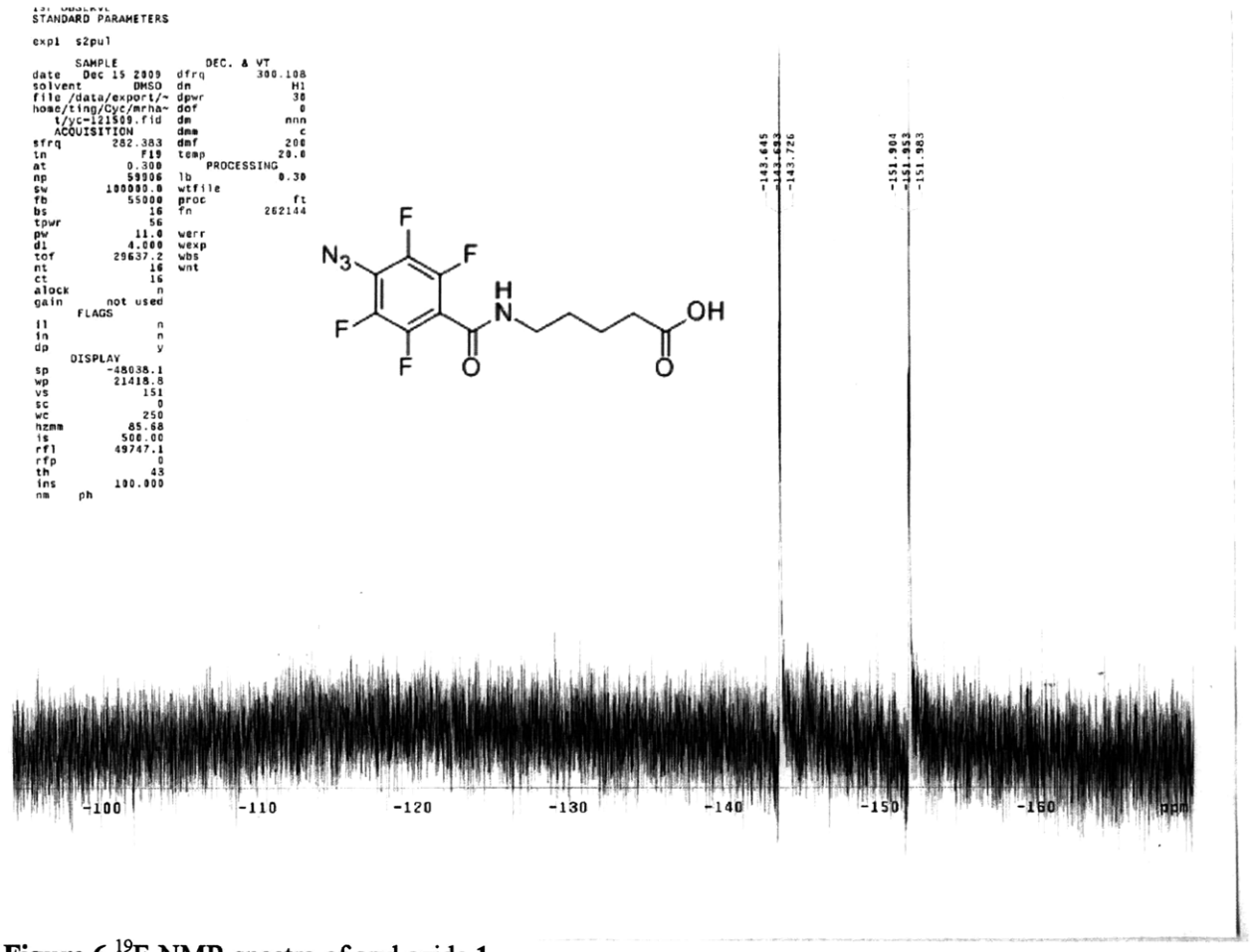


Figure 6 ^{19}F-NMR spectra of aryl azide **1**

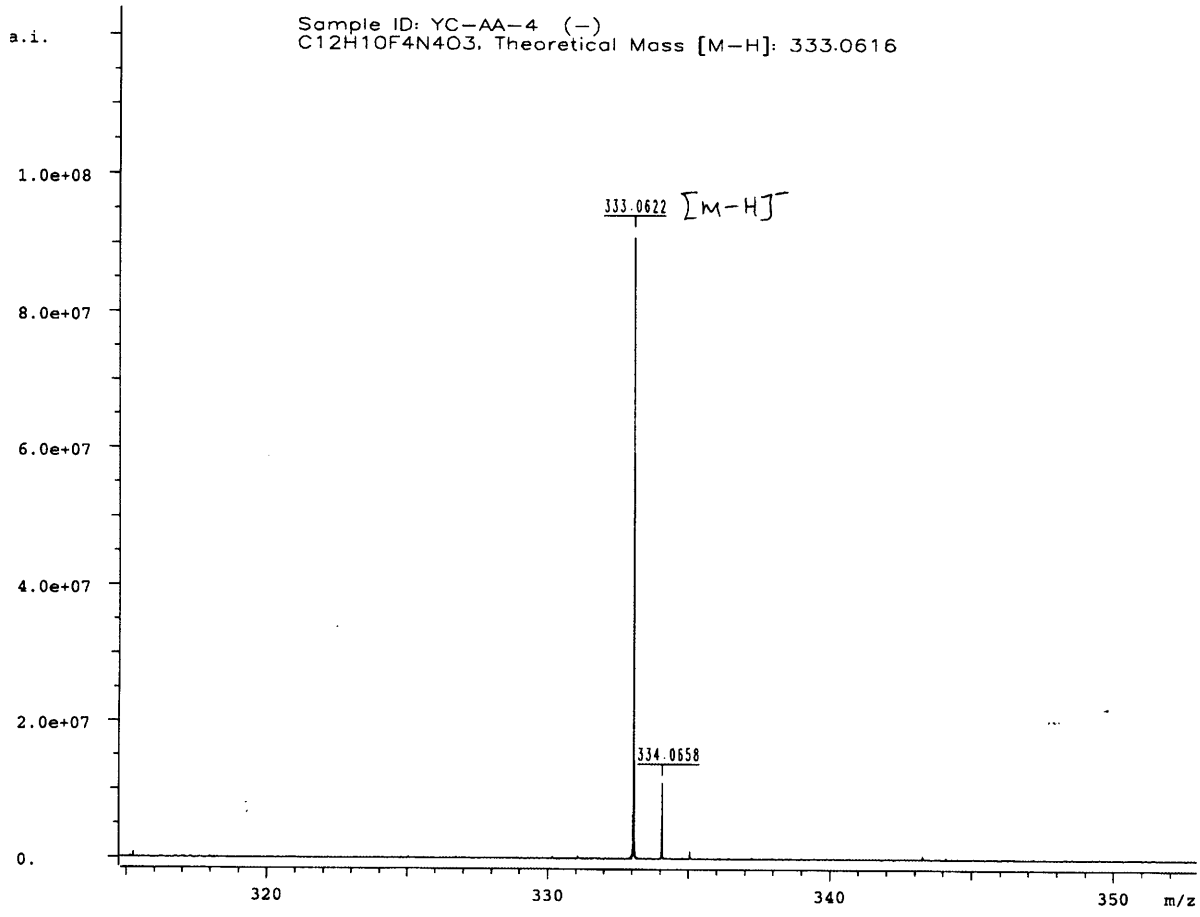


Figure 7 HR-MS spectra of aryl azide **1**

YOON-AA CHOI

EDUCATION

Ph. D. in Biochemistry, *Massachusetts Institute of Technology*, Cambridge, MA

Advisor: Prof. Alice Y. Ting Sep 2004- Feb 2010

B. S. in Chemistry, *Korea Advanced Institute of Science and Technology,* Daejon, Korea

Advisor: Prof. Sukbok Chang Mar 2000- Feb 2004

RESEARCH EXPERIENCE

Research Assistant. Topic: "Site-specific incorporation of photoaffinity probes for detection of transient protein-protein interactions *in vitro* and in living cells

"Yeast display evolution of a peptide substrate for yeast biotin ligase and application to two-color quantum dot labeling of cell surface proteins"

Laboratory of Professor Alice Ting, Department of Chemistry, MIT, December, 2004-2009

Research Assistant. Topic: "Oxidative cleavage of alkynes using $K_2OsO_4 \cdot 2H_2O$ as an oxidant and total synthesis of indolizidine-167B," Organo Transition Metal Catalysis and Materials Laboratory, Department of Chemistry, KAIST, December, 2002 –June, 2003

HONORS AND AWARDS

Samsung Scholarship for Graduate Studies.The Samsung Lee Kun Hee Scholarship Foundation

PUBLICATIONS

Park, S., **Choi, Y.**, Han, H., Yang, S., and Chang, S. (2003) Rh-Catalyzed one-pot and practical transformation of aldoximes to amides. *Chem. Comm.* 15, 1936 – 1937

Chen, I., **Choi, Y. –A.**, and Ting, A. Y. (2007) 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.* 129, 6619-6625.

Slavoff, S. A., Chen, I., **Choi, Y.-A.**, and Ting, A. Y. (2008) Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *J. Am. Chem. Soc.*130, 1160-1162.

Baruah, H., Puthenveetil, S., **Choi, Y.-A.**, Shah, S., and Ting, A. Y. (2008) An engineered aryl azide ligase for site-specific mapping of protein-protein interactions through photocrosslinking. *Angewandte Chemie International Edition* 2008, 47, 7018-7021.