

Directed evolution of split APEX peroxidase

By

Yisu Han

B.S. Chemistry (2013)
University of California, Berkeley

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

at the

Massachusetts Institute of Technology

September 2018

© 2018 Massachusetts Institute of Technology
All rights reserved

Signature redacted

Signature of the author: _____

Department of Chemistry
August 21, 2018

Signature redacted

Certified by: _____

Alice Y. Ting
Professor of Genetics, Biology, and Chemistry (by courtesy), Stanford University
Thesis Supervisor

Signature redacted

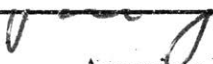
Accepted by: _____

Robert W. Field
Haslam and Dewey Professor of Chemistry Chair,
Department Committee on Graduate Students

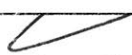


This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:


Signature redacted

 Bradley L. Pentelute
Associate Professor of Chemistry

Signature redacted

 Alice Y. Ting
Professor of Genetics, Biology, and Chemistry (by courtesy), Stanford University
Thesis Supervisor

Signature redacted

 Elizabeth M. Nolan
Associate Professor of Chemistry

Directed evolution of split APEX peroxidase

By

Yisu Han

Submitted to the Department of Chemistry
on August 21, 2018 in partial fulfillment of the
Requirements for the Degree of the Doctor of Philosophy

ABSTRACT

APEX is an engineered peroxidase that catalyzes the oxidation of a wide range of substrates, facilitating its use in a variety of applications, from subcellular staining for electron microscopy to proximity biotinylation for spatially restricted proteomics and transcriptomics. While this strategy has provided access to many cellular regions and organelles, there are still many compartments and structures that cannot be accessed; this strategy is limited by the specificity of genetic targeting; there are cellular regions that cannot be exclusively targeted by a single genetic tag (Chapter 1).

To further advance the capabilities of APEX and address the need for an interaction-dependent proximity labeling tool, this thesis describes the development of a split APEX2 system. Short enzymatic reconstitution times are also desired, to further ensures both organelles' morphological integrity. Thus, it is critical that split APEX2 reconstitute peroxidase activity both rapidly and robustly. We first performed two subsequent rounds of structure-guided screening to determine the most optimal cut site (Chapter 2). We then used directed evolution on the top candidate pair to engineer a split APEX tool (sAPEX). Selections were performed via FACS on yeast-displayed fragment libraries, and 20 rounds of evolution produced a 200-amino acid N-terminal fragment (with 9 mutations relative to APEX2) called "AP" and a 50-amino acid C-terminal fragment called "EX". AP and EX fragments were each inactive on their own, but reconstituted to give peroxidase activity when driven together by a molecular interaction (Chapter 3). Our resulting split APEX2 fragment pair has significantly diverged from its parental sequence and shows interaction-dependent reconstitution in multiple contexts in living mammalian cells (Chapter 4).

Our split APEX tool adds to the proximity labeling toolkit (Chapter 5 and 6), and in the future, should extend the utility of APEX-based approaches to new areas of biology at higher spatiotemporal resolution.

Thesis Supervisor: Alice Y. Ting

Title: Professor of Genetics, Biology, and Chemistry (by courtesy), Stanford University

Acknowledgements

First, I would like to thank Alice, my thesis advisor, for giving me the opportunity to join her lab. I was able to explore my passions in science and methodology development, and coming from a synthetic chemistry background, I am grateful that she had taken a chance on me to allow me to learn a whole new field. The past five years in her lab, I have learned the most about being a rigorous scientist. Her mentorship provided insight into how to pragmatically design the best experiments, with the most rigorous questions - asking the right questions, to generate the most impactful and significant data. Furthermore, I would also like to thank Matt Shoulders, Bradley Pentelute, and Elizabeth Nolan for being on my thesis committee. Their mentorship and guidance through annual meetings and interactions within classes have been memorable and helpful.

Furthermore, I would like to thank all the mentors I've had in my life who have given me a chance to explore science. From my high school chemistry teacher, Mrs. Celia Reidler, to the lab at Johns Hopkins Applied Physics Laboratory (George, Lance, Andrew, Tracy, Edd, Carl, Sarah) that allowed to do real bench work and solidify my love for science. Thank you for trusting me, believing in me, and encouraging me. I remember the care packages you've sent over the years. I also have to thank my undergrad mentors Carolyn Bertozzi and Chelsea Gordon, for emboldening me to apply to my dream grad schools. Thank you for giving me a chance to learn and expand my passions.

I have also been blessed to meet some of the most amazing people during my time in the Ting lab. I would like to thank the following people for their friendship, conversation and companionship, hot chocolate dates, cookie runs, and a lot of commiseration over food over the years: Cathy Amaya, Tess Branon, Kelvin Cho, Kurt Cox, Robert Coukos, Elbeg Erdenee, Shuo Han, Victoria Hung, Chai Kaewsapsak, Tina Kim, Stephanie Lam, Jeff Martell, Monica Neugebauer, Oom Pattarabanjird, Mateo Sanchez, Wenjing Wang, and Boxuan Zhao. I would especially like to extend my gratitude towards Jeff Martell and Wenjing Wang for being the most valuable wealth of resource; you guys have been amazing mentors and teachers in not just how to do science and troubleshoot, but how to be a scientist and a good person.

Graduate school was not always easy, and I would also like to thank my furbabies LC and Ollie for providing me a constant source of comfort and fluff. I would also like to extend the deepest gratitude towards my boyfriend, Stephen Jedidiah Van Wyck. I met you just as I was reaching the last few months of my graduate career. I appreciate your curiosity and passion for science and nature, your calm demeanor, and your kindness. Thank you for bringing me flowers and feeding me dumplings, boba, hot chocolate, and In-n-Out. For pushing me, (kindly), to finish my thesis with accountability texts. Your rational voice-of-reason resonates love whenever I felt overwhelmed and you provided waves of calm. Thanks for being the most supportive and best surprise to ever come into my graduate life. You bring more light into my life every day. I love you.

Finally, I am most thankful to my parents and extended family. Without their sacrifices, their motivation; their guidance, and love, I would be nowhere. My parents provide me the unwavering support to pursue my dreams. They are there to lift me up when I trip on rocky roads, but also my biggest cheerleaders when I do succeed. I am so thankful for my role models to also be my best support crew. Thank you. I love you. This thesis is dedicated to my parents, for none of this would be possible without them.

Table of Contents

Title Page.....	1
Signature Page.....	2
Abstract	3
Acknowledgements.....	4
Table of contents.....	5
List of figures.....	7
List of tables.....	9
List of abbreviations.....	10
Chapter 1. Introduction to APEX	12
Background and development.....	13
Applications	18
Splitting APEX to make a protein complementation assay (PCA).....	20
Discussions.....	21
References	22
Chapter 2. Rationally-selected screening of APEX2 cut sites	27
Introduction.....	28
Amplex UltraRed Screen of rationally-selected APEX2 cut sites for enzymatic reconstitution in HEK 293T cells.....	29
Live biotin-phenol labeling assay of potential split APEX pairs.....	32
Discussion.....	34
Experimental methods.....	36
References	41
Chapter 3. Directed Evolution of split APEX	42
Introduction.....	43
Establishing a platform.....	44
Yeast display evolution of AP (large fragment)	49
FACS-based negative selections for reduced fragment affinity.....	54
Modulating heme concentrations	56
Discussion.....	60

Experimental methods.....	63
References	69
Chapter 4. Testing different clones of split APEX in Mammalian Cells.....	70
Introduction.....	71
Determining the best clone from each generation of directed evolution.....	71
Comparing finalized split APEX (AP +EX) against earlier generations	90
Comparing a published split APEX2 ¹ , our original split APEX2, and evolved split APEX2.....	95
Discussion.....	98
Experimental methods.....	99
References	104
Chapter 5. Applying sAPEX to the mito-ER junction for EM imaging and proteomic mapping.....	105
Introduction to the mito-ER junction.....	106
Applying sYFP and sGFP at the mito ER junction.....	108
Optimizing sAPEX targeting and expression levels for mito-ER applications.....	113
Preliminary testing of sAPEX at the mito-ER junction for proteomics in COS7 cells.....	118
Using sAPEX for DAB labeling across the MITO-ER.....	123
Discussion.....	125
Experimental Methods.....	126
References.....	131
Chapter 6. Reconstitution of sAPEX on a target RNA motif.....	134
Introduction.....	135
Applying sAPEX and optimizing expression levels of AP and EX fragment	137
Discussion.....	148
Experimental methods.....	149
References	154

List of Figures

Figure 1-1: The mechanism of Class I peroxidases, model structures obtained from the Gumiero, et al. paper ²	14
Figure 1-2: The binding site of horseradish peroxidase (HRP) with ferulic acid (FA) in the active site.....	15
Figure 1-3: APEX2 is an engineered product of structure-guided mutagenesis and yeast display-directed evolution.....	16
Figure 1-4: APEX2 can be utilized as both a reporter for EM and proteomics.....	18
Figure 1-5. Overview of protein complementation assays (PCAs)	20
Figure 2-1. Selection of potential split APEX cut sites.....	28
Figure 2-2. split APEX cut site screening platform.....	29
Figure 2-3. Screening of potential split APEX cut sites.....	30
Figure 2-4. Focused Amplex UltraRed screening of potential sAPEX cut sites with single fragment controls.....	31
Figure 2-5. Testing split APEX pairs using biotin-phenol as substrate.....	33
Figure 2-6. Biotin-phenol assay of the 3 most promising split APEX2 cut site pairs at the mitochondrial junction.....	35
Figure 3-1. Displaying split APEX2 candidates as circular permutations on the yeast surface....	44
Figure 3-2. Displaying split APEX2 pairs as fusions to different yeast surface proteins.....	47
Figure 3-3. Comparing split display (SD200) and circularly permuted (CP89) platforms to full length APEX2 fused to Aga2p.....	48
Figure 3-4. Generation 1 yeast display-based directed evolution.....	50
Figure 3-5. Generation 2 yeast display-based directed evolution.....	52
Figure 3-6. Generation 3 yeast display-based directed evolution.....	53
Figure 3-7. Generation 4 yeast display-based directed evolution.....	55
Figure 3-8. Using heme biosynthesis inhibitor, succinyl acetone, to reduce CP89 peroxidase activity.....	57
Figure 3-9. Testing exogenous heme supplementation prior to EX-GFP incubation.....	59
Figure 3-10. Yeast-display directed evolution and results.....	62
Figure 4-1. Amplex UltraRed Assay of AP-0 mutants derived from first generation of yeast display evolution.....	73
Figure 4-2. Biotin-phenol labeling on live HEK 293T cells comparing the top mutants of Generation 1 evolution against APEX2, and starting template AP-0.....	75
Figure 4-3. Amplex UltraRed and Biotin-phenol labeling of AP-1 mutants derived from second generation of yeast display evolution.....	78
Figure 4-4. DAB labeling of AP-1 mutants derived from second generation of yeast display evolution.....	79
Figure 4-5. Biotin-phenol labeling comparing AP-0, AP-1 AP-2 in HEK 293T cells stably expressing EX-HA-FRB ERM.....	81
Figure 4-6. Biotin-phenol labeling of AP-2 mutants derived from third generation of yeast display evolution in stable EX-FRB-ERM cells.	82
Figure 4-7. Diaminobenzidine (DAB) labeling of AP-2 mutants derived from third generation of yeast display evolution in stable EX-FRB-ERM cells.....	84

Figure 4-8. Biotin-phenol and Amplex UltraRed labeling of AP-3 mutants derived from fourth generation of yeast display evolution in monoclonal stable EX-FRB-ERM COS7 cells.....	86
Figure 4-9. Biotin-phenol labeling of finalized split APEX in monoclonal stable EX-FRB-ERM COS7 cells infected with FKBP-AP.....	86
Figure 4-10. Western blot analysis of biotin-phenol labeling of finalized split APEX in stable FRB-EX HEK 293T.....	89
Figure 4-11. Comparing the different generations of evolved split APEX clones in mammalian cells.....	91
Figure 4-12. Comparison of sAPEX variants in the mammalian cytosol, with biotin-phenol labeling as readout of peroxidase activity.....	93
Figure 4-13. Examining accumulated AP mutations in full length APEX2.....	94
Figure 4-14. Comparing a published split APEX2 ¹⁰ , our original split APEX2, and evolved split APEX2.....	96
Figure 5-1: ER-mitochondria encounter structure (ERMES) consists of four proteins.....	107
Figure 5-2. Testing split GFP at the mito-ER junction.....	109
Figure 5-3. Testing split YFP at the mito-ER junction.....	111
Figure 5-4. Testing mito and ER targeting sequences fused to the FKBP/FRB scaffold.....	114
Figure 5-5. Testing new mito and ER targeting sequences fused to the FKBP/FRB scaffold for sYFP reconstitution	116
Figure 5-6. Testing sAPEX at mito-ER contact sites.....	118
Figure 5-7. Examining sAPEX mito-ER targeting and morphology.....	119
Figure 5-8. Western blot analysis of live biotin-phenol labeling of sAPEX targeted to the mito-ER	122
Figure 5-9. DAB labeling of sAPEX targeted to the mito-ER.....	124
Figure 6-1. Testing MS2 stem-loop coating protein fused to full length APEX2 to target it to an RNA binding site motif.....	136
Figure 6-2. Testing split APEX targeted to an RNA binding site motif.	138
Figure 6-3. Noncoding RNA construct design.....	141
Figure 6-4. Streptavidin-enrichment of biotin-phenol labeled HEK 293T cells expressing split APEX.....	143
Figure 6-5. Testing Split APEX2 P2A constructs.....	145
Figure 6-6. Testing HEK 293T cells stably expressing PCP-EX against double stable cells expressing PCP-EX and MCP-AP.....	147

List of tables

Table 2: Plasmids used in Chapter 2.....	38
Table 3: Plasmids used in Chapter 3.....	68
Table 4: Plasmids used in Chapter 4.....	103
Table 5: Plasmids used in Chapter 5.....	131
Table 6: Plasmids used in Chapter 5.....	154

List of abbreviations

8-oxo-dGTP	8-oxo-2'-deoxyguanosine-5'-triphosphate
AF	alexafluor
Aga1p	yeast surface glycol protein A-agglutinin, one of the two-subunits
Aga2p	yeast surface glycol protein A-agglutinin, one of the two-subunits
AP	evolved, finalized N-terminal large fragment of sAPEX
AP-0	starting template, unevolved N-terminal large fragment of split APEX
AP-1	evolved first generation N-terminal split APEX fragment
AP-2	evolved second generation N-terminal split APEX fragment
AP-3	evolved third generation N-terminal split APEX fragment
APEX	first generation enhanced APX
APEX2	second generation APEX
APX	wild-type ascorbate peroxidase
ATP	adenosine triphosphate
AU	arbitrary unit
BP	biotin-phenol
BirA	E. coli biotin ligase
BSA	bovine serum albumin
CAS9	CRISPR associated protein 9
CFP	cyan fluorescent protein
CMV	Cytomegalovirus promoter
COS7	african green monkey kidney cells
CP	circularly permuted
CRISPR	clustered regularly-interspaced short palindromic repeats
DAB	3,3'-diaminobenzidine
DMEM	Dulbecco's modified Eagle's medium
DNA	deoxyribonucleic acid
DPBS	Dulbecco's phosphate-buffered saline
dPTP	2'-deoxy-P-nucleoside-5'-triphosphate
DTT	dithiothreitol
ER	endoplasmic reticulum
ERM	endoplasmic reticulum membrane
EX	C-terminal small fragment of sAPEX
EM	electron microscopy
FACS	fluorescence activated cell sorting
FBS	fetal bovine serum
FKBP	FK506 binding protein
Flag	synthetic epitope tag
FP	fluorescent protein
FRB	FKBP-12, rapamycin binding domain
GFP	green fluorescent protein
HA	epitope tag derived form human influenza hemagglutinin surface glycoprotein
H ₂ O ₂	hydrogen peroxide
HEK 293T	human embryonic kidney cells

His6.....	histidine hexamer protein tag
HRP	horseradish peroxidase
IMS.....	mitochondrial intermembrane space
KDEL.....	endoplasmic reticulum retention sequence
MAVS	mitochondrial antiviral signaling protein
MCP.....	MS2 stem loop coating protein
MEM.....	minimum essential medium
Mito-ER.....	mitochondrial-endoplasmic reticulum
MS	mass spectrometry
MW.....	molecular weight
Myc.....	epitope tag derived from c-myc
NEB.....	New England BioLabs
NES.....	nuclear export signal
NLS.....	nuclear localization signal
Ni-NTA.....	nickel nitrilotriacetic acid
OMM.....	outer mitochondrial membrane
PCA.....	protein complementation assay
PCP	PP7 stem loop coating protein
PCR.....	polymerase chain reaction
PMSF.....	phenylmethylsulfonylfluoride
PPI.....	protein-protein interaction
RBP.....	RNA binding protein
RIPA.....	radioimmunoprecipitation assay
RNA.....	ribonucleic acid
SA-HRP.....	streptavidin-conjugated horse radish peroxidase
sAPEX.....	finalized evolved split APEX PCA system consisting of AP and EX
SD.....	split display
SGCAA.....	synthetic galactose plus casein amino acid
SDCAA	synthetic dextrose plus casein amino acid
SDS.....	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TEV.....	tobacco etch virus
V5.....	epitope tag derived from paramyxovirus of simian virus 5
WCL	whole cell lysate
YFP.....	yellow fluorescent protein

Chapter 1. Introduction to APEX

Background and development

In order to understand a complex system, it is critical to first identify its components. Typically to understand the components, the region of interest is isolated and examined in detail, and although standard subcellular fractionation increases the spatial information of a mass spectrometry proteomic experiment, the inherent purification process is subject to four major drawbacks. First, the lysis procedure may decrease spatial information depending on the harshness of the protocol. Second, the purification process can introduce false positives from contaminants and false negatives from the loss of material. Third, the purification procedure may be lengthy and laborious, in which the physiology of the region of interest may change. Fourth, there are cellular regions of interest that simply cannot be accessed by traditional centrifugation methods. Proximity-dependent labeling in living cells offer a novel way to map proteomes by bypassing complex organelle purification. Furthermore, all information is recorded in the context of a living cell. To address this need, the Alice Ting laboratory engineered and developed APEX2, which is derived from soybean ascorbate peroxidase (APX)², a heme-dependent class I peroxidase. Wild type APX is a dimer that scavenges hydrogen peroxide in the cytosol through the oxidation of ascorbate to monodehydroascorbate, which reduces hydrogen peroxide to water and prevents oxidative damage due to hydrogen peroxide and resultant hydroxyl radicals³⁻⁷. In addition to recognition of ascorbate as a reducing agent, ascorbate peroxidase has some promiscuity for the oxidation of aromatic substrates^{4,6-9}.

Class I peroxidases generally lack disulfide bonds, and Ca^{2+} dependence, which allows for peroxidase activity within the cytosol. The heme co-factor is not covalently bound but is necessary for catalytic activity. These heme-dependent peroxidases are a class of enzymes that catalyze the H_2O_2 -dependent oxidation of a wide variety of substrates using a 2-equivalent oxidized intermediate known as Compound I see equations below:

1. Peroxidase + $\text{H}_2\text{O}_2 \rightarrow$ Compound I + H_2O
2. Compound I + $\text{Substrate}_{\text{reduced}} \rightarrow$ Compound II + $\text{Substrate}_{\text{oxidized}}$
3. Compound II + $\text{Substrate}_{\text{reduced}} \rightarrow$ Peroxidase + $\text{Substrate}_{\text{oxidized}} + \text{H}_2\text{O}$

The proposed mechanism⁴ lead to future studies on in which x-ray crystal structures provided insight towards the mechanism for this class of peroxidases as well as possible intermediate structures. **Figure 1-1** depicts the how peroxidases utilizes the heme cofactor to oxidize substrates.

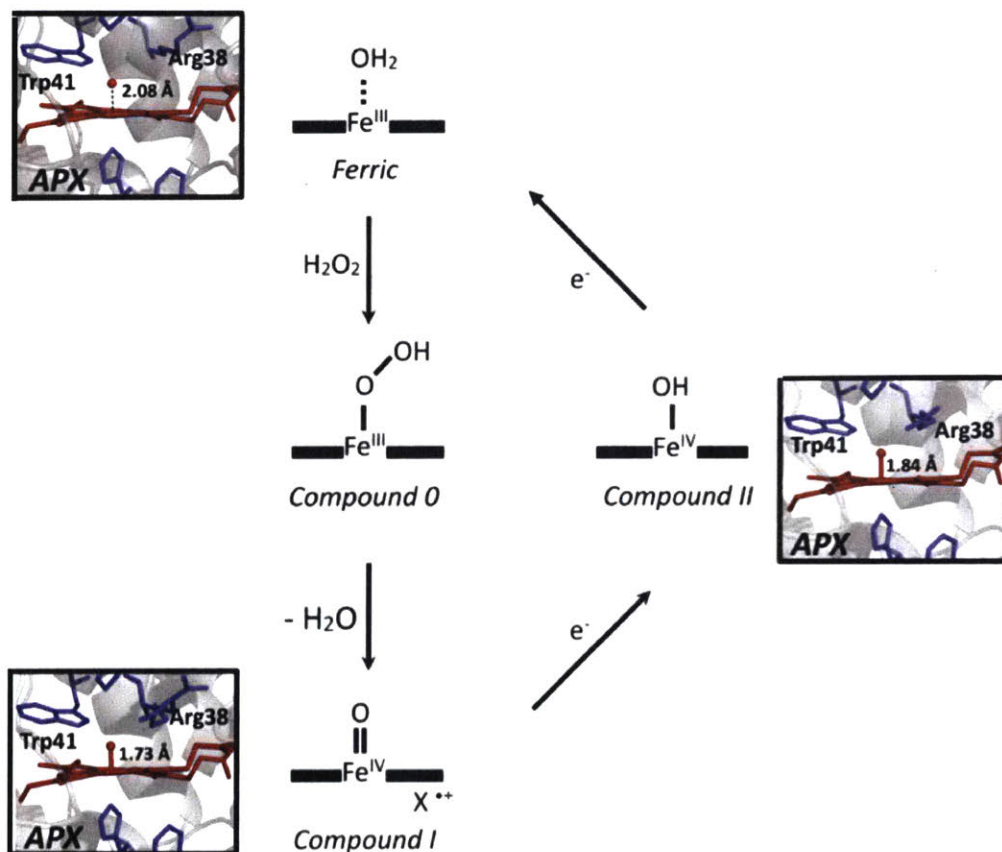


Figure 1-1: The mechanism of Class I peroxidases, model structures obtained from the Gumiero, et al. paper³. APX originates in a resting ferric state (top), but upon replacement of the presence of H_2O_2 , the occupancy changes the ferric to ferric hydroperoxyl, also known as Compound 0 (center). The activated compound I causes a loss of 1 water, and 1 electron (bottom). Once reducing substrate binds to the active site again, it donates one proton and one electron to Compound 1. Compound I transforms into Compound II (right), which lacks the radical cation in the porphyrin ring. After a second equivalence of a reducing agent binds to the active site, the cycle repeats and the donation of a proton and electron yield another oxidized substrate product and finally returns the enzyme back into the ferric resting state.

The scavenging of APX binds hydrogen peroxide and replacing the water, which otherwise occupies the binding site. The starting compound 0 structure still contains Fe in the +3 oxidation state. However once reaching the activated compound I form, in which the Fe is in a +4 oxidation state, the sequential binding and donation of proton and electron by a reducing agent, such as biotin-phenol for instance, can return the enzyme back to its original ferric group state (**Figure 1-1**).

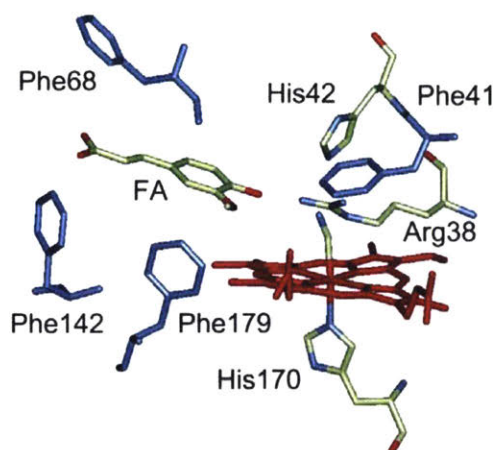


Figure 1-2: The binding site of horseradish peroxidase (HRP) with ferulic acid (FA) in the active site. The figure is adapted from Veitch¹⁰ and shows the cartoon representation taken from the x-ray crystal structure of the ternary complex: the substrate (ferulic acid), the cyanided-ligated HRP (variant C), and the heme co-factor necessary for catalysis (this is the molecule in red). The catalytic residues, Arg38, His42, and His170 are also shown, but most importantly the four aromatic phenylalanines are shown from the binding site. This hydrophobic patch of aromatic rings could contribute the high peroxidase activity of HRP towards a large range of aromatic substrates.

Using structure-guided mutagenesis, members of the Ting lab engineered APEX from APX through 3 mutations – K14D, E112K, W41F. The mutations K14D and E112K are charge reversal mutations that was designed to monomerize APEX (27 kDA), rather than its wild-type form of constitutive dimer. The W41F mutation was modeled after the active site of horseradish peroxidase (HRP) to improve the activity towards aromatic substrates. HRP is a prototypical class III plant peroxidase with a wide substrate tolerance, oxidizing almost any phenol or aniline to the corresponding radicals and releasing them from the active site. The substrate binding pocket of HRP contains a hydrophobic patch with 4 phenylalanines (**Figure 1-2**), thus leading to the W41F

mutation in APX. By replacing the indole with an aromatic phenyl-group increases the pi-pi stacking within the substrate pocket^{4,5}. Through the utilization of yeast display directed evolution on first generation APEX, the Ting lab engineered second generation APEX2, which introduces the fourth mutation A134P, which further improves enzymatic activity for aromatic substrates (Figure 1-3 for summary of engineering).

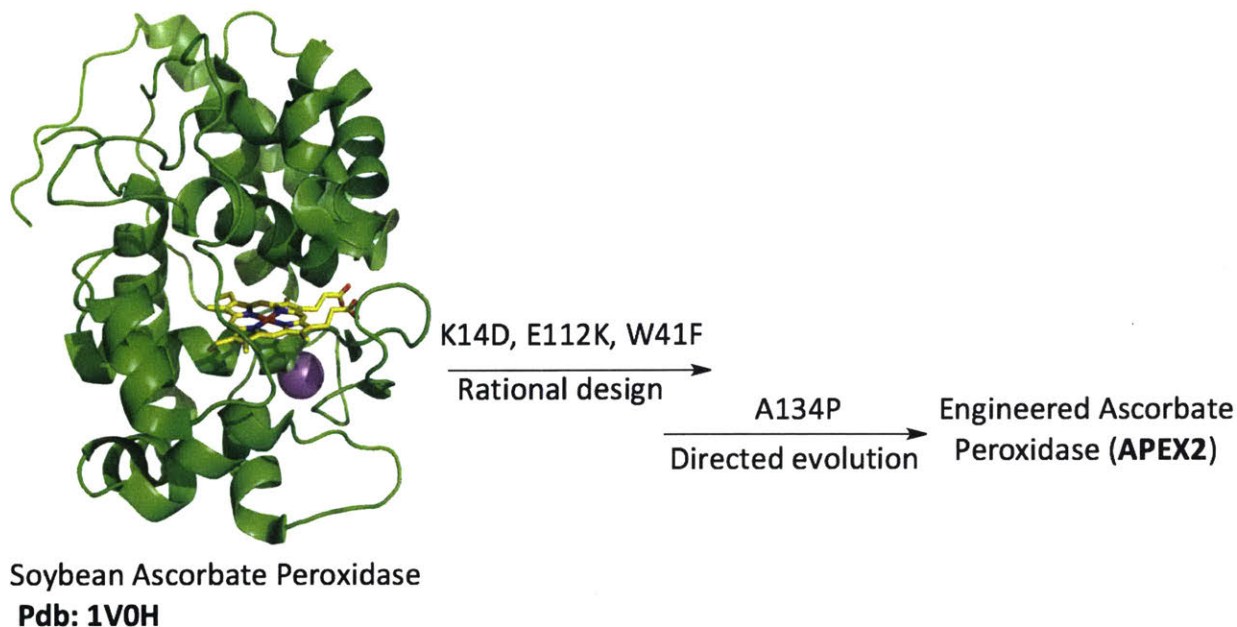


Figure 1-3: APEX2 is an engineered product of structure-guided mutagenesis and yeast display-directed evolution. The starting template of soybean APX underwent one round of rational design to monomerize and enhance peroxidase activity towards aromatic substrates. Generation 1 APEX underwent a second round of engineering where the single mutation A134P provided enhanced peroxidase activity for aromatic substrates as well. This finalized engineering product is APEX2

Proximity labeling methodology requires the use of a genetically targeted enzyme that can promiscuously covalently tag proximal endogenous proteins with a chemical handle, such as biotin, for purification and further analysis without performing any organelle purification. When catalyzed by the addition of H₂O₂, APEX2 uses its enhanced peroxidase activity to perform one electron oxidation on the aromatic reducing substrate, biotin-phenol, and generates a short-lived biotin-phenoxy radical that can diffuse out of the active site to react with proximal surface

exposed electron side chains such as tyrosine residues to covalently attached a biotin. The generated biotin-phenoxy radical product is very reactive has a half-life < 1 ms, resulting in a small labeling radius of APEX2 (< 20 nm) and only proteins proximal to APEX2 will be sufficiently tagged¹¹. Furthermore, proximity labeling experiments by APEX2 are performed with high temporal resolution. APEX2, with a $k_{cat} = 299$ s⁻¹, can generate sufficient quantities of biotin-phenoxy radicals to tag proximal proteins within 1 minute. This feature of APEX2 enables dynamic analysis of protein interaction networks, such as the mapping of GPCR signaling pathways in real time^{12,13}. Furthermore, the biotin-phenoxy radical not only reacts with proteins, but also nucleic acids, enabling mapping of the spatial distribution of endogenous RNAs throughout the cell¹⁴. This feature of APEX2, in addition to its versatile ability to catalyze the H₂O₂-dependent one-electron oxidation of a wide variety of small molecule substrates, has led to its widespread use for a variety of applications, including proteomic mapping of organelles^{11,15-18}, proximity labeling of protein interactomes^{12,13,19}, spatial mapping of cellular RNA¹⁴, electron microscopy^{2,20-25}, H₂O₂ sensing²⁶, and protein topology determination^{2,11,25}.

Two other types of enzymes have been commonly utilized for proximity labeling of proteins, promiscuous biotin ligase and HRP. Roux and coworkers used a promiscuous mutant of *Escherichia coli* biotin ligase – pBirA. (BirA + R118G = pBirA) in a method called BioID to biotinylate proteins proximal to the pBirA fusion to obtain proteomes of the mammalian nuclear lamina, trypanosome biolobe, nuclear pore, and centriole²⁷⁻³¹. The proposed mechanism is that pBirA accepts biotin and ATP to generate a biotin adenylate ester, which diffuses from the active site to biotinylate lysine residues of proximal proteins. However, this enzyme has slow kinetics: the rate-limiting step, the dissociation of the biotin adenylate ester from the active site, has a $k_{off} = 0.119$ s⁻¹. pBirA requires 18-24 hours of labeling to generate product levels comparable to one minute of APEX2 labeling^{25,27}. This extended biotin incubation time could lead to toxicity. Furthermore, the biotin-adenylate ester has a half-life of minutes, yielding a larger labeling radius and further loss of spatiotemporal resolution when compared to phenoxy radicals generated by APEX2¹¹. Additionally, due to the need for such length labeling times, the BioID methodology is limited when investigating dynamic biological processes.

Unlike pBirA, HRP has much faster kinetics for aromatic substrates. The substrate binding pocket of HRP contains a hydrophobic patch with 4 phenylalanines, thus inspiring the W41F mutation in APX. This hydrophobic region in the substrate binding pocket, combined with its

highly solvent exposure nature may explain HRPs robust activity in oxidizing a broad array of aromatic substrates^{32,33}. However, HRP has four disulfide bonds and requires Ca^{2+} ^{11,34,35}. As a result, HRP is inactive in the reducing and Ca^{2+} scarce environment of the cytosol. Contrastingly, APEX2 is active within the cytosol since it does not require disulfide bonds or Ca^{2+} ^{2,11}.

Applications

In general, the use of APEX2 begins with fusing it to a protein or peptide in order to target it to a subcellular region or macromolecular complex of interest. For instance, we have targeted APEX2 to the outer mitochondrial membrane (OMM) and the endoplasmic reticulum membrane (ER membrane, or ERM) of mammalian cells by fusing the APEX2 gene to the C-terminal 31 amino acids of the native OMM mitochondrial antiviral signaling protein (MAVS), or to the N-terminal 27 amino acids of the native ERM protein P450 oxidase 2C1, respectively^{16,25}. These constructs were used for both EM²⁵ and proteomic analysis¹⁶ of the OMM and ERM environment. In the presence of hydrogen peroxide and DAB, APEX2 catalyzes the polymerization of DAB in

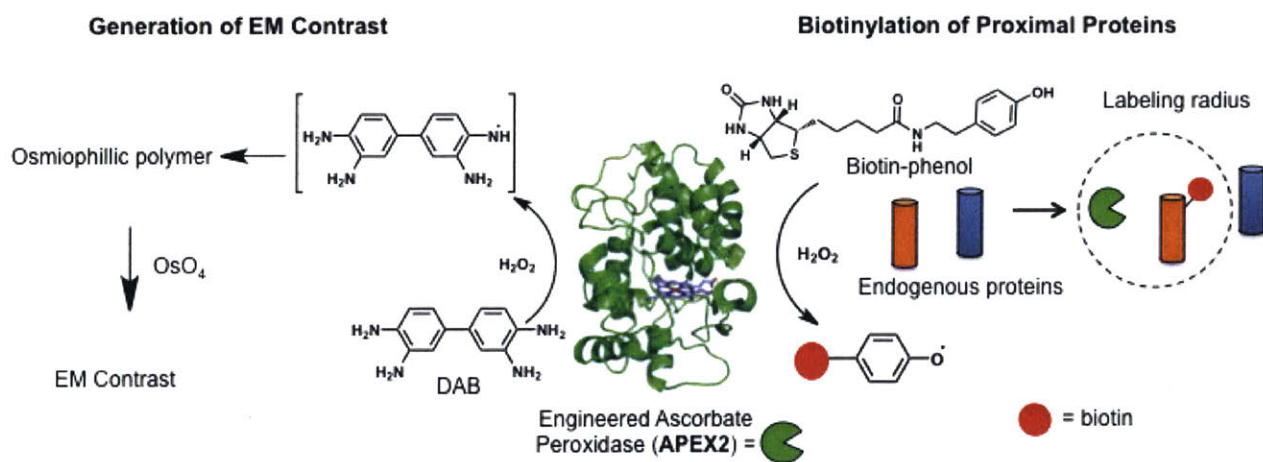


Figure 1-4: APEX2 can be utilized as both a reporter for EM and proteomics. In the presence of diaminobenzidine (DAB) and H_2O_2 , APEX2 can oxidize DAB, initiating oxidative polymerization to form a locally deposited precipitate that provides EM contrast with subsequent treatment of osmium. For proteomic applications, in the presence of biotin-phenol and H_2O_2 , APEX2 generates a biotin phenoxyl radical that proximally labels proteins. The covalently attached biotin allows for streptavidin-enrichment of these proximal proteins for further analysis.

fixed cells². The localized DAB polymer gives EM contrast upon further treatment with electron-rich osmium⁴. Alternatively, in the presence of hydrogen peroxide and biotin-phenol, APEX2 in live cells generates a short-lived, membrane-impermeable phenoxyl radical that biotinylates surface-exposed tyrosine residues in a proximity dependent manner^{11,25} (**Figure 1-4**). Biotinylated proteins can then be enriched and identified utilizing tandem MS¹⁵.

While this strategy has provided access to many cellular regions and organelles, there are numerous compartments and structures that cannot be cleanly accessed by such an approach. This technique is limited by the specificity of genetic targeting; there are cellular regions that cannot be exclusively targeted by a single genetic tag. For example, there is great interest in the biology of organelle-organelle contact sites, such as the junctions between mitochondria and ER, which participate in calcium signaling^{36,37}, lipid synthesis³⁸⁻⁴¹, and mitochondrial fission^{42,43}. These junctions have been visualized by microscopy⁹, and the complex that tethers the two organelles has been identified in yeast as the ER-mitochondria encounter structure (ERMES)⁴⁴.

Yet all APEX2 fusion constructs we have evaluated or considered, such as to the proteins dynamin-1-like protein, Drp1⁴², mitofusin2, Mfn2⁴⁵⁻⁴⁷, synaptojanin 2 binding protein, SYNJ2BP1¹⁶, and PDZ domain containing protein 8, PDZD8⁴⁸ would also target the peroxidase to protein pools outside of mito-ER contacts, such as into the cytosol⁴⁹, along the cytoskeleton⁵⁰, and over the entire OMM¹⁶. As a result, one cannot use conventional APEX2 fusion approaches to specifically target the mito-ER junctions.

Another application for which the APEX2 genetic fusion strategy may be unsuitable is for profiling the interactome of specific cellular RNAs. While there are several robust methods to identify RNA interaction partners for specific proteins of interest⁵¹⁻⁵³, the converse problem—identifying proteins that interact with a particular RNA species—is much more challenging. One could envision fusing the APEX2 protein to a high-affinity RNA-binding protein (RBP; for example, the bacteriophage MS2 coat protein⁵⁴) allowing the peroxidase to be ectopically targeted to transcripts that are tagged with that RBP's cognate RNA motif. However, a major technical challenge would be the large pool of excess, catalytically active APEX2-RBP fusion protein that is not docked to the tagged RNA, resulting in background labeling that masks the specific signal.

A general solution to these problems and related ones could be a split form of APEX2, in which two inactive fragments of APEX2 reconstitute to give an active peroxidase. One could use an intersectional approach to restrict APEX2 activity specifically to sites of interest, such as mito-ER contacts only, or specific RNA binding sites only – thus eliminating the background labeling from protein overexpression and off-targeting.

Splitting APEX to make a protein complementation assay (PCA)

Split protein sensors, also referred to as protein fragmentation complementation assays (PCA), are employed to improve specificity for a variety of applications from protein-protein interaction (PPI) detection to genome editing^{55,56}. This strategy relies on the conditional reconstitution of enzymatic activity using a protein that has been strategically split into two fragments⁵⁵. This principle was first observed in 1958 when Richards noted that the two fragments of Ribonuclease, generated by the first subtilisin proteolysis step, results in almost native level enzymatic activity⁵⁷. In a PCA, a reporter protein is split into two inactive fragments, and each fragment is genetically fused to a separate protein of interest (**Figure 1-5**). Interaction of the two proteins of interest drives the reconstitution of the reporter fragments, restoring protein activity.

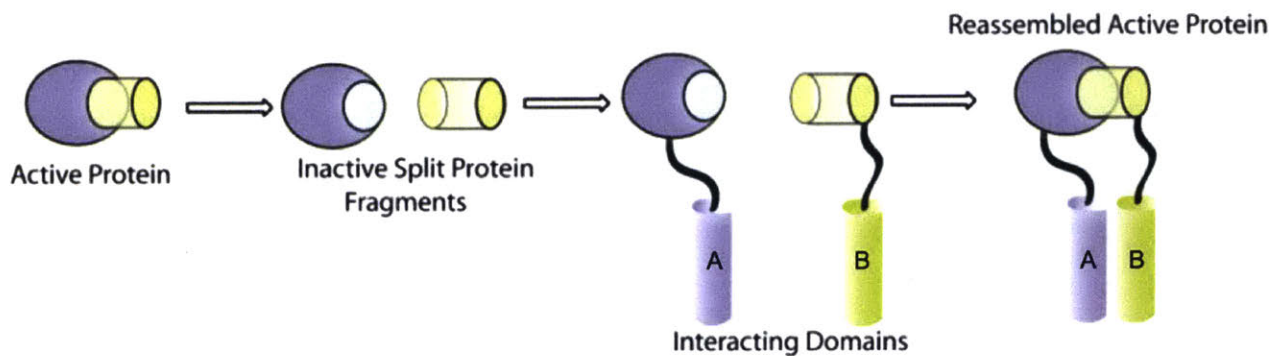


Figure 1-5. Overview of protein complementation assays (PCAs) A reporter protein is divided into two inactive fragments, and these fragments are each fused to one protein, A or B, in a pair of proteins whose interaction one wishes to detect. In the absence of a PPI, the reporter fragments remain unfolded, and no reporter protein signal is observed. When a PPI occurs between A and B, it increases the local concentration of the fragments and drives their refolding into a reconstituted

complex with the same activity as the parent reporter protein. Figure is adapted from *Ghosh, et al. in Split-Protein systems: Beyond Binary Protein-Protein Interactions*⁵⁵

Split ubiquitin was the first intentionally designed split protein system; fragments only reconstituted when fused to interacting protein pairs due to the increase in local concentration⁵⁸. The reassembled ubiquitin is then cleaved after the C-terminus, which provides a read out for the original non-covalent protein-protein interaction²⁶. Since then, split proteins and split enzymes have been developed for many workhorses of cell biology, including green fluorescent protein^{59,60}, HRP⁶¹, dihydrofolate reductase⁶², ubiquitin⁵⁸, luciferase⁶³⁻⁶⁵, beta-galactosidase⁶⁶⁻⁶⁹, TEV protease⁷⁰, and Cas9^{56,71-73}. These diverse split proteins have enzymatic activities that include site-specific proteolysis, cleavage of reconstituted ubiquitin, genome editing, and provide fluorescence, positron emission tomography, and host survival readouts²³. Many applications of these split protein systems are driven together by *in vivo* macromolecular interactions to screen for interactions²³. Other applications extend the utility of split proteins via temporally-controlled reconstitution through fusion with FK506 binding protein (FKBP) and FKBP-12 rapamycin binding domain (FRB), which undergoes chemically inducible dimerization upon addition of rapamycin²³. Our lab has initiated efforts to split both APEX2 and HRP. A former graduate student, Jeff Martell, has successfully published on his engineered split HRP. Jeff also started to split first generation APEX by screening and assaying several rationally selected cut sites. Jeff and I have collaborated to rationally select and test APEX2 cut sites (Chapter 2).

Discussion

Splitting APEX2 presents new challenges, however. First, APEX2 requires a heme cofactor for its activity, and most cut sites would split apart the heme-binding pocket. Second, in order for split APEX2 to be useful for a broad range of applications, the inactive fragments should have relatively low affinity for one another, such that reconstitution only occurs when the fragments are driven together by a molecular interaction. Not many known split proteins have low-affinity fragments, and it is challenging to engineer such a property in conjunction with high activity upon reconstitution.

The ideal successful split APEX2 system requires two low affinity fragments that reconstitute enzymatic activity rapidly only in a proximity dependent manner. For instance, for an

application at the organelle-organelle contact site, mito-ER, each fragment alone must not only remain monomeric and inactive, but also avoid perturbing the localization and expression of native mitochondrial and ER proteins. They also must be well-folded and non-aggregating for proper trafficking and localization. Their mutual low affinity is necessary to prevent artificially induced mito-ER junctions.

Short enzymatic reconstitution times are also desired, to further ensures both organelles' morphological integrity. Thus, it is critical that split APEX2 reconstitute peroxidase activity both rapidly and robustly. The development of split APEX2 requires systematic screening to determine an appropriate fragmentation site, as well as yeast-display directed evolution to improve expression, enzymatic activity, and dynamic range.

To address the need for an interaction-dependent proximity labeling tool, this thesis describes the development of a split APEX2 system. We used a combination of rational design (Chapter 2) and a novel yeast display-based directed evolution approach that incorporates positive and negative selection steps (Chapter 3) to evolve sAPEX in order to obtain targeting of regions that are not membrane bound and/or cannot be targeted by a single gene fusion. Our engineering efforts focused on attempting to minimize background association of the peptide fragments while maintaining high peroxidase activity upon reconstitution. Our resulting split APEX2 fragment pair has significantly diverged from its parental sequence and shows interaction-dependent reconstitution in multiple contexts in living mammalian cells (Chapter 4). Comparison against competitor split APEX⁷³ demonstrated the superiority of the reconstitution of our split APEX tool after directed evolution (Chapter 4). Our split APEX tool adds to the proximity labeling toolkit (Chapter 5 and 6), and in the future, should extend the utility of APEX-based approaches to new areas of biology at higher spatiotemporal resolution.

References

1. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
2. Gumiero, A., Metcalfe, C. L., Pearson, A. R., Raven, E. L. & Moody, P. C. E. Nature of the ferryl heme in compounds I and II. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.183483

3. Jones, D. K., Dalton, D. A., Rosell, F. I. & Raven, E. L. Class I heme peroxidases: Characterization of soybean ascorbate peroxidase. *Arch. Biochem. Biophys.* (1998). doi:10.1006/abbi.1998.0941
4. Sharp, K. H., Mewies, M., Moody, P. C. E. & Raven, E. L. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat. Struct. Biol.* **10**, 303–307 (2003).
5. Lad, L., Mewies, M. & Raven, E. L. Substrate binding and catalytic mechanism in ascorbate peroxidase: Evidence for two ascorbate binding sites. *Biochemistry* (2002). doi:10.1021/bi0261591
6. Gumiero, A., Murphy, E. J., Metcalfe, C. L., Moody, P. C. E. & Raven, E. L. An analysis of substrate binding interactions in the heme peroxidase enzymes: A structural perspective. *Archives of Biochemistry and Biophysics* (2010). doi:10.1016/j.abb.2010.02.015
7. Efimov, I. *et al.* The redox properties of ascorbate peroxidase. *Biochemistry* (2007). doi:10.1021/bi7006492
8. Macdonald, I. K., Badyal, S. K., Ghamsari, L., Moody, P. C. E. & Raven, E. L. Interaction of ascorbate peroxidase with substrates: A mechanistic and structural analysis. *Biochemistry* (2006). doi:10.1021/bi0606849
9. Veitch, N. C. Horseradish peroxidase: A modern view of a classic enzyme. *Phytochemistry* (2004). doi:10.1016/j.phytochem.2003.10.022
10. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science (80-.)*. **339**, 1328–1331 (2013).
11. Paek, J. *et al.* Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell* **169**, 338–349.e11 (2017).
12. Lobingier, B. T. *et al.* An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell* **169**, 350–360.e12 (2017).
13. Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L. & Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *doi.org* 153098 (2017). doi:10.1101/153098
14. Hung, V. *et al.* Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol. Cell* **55**, 332–341 (2014).
15. Hung, V. *et al.* Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* **6**, (2017).
16. Loh, K. H. *et al.* Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. *Cell* **166**, 1295–1307.e21 (2016).
17. Chen, C.-L. *et al.* Proteomic mapping in live Drosophila tissues using an engineered ascorbate peroxidase. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1–6 (2015).
18. Han, S. *et al.* Proximity Biotinylation as a Method for Mapping Proteins Associated with mtDNA in Living Cells. *Cell Chem. Biol.* **24**, 404–414 (2017).
19. Shvets, E., Bitsikas, V., Howard, G., Hansen, C. G. & Nichols, B. J. Dynamic caveolae exclude bulk membrane proteins and are required for sorting of excess glycosphingolipids. *Nat. Commun.* **6**, (2015).
20. Ludwig, A., Nichols, B. J. & Sandin, S. Architecture of the caveolar coat complex. *J. Cell Sci.* jcs.191262 (2016). doi:10.1242/jcs.191262
21. Liu, L. K., Choudhary, V., Toulmay, A. & Prinz, W. A. An inducible ER-Golgi tether facilitates ceramide transport to alleviate lipotoxicity. *J. Cell Biol.* **216**, 131–147 (2017).
22. Hyenne, V. *et al.* RAL-1 controls multivesicular body biogenesis and exosome secretion. *J. Cell Biol.* **211**, 27–37 (2015).

23. Joesch, M. *et al.* Reconstruction of genetically identified neurons imaged by serial-section electron microscopy. *Elife* **5**, (2016).
24. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).
25. Dwyer, D. J. *et al.* Antibiotics induce redox-related physiological alterations as part of their lethality. *Proc. Natl. Acad. Sci.* **111**, E2100–E2109 (2014).
26. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* (2012). doi:10.1083/jcb.201112098
27. Belousov, V. V. *et al.* Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nat. Methods* (2006). doi:10.1038/nmeth866
28. Li, X. W. *et al.* New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. *J. Biol. Chem.* (2014). doi:10.1074/jbc.M113.529578
29. Lee, N., Moss, W. N., Yario, T. A. & Steitz, J. A. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160**, 607–618 (2015).
30. Adams, J. C. Biotin amplification of biotin and horseradish peroxidase signals in histochemical stains. *J. Histochem. Cytochem.* (1992). doi:10.1177/40.10.1527370
31. Gajhede, M., Schuller, D. J., Henriksen, A., Smith, A. T. & Poulos, T. L. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. *Nat. Struct. Biol.* (1997). doi:10.1038/nsb1297-1032
32. Veitch, N. C., Gao, Y., Smith, A. T. & White, C. G. Identification of a critical phenylalanine residue in horseradish peroxidase, Phe179, by site-directed mutagenesis and 1H-NMR: Implications for complex formation with aromatic donor molecules. *Biochemistry* (1997). doi:10.1021/bi9718402
33. Jiang, S. *et al.* A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. *Proteomics* (2012). doi:10.1002/pmic.201100551
34. Hopkins, C., Gibson, A., Stinchcombe, J. & Futter, C. Chimeric molecules employing horseradish peroxidase as reporter enzyme for protein localization in the electron microscope. *Methods Enzymol.* (2000). doi:10.1016/S0076-6879(00)27265-0
35. Szabadkai, G. *et al.* Chaperone-mediated coupling of endoplasmic reticulum and mitochondrial Ca²⁺ channels. *J. Cell Biol.* **175**, 901–911 (2006).
36. Rizzuto, R. *et al.* Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca²⁺ responses. *Science (80-.).* **280**, 1763–1766 (1998).
37. Kornmann, B., Osman, C. & Walter, P. The conserved GTPase Gem1 regulates endoplasmic reticulum-mitochondria connections. *Proc. Natl. Acad. Sci.* **108**, 14151–14156 (2011).
38. Kornmann, B. *et al.* An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science (80-.).* **325**, 477–481 (2009).
39. Lewin, T. M., Van Horn, C. G., Krisans, S. K. & Coleman, R. A. Rat liver acyl-CoA synthetase 4 is a peripheral-membrane protein located in two distinct subcellular organelles, peroxisomes, and mitochondrial-associated membrane. *Arch. Biochem. Biophys.* **404**, 263–270 (2002).
40. Rusiñol, A. E., Cui, Z., Chen, M. H. & Vance, J. E. A unique mitochondria-associated membrane fraction from rat liver has a high capacity for lipid synthesis and contains pre-

- Golgi secretory proteins including nascent lipoproteins. *J. Biol. Chem.* **269**, 27494–27502 (1994).
41. Friedman, J. R. *et al.* ER Tubules Mark Sites of Mitochondrial Division. *Science* (80-). **334**, 358–362 (2011).
 42. Murley, A. *et al.* ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. *Elife* **2013**, (2013).
 43. Murley, A. *et al.* ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. *Elife* **2013**, (2013).
 44. De Brito, O. M. & Scorrano, L. Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* **456**, 605–610 (2008).
 45. Filadi, R. *et al.* Mitofusin 2 ablation increases endoplasmic reticulum–mitochondria coupling. *Proc. Natl. Acad. Sci.* **112**, E2174–E2181 (2015).
 46. Cosson, P., Marchetti, A., Ravazzola, M. & Orci, L. Mitofusin-2 Independent Juxtaposition of Endoplasmic Reticulum and Mitochondria: An Ultrastructural Study. *PLoS One* **7**, (2012).
 47. Hirabayashi, Y. *et al.* ER-mitochondria tethering by PDZD8 regulates Ca²⁺ dynamics in mammalian neurons. *Science* (80-). **358**, 623–630 (2017).
 48. Smirnova, E., Griparic, L., Shurland, D.-L. & Blik, A. M. van der. Dynamin-related Protein Drp1 Is Required for Mitochondrial Division in Mammalian Cells. *Mol. Biol. Cell* **12**, 2245–2256 (2001).
 49. Henning, M. S. *et al.* PDZD8 is a novel moesin-interacting cytoskeletal regulatory protein that suppresses infection by herpes simplex virus type 1. *Virology* **415**, 114–121 (2011).
 50. Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, (2016).
 51. Sibley, C. R. Individual nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) to determine protein–RNA interactions. in *Methods in Molecular Biology* **1649**, 427–454 (2018).
 52. Garzia, A., Morozov, P., Sajek, M., Meyer, C. & Tuschl, T. PAR-CLIP for discovering target sites of RNA-binding proteins. in *Methods in Molecular Biology* **1720**, 55–75 (2018).
 53. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).
 54. Shekhawat, S. S. & Ghosh, I. Split-protein systems: Beyond binary protein-protein interactions. *Current Opinion in Chemical Biology* (2011). doi:10.1016/j.cbpa.2011.10.014
 55. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology* **33**, 139–142 (2015).
 56. Richards, F. M. ON THE ENZYMIC ACTIVITY OF SUBTILISIN-MODIFIED RIBONUCLEASE. *Proc. Natl. Acad. Sci.* (1958). doi:10.1073/pnas.44.2.162
 57. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10340–4 (1994).
 58. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
 59. Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein [12]. *Journal of the American Chemical Society* **122**, 5658–5659 (2000).

60. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).
61. Pelletier, J. N., Campbell-Valois, F.-X. & Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci.* **95**, 12141–12146 (1998).
62. Paulmurugant, R. & Gambhir, S. S. Monitoring protein-protein interactions using split synthetic renilla luciferase protein-fragment-assisted complementation. *Anal. Chem.* **75**, 1584–1589 (2003).
63. 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.* **101**, 12288–12293 (2004).
64. Ozawa, T., Kaihara, a, Sato, M., Tachihara, K. & Umezawa, Y. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal. Chem.* **73**, 2516–2521 (2001).
65. Olson, K. R. & Eglén, R. M. β Galactosidase Complementation: A Cell-Based Luminescent Assay Platform for Drug Discovery. *Assay Drug Dev. Technol.* **5**, 137–144 (2007).
66. Broome, A. M., Bhavsar, N., Ramamurthy, G., Newton, G. & Basilion, J. P. Expanding the utility of β -galactosidase complementation: Piece by piece. *Mol. Pharm.* **7**, 60–74 (2010).
67. Ullmann, A., Jacob, F. & Monod, J. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**, 339–343 (1967).
68. Rossi, F., Charlton, C. a & Blau, H. M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8405–10 (1997).
69. Wehr, M. C. *et al.* Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods* **3**, 985–993 (2006).
70. Wright, A. V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci.* **112**, 2984–2989 (2015).
71. Truong, D. J. J. *et al.* Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458 (2015).
72. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).
73. Xue, M. *et al.* Optimizing the fragment complementation of APEX2 for detection of specific protein-protein interactions in live cells. *Sci. Rep.* **7**, (2017).

Chapter 2: Rationally-selected screening of APEX2 cut sites

Introduction

We first sought to identify potentially fruitful cut sites in the APEX2 enzyme; previously, Jeff Martell had performed a cut site screen with first-generation APEX. Together with his guidance, we rationally selected possible fragmentations of APEX2 for screening. There were several considerations in picking which cut sites to screen. First, cut sites were not made within regions of known secondary structure as the disruption could lead to aggregation and/or degradation. Second, with the exception of cut sites on the first N-terminal loop, all fragments lack a fully formed heme-binding pocket. Third, the new N and C-termini generated from cut sites have side chains that do not contain hydrophobic residues to avoid collapse and burial of the termini and significant alteration of secondary structure. In the first round, we scanned all potential loop regions; the second round focused on the successful loops and more exhaustively screening cut sites along those loops.

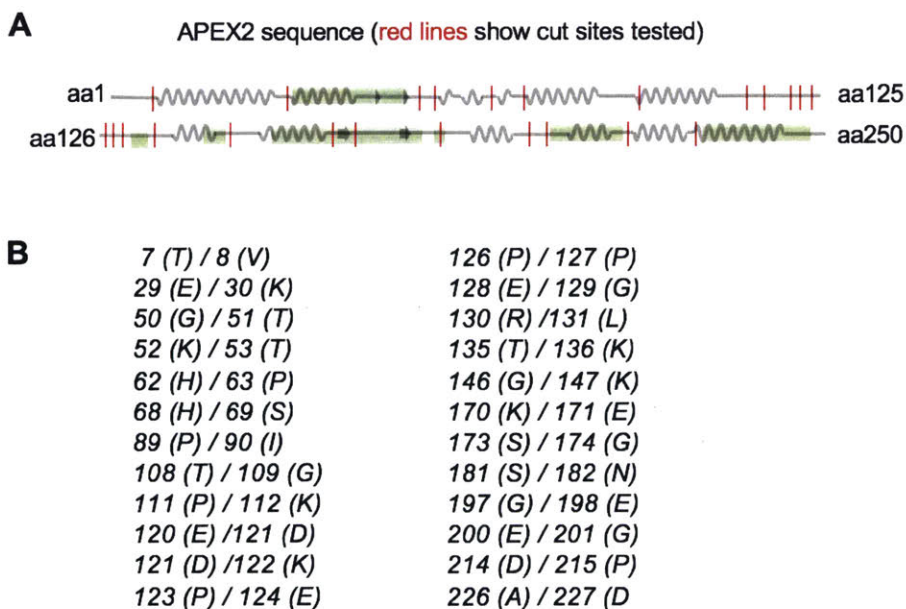


Figure 2-1. Selection of potential split. APEX cut sites. (A) The first screen tested 24 different cut sites. Their locations in the APEX2 protein sequence are indicated by the red vertical lines. Squiggles denote alpha helices. Grey arrows denote beta sheets. Areas shaded green are part of the heme-binding pocket. (B) The 24 different cut sites; newly generated N- and C- terminal amino acid identity abbreviated in parenthesis.

Amplex UltraRed Screen of rationally-selected APEX2 cut sites for enzymatic reconstitution in HEK 293T cells

Using these guidelines, we selected 24 preliminary cut sites within solvent-exposed loops and turns between secondary structural elements (alpha helices and beta sheets) based on the crystal structure of wild-type ascorbate peroxidase⁵ (**Figure 2-1**). We sought to find the best cut sites in the APEX2 enzyme using a chemically-inducible protein association system as a test platform. We cloned each of the 24 fragment pairs as fusions to FKBP and FRB, whose interaction can be induced with the small molecule rapamycin (**Figure 2-2a, 2-2b**). We introduced the fragment pairs into HEK 293T cells and evaluated peroxidase activity in the presence or absence of rapamycin. Catalytic activity was measured using an established assay based on the membrane-permeable fluorogenic probe Amplex UltraRed which is colorless, but upon peroxidase-catalyzed oxidization, generates the red-fluorophore Resorufin^{74,75} (**Figure 2-2c**). Ideal split APEX2 candidates are those that display strong Resorufin signal only with the addition of rapamycin, indicating low affinity fragments and robust reconstitution.

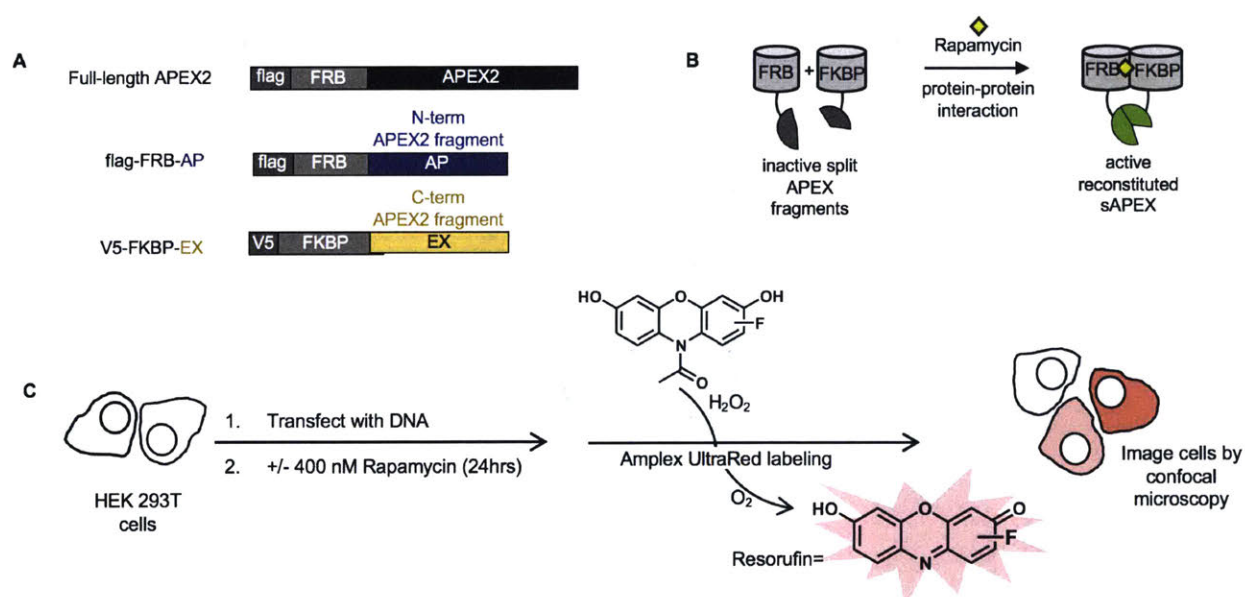


Figure 2-2. split APEX cut site screening platform. (A) N- and C-terminal sAPEX fragments selected for testing were fused to FRB and FKBP, respectively. Full length APEX2 was used cloned and used as both a benchmark and a positive control. (B) Schematic overview of split

APEX (sAPEX). Two inactive fragments (grey) can reconstitute to give active peroxidase (green) when driven together by a protein-protein interaction (PPI) such as that of FKBP and FRB under the presence of rapamycin, yellow diamond. (C) Cartoon of the Amplex UltraRed assay – Pairs of constructs were introduced into HEK 293T cells by transient transfection, along with a CFP-NLS (nuclear localization signal) co-transfection marker. Cells were either treated with rapamycin for 20-24 h (left) or left untreated (right) in (Figure 2-3). Subsequently Amplex UltraRed, a fluorogenic small-molecule peroxidase substrate, and H₂O₂ were added for 25 minutes, after which cells were fixed and imaged. Resorufin is the fluorescent product of Amplex UltraRed oxidation and indicates peroxidase activity.

Of the 24 fragment pairs tested, seven produced significant Resorufin product, indicative of reconstituted activity (Figure 2-3a). With the exception cut sites 7/8 and 29/30, all of the fragments we screened are expected to lack a completely formed heme-binding pocket. Cut sites within the heme binding cavity (130/131, 134/135, 146/147, 170/171, 173/174, and 181/182) exhibited the most dramatic reduction in catalytic function; these fragment pairs completely failed to reconstitute activity (Figure 2-1a; 2-3a).

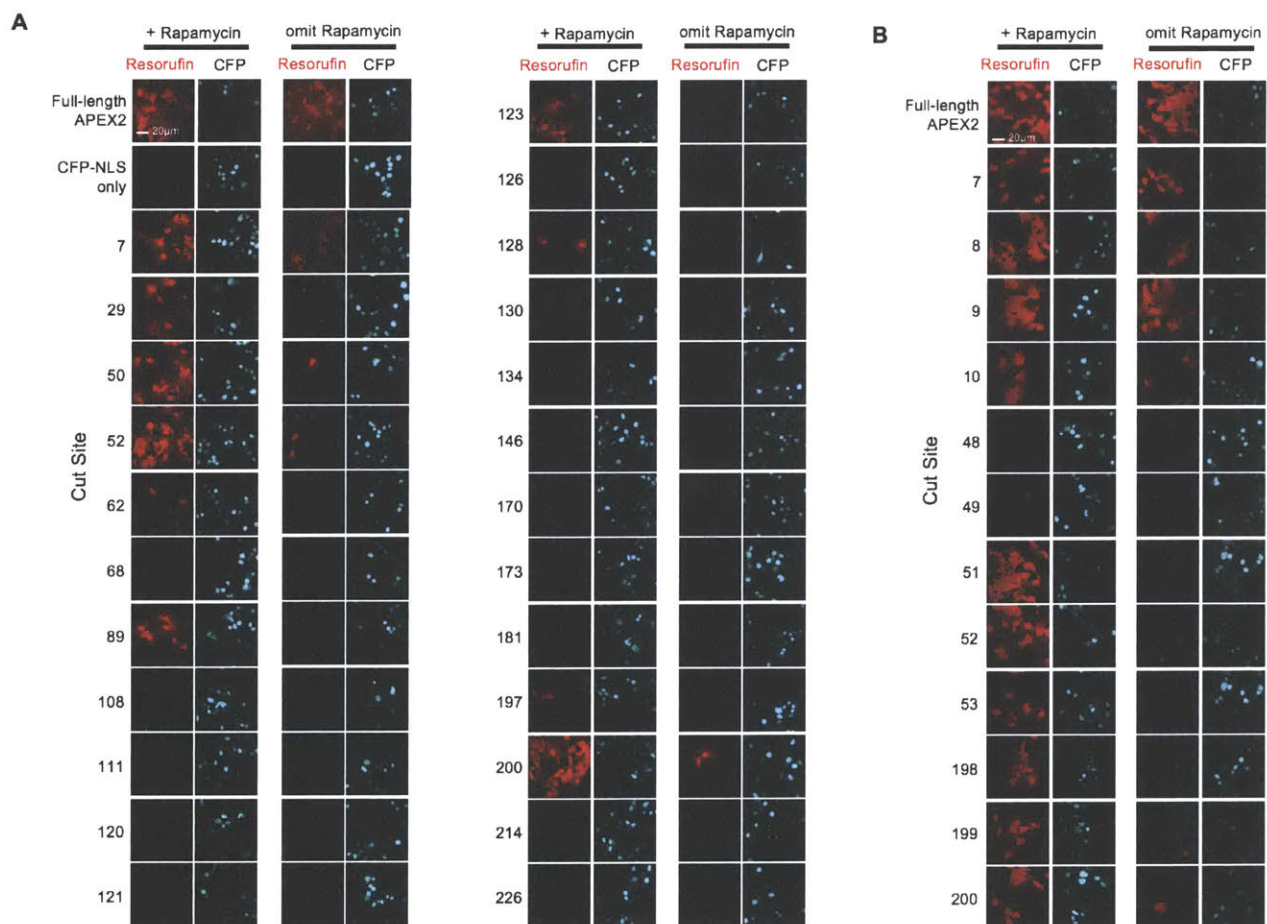


Figure 2-3. Screening of potential split APEX cut sites. (A) Initial Amplex UltraRed screen of cut sites, fragmentation occurs after the listed amino acid. For instance, cut site 7 APEX2 between aa 7 and 8. Scale bars, 20 μ m. performed. (B) Second cut site screen, focused on residues surrounding T7, G50, and E200. Same assay as in (A) with a total of two replicates.

To more finely map the optimal cut sites, we performed a second round of screens, evaluating cut sites 1-3 residues away from the most promising sites identified in our initial screen (Figures 2-3B). This yielded three optimal starting positions—51/52, 89/90, and 200/201— all of which fell in solvent-exposed loops distal (>15 angstroms away) from the APEX2 active site and heme-binding pocket. To test if any of the fragments contained peroxidase activity on its own, fragment pairs that generated the most robust resorufin signals were tested with single fragment controls. Omission of either fragment showed that the fragments alone lack any detectable peroxidase activity at all cut sites except for the cut site 7 (Figure 2-4).

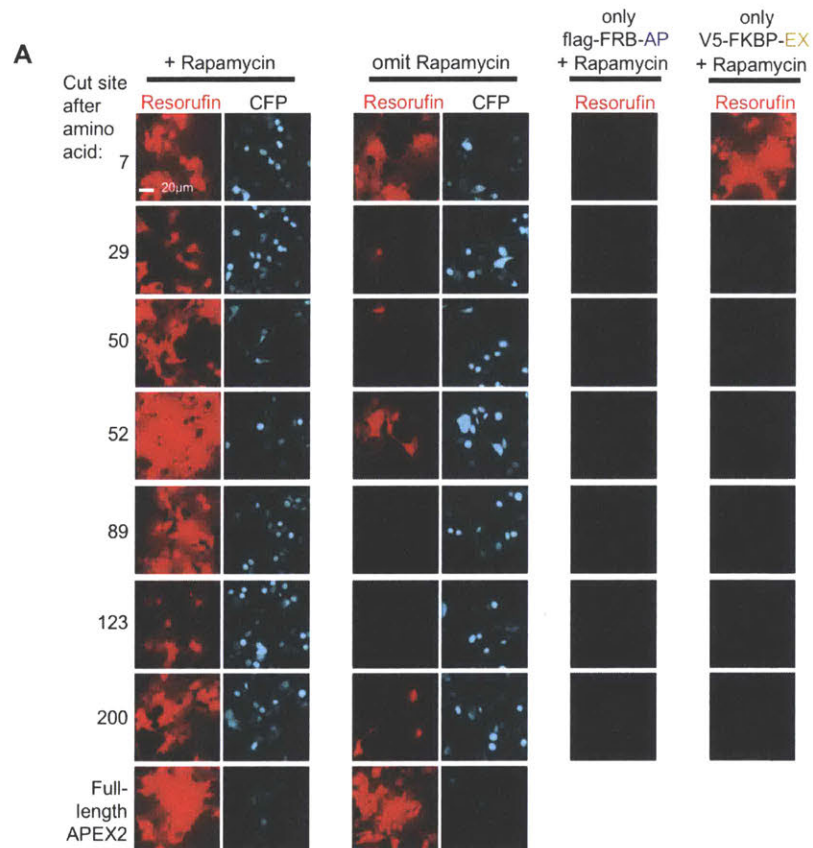


Figure 2-4. Focused Amplex UltraRed screening of potential sAPEX cut sites with single fragment controls. (A) Pairs of constructs were introduced into HEK 293T cells by transient transfection, along with a CFP-NLS (nuclear localization signal) co-transfection marker. Here, for more stringent examination of promising cut sites, each fragment was transfected for individual expression in HEK 293T cells. Cells were either treated with rapamycin for 24 h (left column, and individual fragments) or left untreated (second column). Subsequently Amplex UltraRed, a fluorogenic small-molecule peroxidase substrate, and H₂O₂ were added for 25 minutes, after which cells were fixed and imaged. Resorufin is the fluorescent product of Amplex UltraRed oxidation and indicates peroxidase activity. Scale bars, 20 μ m. The experiment has one biological replicate.

Live biotin-phenol labeling assay of potential split APEX pairs

Amplex UltraRed is an informative assay for a high-throughput screen of peroxidase activity because the enzymatically amplified generation of a fluorescent signal provides a sensitive and facile method for detecting split APEX2 reconstitution. However, in an application such as a proteomics experiment, split APEX2 would need to oxidize biotin-phenol to generate a biotin-phenoxy radical. A more representative assay to determine if a cut site has sufficient enzymatic activity would be to utilize split APEX2 to biotinylate proteins and detect the amount of biotinylation via immunostaining.

Using the more stringent biotinylation assay, HEK 293T cells were transiently transfected to express candidate split APEX pairs in the cytosol. APEX2-NES was used as positive control and cut site pairs 49 and 53 were also included in the experiment to observe the extent of detectable biotinylation from inactive and low performing reconstituted peroxidases. Cut sites 7, 10, 199 and 200 had strong neutravidin-AF647 signal but low levels were also detected in the absence of rapamycin (Figure 2-5). Expectedly, cut sites 49 and 53 had no detectable neutravidin 647 signal, indicating that these cut sites did not sufficiently reconstitute enzymatic activity. However, cut site 51, 52, 89 all had biotin phenol activity with low background. The biotin-phenol assay reaffirms that three optimal starting positions—51/52, 89/90, and 200/201 – discovered from the Amplex UltraRed assays can also oxidize the biotin-phenol substrate. However, despite fairly promising initial results in the highly sensitive Amplex UltraRed assay (**Figure 2-3; 2-4**), all three split APEX2 pairs produced far weaker BP staining than full-length APEX2 (**Figure 2-5**) and the discrepancy between full length activity and the reconstituted activities was more apparent.

While exploring the extent of split APEX2 reconstitution in the cytosol makes for facile

imaging by microscopy, the mito-ER junction as a cellular context for split APEX2 enzymatic reconstitution is more exciting (more information in chapter 5). Having determined three

A

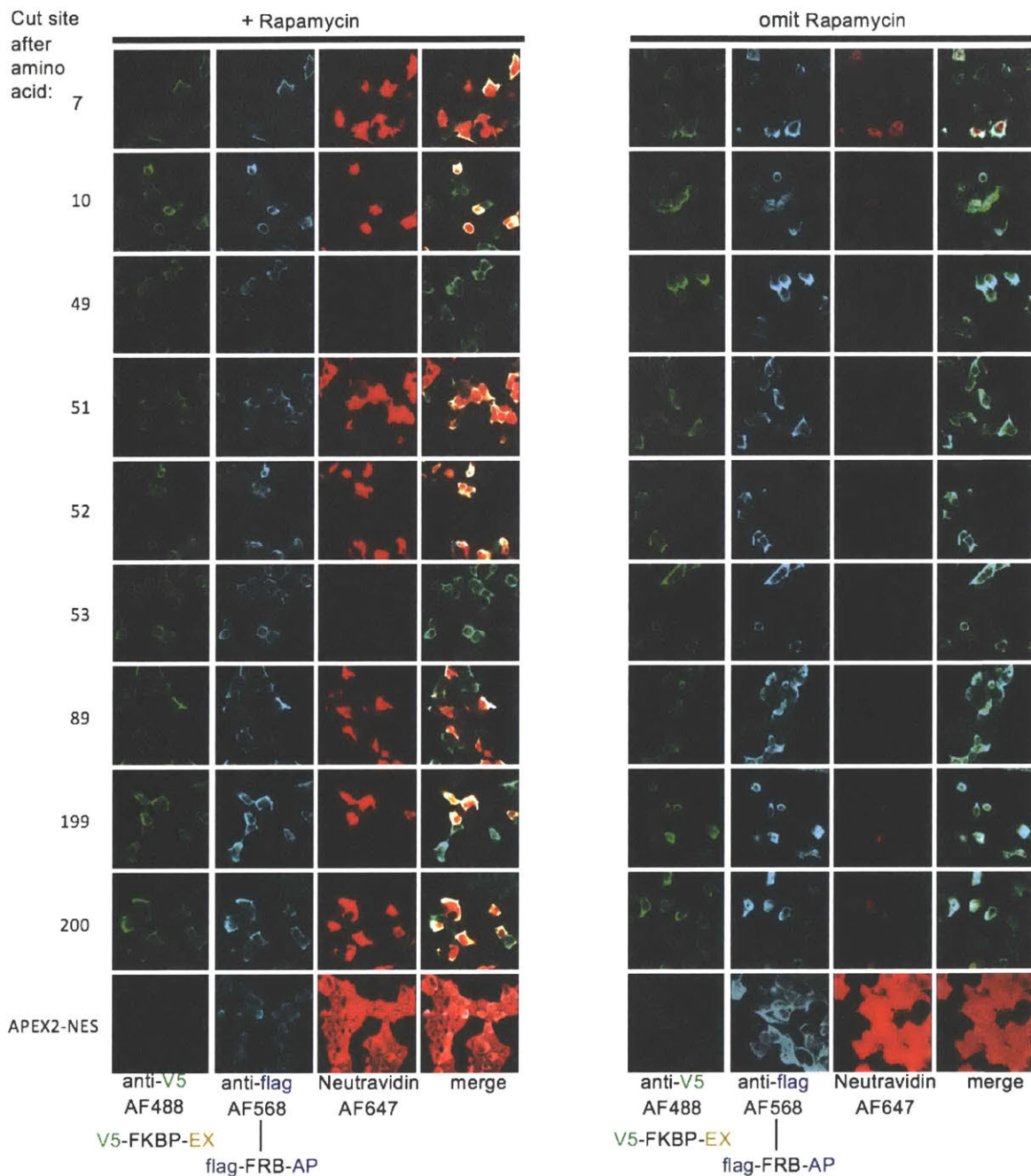


Figure 2-5. Testing split APEX pairs using biotin-phenol as substrate (A) HEK 293T cells were transfected with complementary split APEX2 fragments. After protein synthesis in media

containing +/- 400nM rapamycin for 20 hours, cells were incubated with 100 μ M biotin-phenol. Hydrogen peroxide was added ($C_f = 1$ mM) for 1 minute. Cells were immediately fixed, permeabilized, immunostained for biotinylation and expression, and then imaged via confocal microscopy. Neutravidin-647(nAv647) antibody was used to detect the extent of biotinylation achieved upon incubation in the presence and absence of rapamycin.

promising cut sites that lead to reconstitution of robust peroxidase activity in the cytosol predominately in the presence of a protein-protein interaction; these three fragment pairs were then targeted to the mitochondria and the endoplasmic reticulum. For a preliminary examination of reconstituted enzymatic activity at the mito-ER interface, the C(1-29) targeting sequence from cytochrome p450 was chosen for the ER surface membrane⁷⁶ and C-terminal MAVSTM⁷⁷ was chosen for the mito outer membrane. These targeting sequences have been commonly used within the lab for targeting fusion constructs and are known to target specifically to their designated organelles^{76,77}. Construct maps of these mito-ER split APEX2 plasmids is shown in (**Figure 2-6a**). These mito-ER targeted constructs were transfected into HEK293T cells and a live biotin-phenol labeling assay was performed. Cut site 51 and 89 yielded no detectable biotinylation in both the presence and absence of rapamycin. One possible contributing factor could be the poor expression level of the C-terminal fragment for cut site 51 and the poor expression level of both fragments for cut site 89; however, by immunostaining the C-terminal fragment for cut site 200 is poorly expressed in the absence of rapamycin and robust biotinylation is still detected.

Discussion

Two successive rounds of split APEX2 fragmentation screening using the Amplex UltraRed assay yielded three optimal starting positions—51/52, 89/90, and 200/201— all of which fell in solvent-exposed loops distal (>15 angstroms away) from the APEX2 active site. In the cytosol, these fragmentation pairs demonstrated robust reconstituted peroxidase activity in the presence of a PPI with little (or reduced for cut site 200) activity in the absence of a PPI. Importantly, these fragmentations also abolished any single fragment peroxidase activity.

These cut sites were also evaluated using a live biotin-phenol labeling assay targeting the fragments to both the cytosol and the mito-ER junction. Through both assays, cut site 200

consistently exhibited robust reconstituted peroxidase activity in the presence of a PPI, although it had detectable activity in the absence of a PPI as well. However, the less-sensitive biotin-phenol labeling assay dramatically demonstrated a clear differentiation in enzymatic activity between full length and reconstituted APEX2. When targeting all the potential split APEX2 pairs to the mito-ER junction, only cut site 200 had detectable activity, and its activity was not rapamycin dependent. These preliminary results indicate that these existing split APEX2 fragments lack sufficient sensitivity and dynamic range for desired applications.

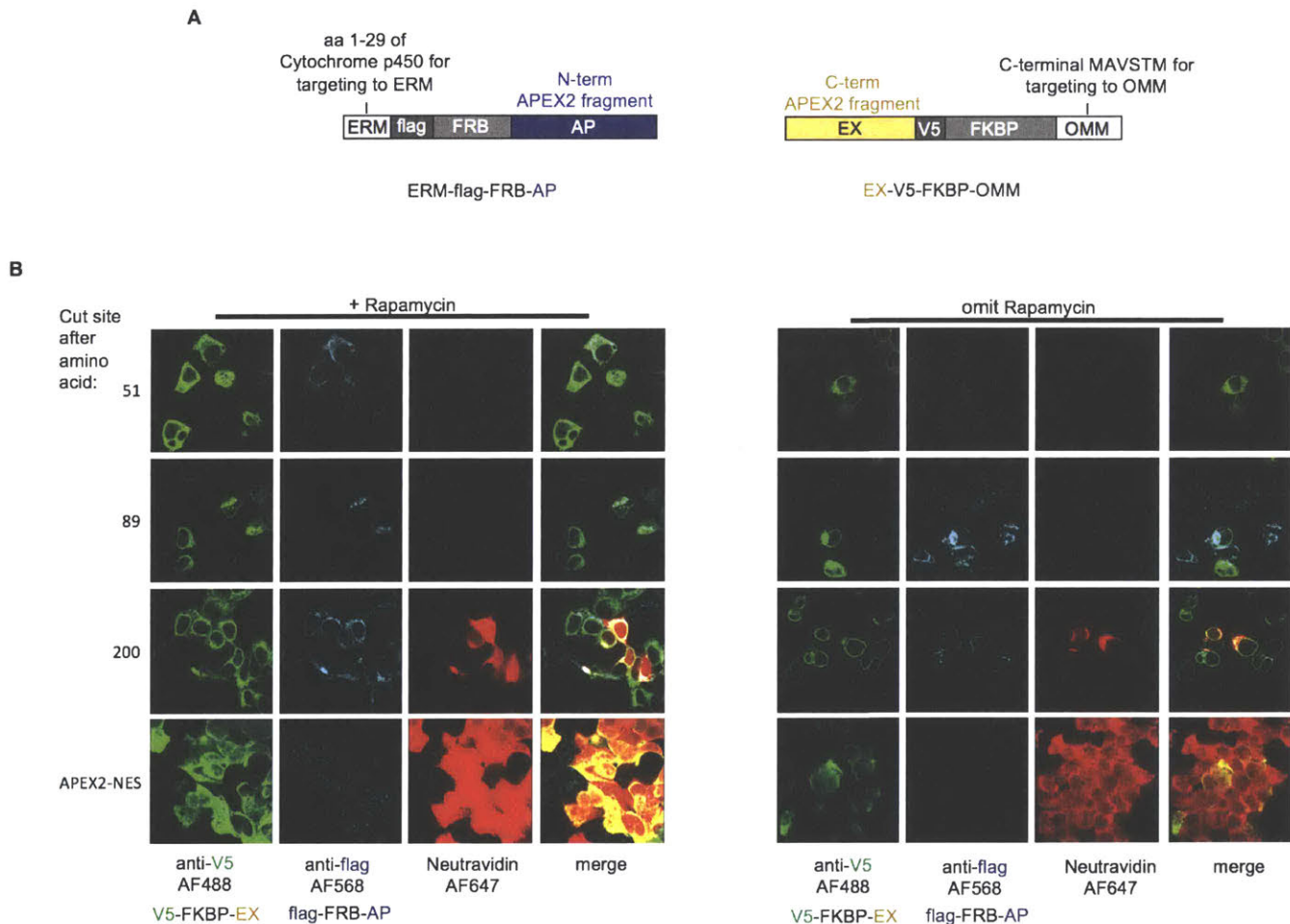


Figure 2-6. Biotin-phenol assay of the 3 most promising split APEX2 cut site pairs at the mito-ER junction (A) Constructs used for targeting split APEX fragment pairs to the ER membrane (ERM) and the outer mitochondrial membrane (OMM), respectively. (B) HEK 293T cells were transfected with complementary split APEX2 fragments. After protein synthesis in media containing +/- 400nM rapamycin for 23 hours, cells were pre-incubated for 30 minutes with biotin-phenol. After a 1 minute live labeling period, cells were immediately fixed, permeabilized,

immunostained for biotinylation and expression, and then imaged via confocal microscopy. Neutravidin-647(nAv647) antibody was used to detect the extent of biotinylation.

Experimental Methods

Mammalian cell culture and transfection

HEK 293T and COS7 cells from ATCC (passages <20) were cultured as monolayers in complete growth media consisting of 50% DMEM (Dulbecco's Modified Eagle medium, Gibco) and 50% MEM (Minimum Essential medium, Gibco) supplemented with 10% w/v FBS (Fetal Bovine Serum, Sigma), 50 units/mL penicillin, and 50 µg/mL streptomycin at 37 °C under 5% CO₂. For confocal imaging experiments, cells were grown on 7 × 7 mm glass coverslips in 48-well plates (Corning). To improve the adherence of HEK 293T cells, the glass coverslips were pretreated with 50 µg/mL fibronectin (Millipore) in Dulbecco's phosphate-buffered saline (DPBS), pH 7.4, for 20 min at 37 °C, followed by two washes with growth media prior to cell plating. Cells were transfected at 70–85% confluence using Lipofectamine 2000 (Life Technologies), with 1.0 µL Lipofectamine 2000 and 200 ng per plasmid per 300,000 cells in serum-free media for 3-4 h, after which transfection media was replaced with fresh complete growth media. Cells were labeled and/or fixed 18–24 h after transfection. The cell line was not authenticated or tested for mycoplasma.

Cloning

See *Table 2* for a list of genetic constructs used in this study, with annotated epitope tag, promoters, resistance, vector, linkers, etc. For cloning the constructs, PCR fragments were amplified using Q5 polymerase (New England BioLabs (NEB)) or PfuUltra II Fusion HS DNA polymerase (Agilent Technologies). The vectors were double-digested and ligated to gel-purified PCR products by T4 ligation or Gibson assembly. Ligated plasmid products were introduced by heat shock transformation into competent XL1-Blue bacteria. The APEX2 gene used for initial cut site screening was amplified from vimentin-APEX2 with codons optimized for mammalian expression. Mutants of AP were generated either using QuikChange mutagenesis (Stratagene) or isolated from individual yeast clones and transferred to mammalian expression vectors using standard cloning techniques.

Amplex UltraRed assay

HEK 293T cells plated on glass coverslips were transfected with 200 ng of each peroxidase fragment and 100 ng of CFP-NLS in tandem as a transfection marker for 3 h, before medium change into complete medium with or without 400 nM rapamycin (Calbiochem). Rapamycin was delivered by 1000-fold dilution into DMEM media of a 400 μ M stock in DMSO, which was maintained at -20 °C for months and thawed as needed. After 20-24 h, cells were moved to ice, washed three times with DPBS, and treated with a solution of 50 μ M Amplex UltraRed (Life Technologies) with 0.02% (6.7 mM) H_2O_2 in DPBS according to previously published protocols^{2,25,61}. Cells were labeled for 25 min on ice. The reaction was quenched by removal of the substrate solution by aspiration, and cells were subsequently washed three times with DPBS, fixed with 4.0% formaldehyde in DPBS on ice for 30 min, and washed 2 more times with DPBS before being imaged by confocal microscopy.

Live biotin-phenol labeling

Genes were introduced into HEK 293T or COS7 cells through either transient transfection with Lipofectamine 2000 or lentiviral infection. After 18–24 h (transfection) or 48 h (lentivirus), biotin-phenol (bp) labeling was initiated by changing the medium to 200 μ L of fresh growth medium containing 500 μ M bp that was sonicated for at least 5 minutes to ensure bp was fully dissolved. Cells were incubated at 37 °C under 5% CO_2 for 30 min according to previously published protocols. To initiate labeling, 2 μ L of 100 mM H_2O_2 was spiked into each well, for a final concentration of 1 mM H_2O_2 , and the plate was immediately gently agitated. To quench the reaction after 1 minute, the bp solution is aspirated, and cells were immediately fixed with 4% formaldehyde with 5 mM Trolox in PBS at room temperature for 5 minutes before continuing fixation on ice for an additional 25 minutes. Cells were then washed with chilled DPBS three times and permeabilized with pre-chilled methanol at -20 °C for 10 min. Cells were washed again three times with DPBS and blocked with 3% BSA in DPBS for 1 h to overnight with rocking at 4 °C. To detect the expression of sAPEX fusions, cells were incubated with primary antibodies mouse anti-V5 (Life Technologies, Cat. No. R96025, 1:1500 dilution) and rabbit anti-HA (Cell Signaling, 1:1000 dilution) in 1% BSA in DPBS for 1 h to overnight at 4 °C followed by 4×5 min washes with chilled DPBS. Cells were then incubated with a 1% BSA in DPBS solution containing

secondary Alexa Fluor 488-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11001, 1:1000 dilution), Alexa Fluor 568-goat anti-mouse IgG (Life Technologies, Cat. No. A-11004, 1:1000 dilution), and homemade streptavidin–Alexa Fluor 647 (1:1000) for 25-45min at 4 °C with rocking. Cells were then washed 4 times for 5 min with chilled DPBS and imaged by confocal microscopy.

Microscopy

Confocal imaging was performed on a Zeiss AxioObserver inverted confocal microscope with 10× air and 40× oil-immersion objectives, outfitted with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), 561 (DPSS) and 640-nm (diode) lasers (all 50 mW). All images were acquired and processed using Slidebook 5.0 or 6.0 software (Intelligent Imaging Innovations), through a 48× or 63× oil-immersion objective CFP(405 laser excitation; 635/85 emission), Alexa Fluor 488 (491 laser excitation; 528/38 emission), Alexa Fluor 568 (561 laser excitation; 617/73 emission), Resorufin (491 laser excitation; 550/585), Alexa Fluor 647 (647 excitation; 680/30 emission), and differential interference contrast (DIC). Acquisition times ranged from 100 to 1000 ms. DAB labeled cells were imaged by bright field; acquisition time was 200 ms. Imaging conditions and intensity scales were matched for each data set presented together.

Plasmid name	plasmid vector	promoter	Expression in	Features
P1	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-7]
P2	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-29]
P3	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-50]
P4	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-52]
P5	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-62]
P6	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-68]
P7	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-89]

P8	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-108]
P9	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-111]
P10	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-120]
P11	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-121]
P12	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-123]
P13	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-126]
P14	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-128]
P15	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-130]
P16	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-135]
P17	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-146]
P18	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-170]
P19	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-173]
P20	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-181]
P21	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-197]
P22	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-200]
P23	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-214]
P24	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-226]
P25	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[8-250]
P26	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[30-250]
P27	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[51-250]
P28	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[52-250]
P29	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[63-250]
P30	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[69-250]
P31	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[89-250]
P32	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[109-250]
P33	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[112-250]
P34	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[121-250]
P35	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[122-250]
P36	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[124-250]

P37	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[127-250]
P38	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[129-250]
P39	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[131-250]
P40	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[136-250]
P41	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[147-250]
P42	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[171-250]
P43	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[174-250]
P44	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[182-250]
P45	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[198-250]
P46	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[201-250]
P47	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[215-250]
P48	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[227-250]
P49	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-APEX2
P50	pcDNA3.1	CMV	Mammalian	CFP-NLS
P51	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-8]
P52	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-9]
P53	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-10]
P54	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-48]
P55	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-49]
P56	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-51]
P57	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-53]
P58	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-198]
P59	pcDNA3.1	CMV	Mammalian	flag-FRB-14aa linker-sAPEX2[aa2-199]
P60	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[9-250]
P61	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[10-250]
P62	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[11-250]
P63	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[49-250]
P64	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[50-250]
P65	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[52-250]

P66	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[54-250]
P67	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[199-250]
P68	pcDNA3.1	CMV	Mammalian	V5-FKBP-13aa linker-sAPEX2[200-250]
p69	pcDNA3.1	CMV	mammalian	C(1-29)-flag-FRB-sAPEX N-terminal fragment
p70	pcDNA3.1	CMV	mammalian	C-terminal fragment-V5-FKBP-MAVSTM

References

1. Sharp, K. H., Mewies, M., Moody, P. C. E. & Raven, E. L. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat. Struct. Biol.* **10**, 303–307 (2003).
2. Mohanty, J. G., Jaffe, J. S., Schulman, E. S. & Raible, D. G. A highly sensitive fluorescent micro-assay of H₂O₂ release from activated human leukocytes using a dihydroxyphenoxazine derivative. *J. Immunol. Methods* **202**, 133–141 (1997).
3. Zhou, M., Diwu, Z., Panchuk-Voloshina, N. & Haugland, R. P. A stable nonfluorescent derivative of resorufin for the fluorometric determination of trace hydrogen peroxide: Applications in detecting the activity of phagocyte NADPH oxidase and other oxidases. *Anal. Biochem.* **253**, 162–168 (1997).
4. Ahn, K., Szczesna-Skorupa, E. & Kemper, B. The amino-terminal 29 amino acids of cytochrome P450 2C1 are sufficient for retention in the endoplasmic reticulum. *J. Biol. Chem.* (1993).
5. Seth, R. B., Sun, L., Ea, C. K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF3. *Cell* (2005). doi:10.1016/j.cell.2005.08.012
6. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
7. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).
8. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).

Chapter 3. Directed evolution of split APEX

Introduction

Low activity of split APEX could be caused by a variety of factors, including poor stability, misfolding, aggregation of the fragments, or distorted geometry of the reconstituted active site and heme binding pocket, which could lead to low catalytic efficiency and/or poor heme recruitment. It is difficult to assess each of these potential problems, and even more difficult to fix them in a rational manner. Hence, we turned to yeast surface-display directed evolution, which we and others have exploited in many cases to improve or alter the properties of enzymes^{25,61,78-81}. We selected yeast display-based directed evolution in particular because compared to other strategies, such as high-throughput screening and phage display, yeast display allows processing of large mutant libraries ($>10^7$) and has large dynamic range; enrichment based on Fluorescence-Activated Cell Sorting (FACS) enables separation of highly active catalysts from moderately active ones, not just from inactive catalysts.

With yeast display, the goal is to find a mutant that has low affinity for each other and high activity upon close-proximity reconstitution. The eukaryotic nature of yeast makes it a suitable candidate for engineering split APEX2; yeast have protein folding and secretory pathways that are highly homologous to mammalian cells. Yeast surface protein A-agglutinin is a two-subunit glycoprotein composed of Aga1p and Aga2p. Aga2p forms two disulfide bonds to Aga1p. While there are no crystal structures of the Aga1p-Aga2p complex, it is known to serve as a mating protein with the binding domain on the C-terminus of Aga2p⁸¹. This feature makes the C-terminus of Aga2p an ideal location for protein fusion because of its steric accessibility to extracellular macromolecules⁸¹. Error-prone PCR is used to generate a library that can be introduced to yeast through electroporation. Utilizing this system, proteins fused to Aga2p have undergone directed evolution for enzymatic activity towards a peptide substrate fused to Aga1p^{81,82}. Aga1p fusion constructs are incorporated through the homologous recombination of the linearized plasmid⁸¹. Using a similar strategy, split APEX2 fragments can be fused to Aga1p and Aga2p to undergo a similar selection. Fluorescence-activated cell sorting (FACS) can be utilized to select for yeast cells that exhibit improved activity utilizing fluorophore-conjugated streptavidin to quantify extent of biotinylation. Similarly, FACS can also be utilized to select for high expression yeast clones by quantifying the amount of protein expressed on each yeast cell utilizing fluorophore-conjugated antibodies to label epitope tags C-terminal to the fused split APEX2 fragment.

Establishing a platform for directed evolution of split APEX

The first step towards utilizing yeast surface display directed evolution is to establish an appropriate display platform. The original starting template of split APEX2 fused to yeast surface must have both detectable activity and room to improve within the platform's detectable range⁷⁹. The three most promising cut sites from the two rounds of cut site screening: 51, 89, and 200 were selected for yeast display.

Displaying circularly permuted split APEX2 on the yeast surface

Originally, we envisioned displaying two fragments on the yeast surface could be accomplished through two major strategies: circular permutation and fusion to Aga2p or a true split system in where one fragment is fused to Aga2p and another on Aga1p. The circularly permuted strategy was first tested because it requires only a single protein fusion to Aga2p (**Figure 3-1A**), which provides the advantage to generate a library in which both fragments can be mutagenized. Secondly, the circularly permuted constructs connect the two fragments with a 3 amino acid linker, causing the environment of this selection system to be that emulating a PPI/close proximity. One potential drawback is that the selection would not select for any mutations that enhance the stability of the two fragment pieces on their own.

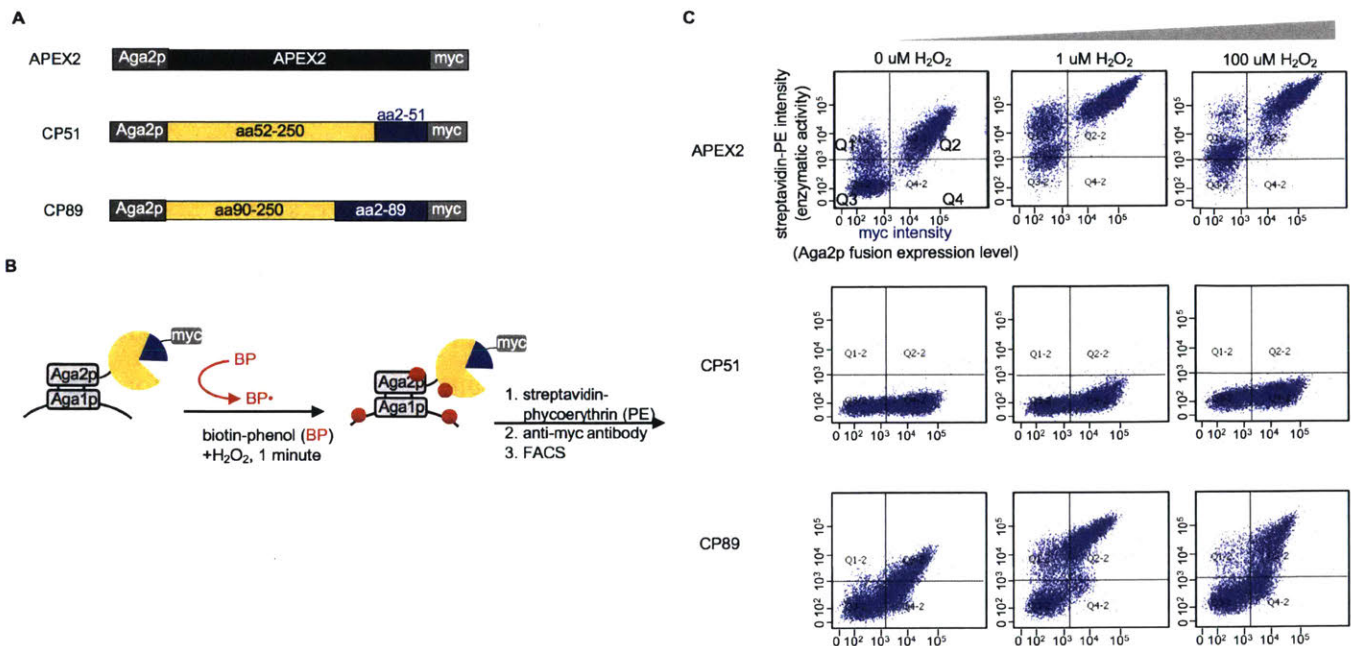


Figure 3-1. Displaying split APEX2 candidates as circular permutations on the yeast surface.

(A) Aga2p yeast surface-display fusion constructs of split APEX2. Cut sites 51 and 89 were circularly permuted – the N-terminal fragment was fused to the C-terminus of the C-terminal fragment with a flexible 3 amino acid linker. (B) Schematic overview of circularly permuted split APEX (CP51 is illustrated) displayed on the yeast surface. Peroxidase activity was detected by treating the cell mixture with biotin-phenol (BP, red dots) and H₂O₂. Cells with high peroxidase activity label themselves with biotin to a high extent²⁵detectable by FACS after streptavidin-phycoerythrin (PE) staining. (C) FACS analysis of constructs in (A) displayed on the yeast surface. For the 1 minute 100 μM biotin-phenol labeling, H₂O₂ was added at various concentrations for 1 minute.

Circular permutation of Aga2p fusion constructs (**Figure 3-1A**) were cloned via Gibson Assembly and transformed into yeast. Yeast were labeled with 100uM biotin phenol and varying concentrations of H₂O₂ prior to staining for FACS analysis (**Figure 3-1B**). Yeast cells that display active reconstituted peroxidase under these conditions should promiscuously biotinylate proteins on their cell surface, which can be quantified using FACS. The lack of detectable streptavidin-PE, which reflects enzymatic activity, from CP51 made it an unsuitable candidate for yeast display. CP89 and APEX2 behaved very similarly - the expression level of CP89 is slightly lower than that of APEX2; but at the overlapping expression level ranges, activity of both constructs was very similar (**Figure 3-1C**). With such high activity, CP89 was not a suitable candidate for a selection scheme either. It would be difficult to differentiate and select for variants that improved activity. These results also discouraged the testing of circular permutation at cut site 200, which consistently had significantly higher peroxidase activity than cut site 89 in mammalian testing. Under these conditions, both circularly permuted CP51 and CP89 did not appear to be feasible as a platform that provided the sufficient dynamic range required for directed evolution.

Displaying split APEX2 fragments on the yeast surface as separate fusion proteins

The second potential strategy is to display the split APEX2 fragments as fusions to different surface proteins, such as the yeast mating proteins Aga1p and Aga2p, which are displayed on the yeast cell surface and are joined by two disulfide bridges. The larger split APEX2 fragment was displayed on the yeast surface as a C-terminal fusion to the mating protein Aga2p, while the complementary fragment is co-displayed as a N-terminal fusion to Aga1p (**Figure 3-2A** see split display 89 and 200 as an example). To test this selection platform, we first cloned yeast Aga1p/Aga2p fusion constructs using the three best sAPEX pairs identified in the Chapter 2

screen. The fragments are allowed to reconstitute for ~20 hours during the induction of Aga2p-fusion protein expression. To read out the resulting peroxidase activity, we initiated APEX2-mediated biotinylation with biotin-phenol and H₂O₂, using standard conditions²⁵.

Upon preliminary testing, cut sites 51 and 89 fragment pairs expressed poorly in yeast and gave no detectable activity on the yeast surface; however, cut site 200 had demonstrated clearly detectable peroxidase reconstitution. Despite the lack of activity from cut site 89, the biotinylation detected by CP89 deemed the cut site still potentially promising in this display platform – further efforts were made with cut site 89 and 200 to test, optimize, and troubleshoot the split fragment yeast-surface display platform. One variation on the platform we tested was swapping the fragment fusion orientation to the Aga1p/Aga2p protein complex. This tested the feasibility of developing a split APEX system in which both fragments could be sequentially evolved. Additionally, from Jeff's experience in utilizing yeast surface display for the directed evolution of split HRP, we hypothesized that reconstitution happened in a *trans* fashion in that fragments reconstitute across different Aga1p/2p complex, and without known crystal structures of the Aga1P and Aga2p protein complex, it is unclear if that protein-protein interaction is geometrically favorable for cut site 89 reconstitution. Longer linkers (sequence obtained from Roger Tsien's lab⁸³) were also tested to address possible geometric constraint limitations. **Figure 3-2B** shows a summary of the alternative split display schemes that were tested.

Constructs were cloned and yeast were transformed with the Aga1p, Aga2p, and both Aga1p/2p constructs for each selection scheme (**Figure 3-2A; 3-2B**). Transformed yeast clones were then subjected to standard 100 μM biotin-phenol labeling conditions. Reconstitution should yield biotinylation of yeast surface proteins, which is detected using a fluorophore-conjugated streptavidin. Immunostaining of Aga1p-only fusions showed robust expression for all the constructs. Immunostaining was also performed on Aga2p-only fusions (data not shown). These experiments provide reassurance that no enzymatic activity is detected when its complementary partner is absent and that there is a basal level of expression without any steric constraints from the fusion Aga1p construct. The Aga2p fusion-containing yeast displayed a larger range of expression levels. Finally, biotin-phenol labeling was performed on yeast transformed to express both fragments of split APEX2 upon induction (**Figure 3-2C**).

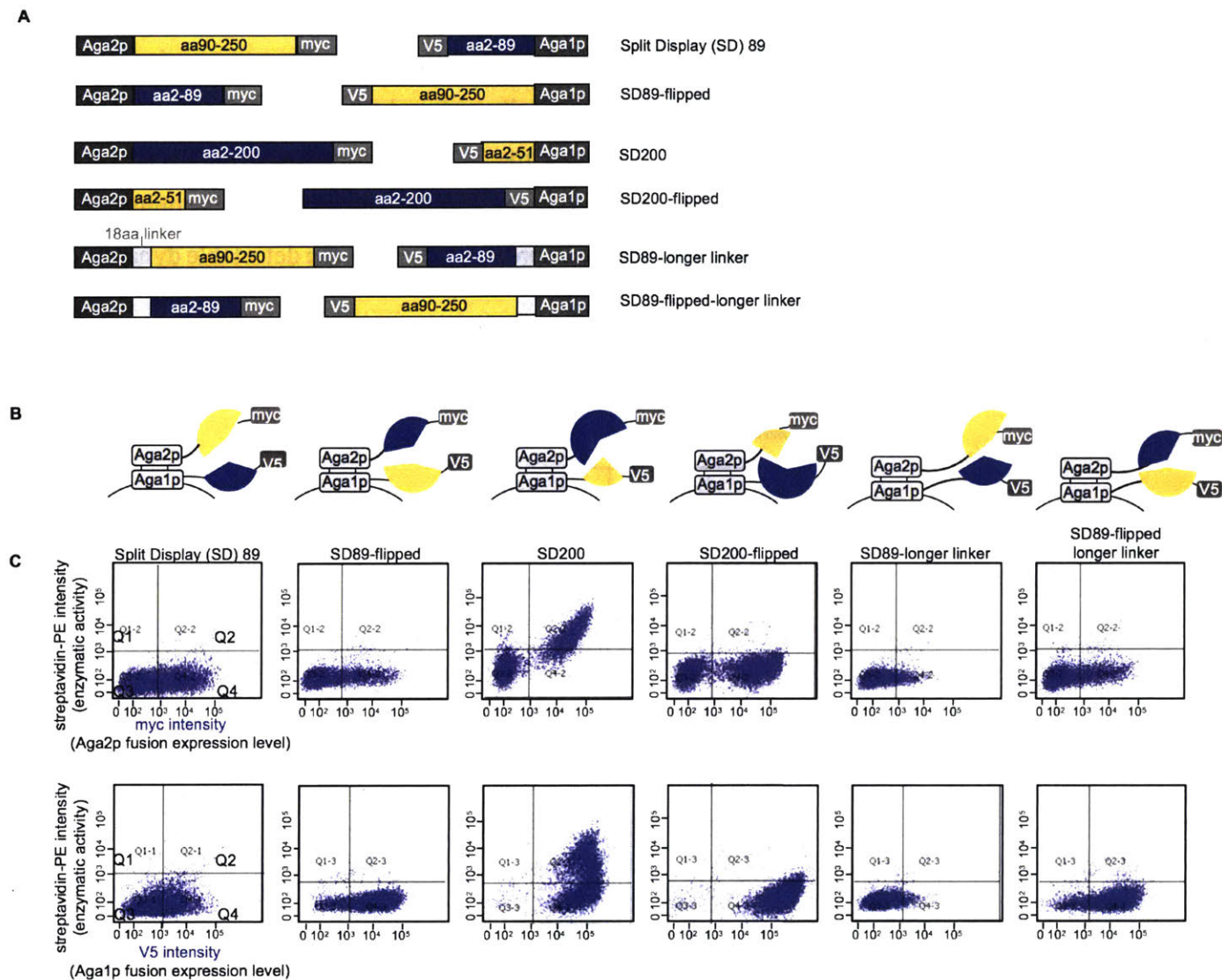


Figure 3-2. Displaying split APEX2 pairs as fusions to different yeast surface proteins. (A) Split Display (SD) indicates both Aga1p and Aga2p construct pairs used for each cut site. SD89-flipped means that the fragments of cut site 89 have the larger fragment as a fusion to Aga1p – thus “flipped”. The longer linker is inserted between the native yeast surface protein and the protein fusion, in hopes to further project the fragment away and alleviate any geometric constraint issues. Each construct also contains an epitope tag which can be utilized to determine protein expression levels **(B)** Cartoons that depict the different platforms tested in **(A)**. **(C)** FACS analysis of yeast in **(B)** that underwent 5 minutes of 100 μ M biotin-phenol labeling. The y-axis of both rows of plots display the extent of reconstituted enzymatic activity. The top row shows the expression level of the protein fragment fused to Aga2p. The bottom row shows the extent of expression of the Aga1p fusion protein.

Strikingly, biotin-phenol labeling was only detected with split fragments with cut site 200. We hypothesize that the detected activity is a result of sufficient expression of both split fragments on the yeast surface, and that the yeast surface provides a sufficient local concentration for reconstitution. Curiously, swapping the fragments drastically reduced the extent of biotinylation despite robust expression of both fragments. However, the SD200 platform provided a robust signal, the opportunity to mutagenize the bulk of the protein, and the lower activity as compared to that of full length APEX2 signifies there is room for improvement within the detectable dynamic range.

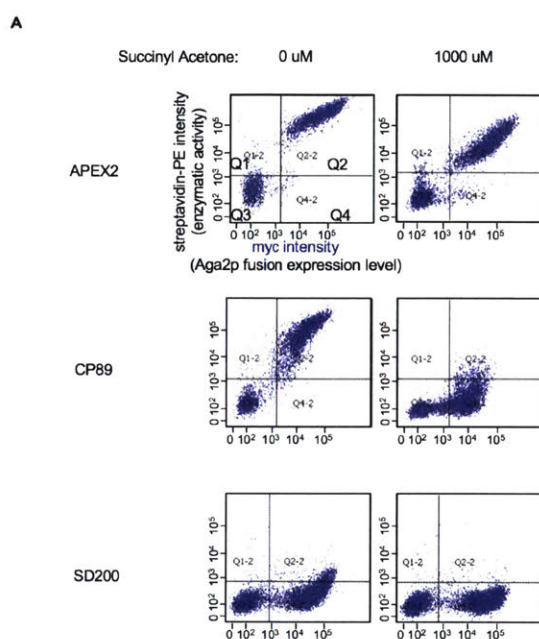


Figure 3-3. Comparing split display (SD200) and circularly permuted (CP89) platforms to full length APEX2 fused to Aga2p. (A) FACS analysis of full length APEX2 was compared to circularly permuted cut site 89 (CP89) and split display cut site 200 (SD200). Yeast were induced for 20 hours in the absence or presence of 1 μ M succinyl acetone in galactose containing induction media. Yeast were then washed and standard 100 μ M biotin-phenol labeling and prepared for analysis by FACS.

As the split display and circularly permuted platforms were being explored in tandem, it was interesting to compare the optimize version of both to each other, as well as against their ideal benchmark level for reconstituted activity – the full length APEX2. APEX2 as a fusion to Aga2p,

CP89, and SD200 were compared using standard biotin-phenol labeling conditions, for 1 minute (**Figure 3-3**). The yeast were either induced using regular galactose containing media, or succinyl acetone was added to an effective concentration of 1 μ M in the inducing media (90%SGCAA to 10%SDCAA) for the entirety of the induction period. (see heme modulation section for more information. While the presence of succinyl acetone did decrease the activity of full length APEX2, extent of biotinylation and expression levels were robust in both conditions. The limited bioavailability of heme significantly impacted CP89 – activity is nearly abolished, and expression level. Despite the yeast being monoclonal, the population is not very tight by FACS analysis. This large variability of activity to expression indicates also a more complex and more “noisy” system when trying to select specific clones from a pool of different yeast.

SD200 has detectable biotinylation and the platform works to provide a large dynamic range for improvement of enzymatic activity. While the addition of succinyl acetone does abolish detectable enzymatic activity, the expression level of the large fragment does not suffer and the population of expressing yeast cells is very tight.

Yeast display evolution of AP (large fragment)

Generation 1

We proceeded with directed evolution of the 200/201 fragment pair using the split display strategy. The C-terminal fragment (amino acids 201-250 of APEX2; henceforth called “EX”) was held constant while the N-terminal fragment (amino acids 1-200 of APEX2, called “AP-0”) (**Figure 3-4A**) was mutagenized using error prone PCR. Sequencing showed that our AP-0 library contained an average of 1.4 amino acid mutations per clone (**Experimental methods**). The ideal library, as according to Wittrup lab protocols⁸¹, is one in which roughly 50% of activity is eliminated³¹. Yeast cells with activity that correlates with expression, are within the upper right quadrant (Q2-2). 3 libraries were combined in a 1:2:1 ratio to further increase diversity and to achieve the 50% reduction in activity (**Figure 3-4B**). The library size listed in **Figure 3-4A** is that of the most diverse library as determined by an estimation of serial dilutions (**Experimental methods**). However, actual effective library size estimates could have been influenced by more complicated factors such as the effective coverage of the library during the actual sort, how many

of them were expressing, number of truly viable clones, duplicates, silent mutations, codons that result in the same amino acid, etc. As a result, the first-round sort utilized a more permissive selection gate that retained all yeast cells that had detectable biotinylation. In subsequent selections, the gate became more stringent and the percentage of yeast retained dropped significantly. We performed a total of four rounds of selection and observed that the activity of the

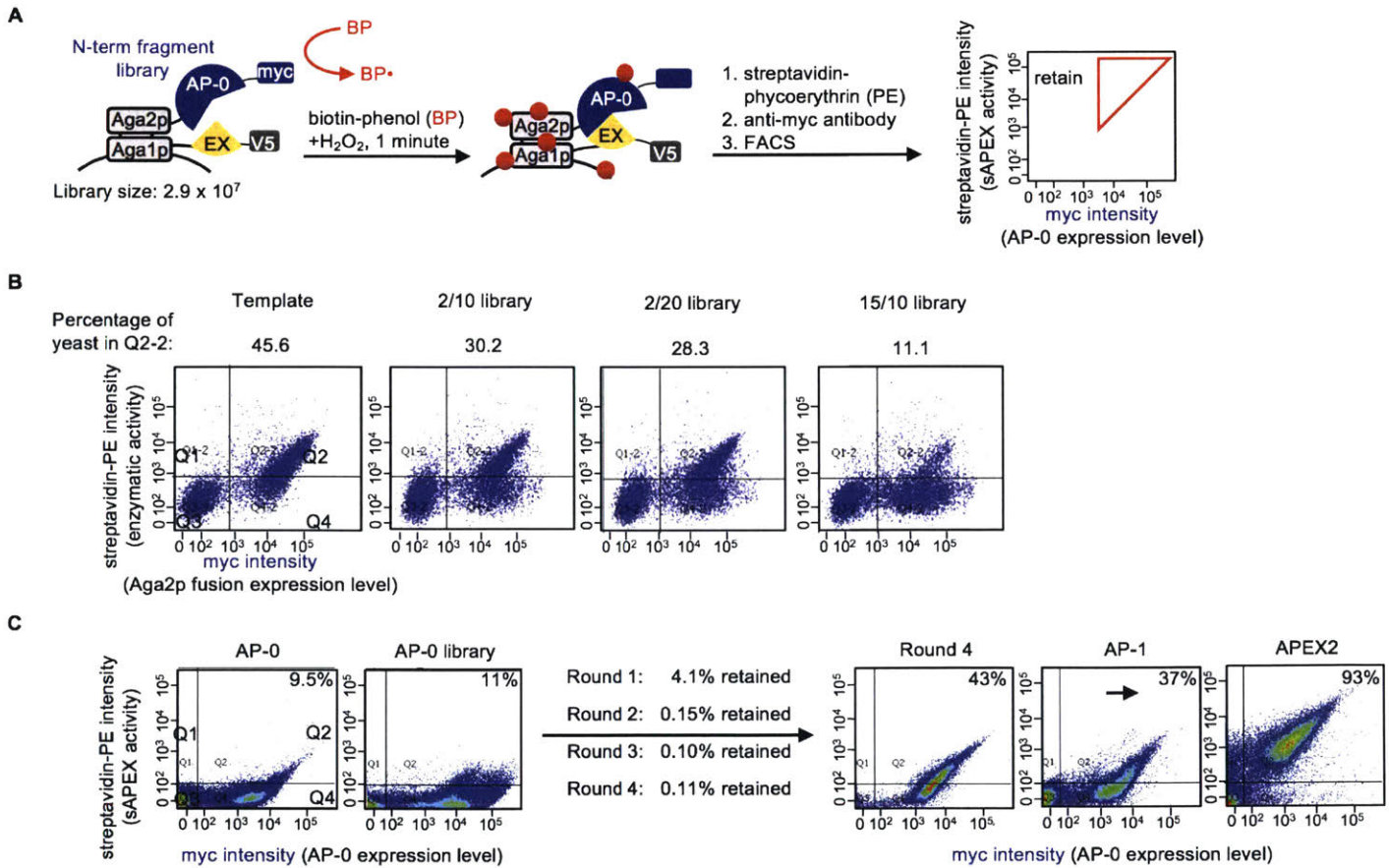


Figure 3-4. Generation 1 yeast display-based directed evolution. (A) Experimental setup. A library of N-terminal fragment AP-0 variants was displayed on the yeast surface via fusion to the mating protein Aga2p. A constant C-terminal fragment, EX, (aa 201-250 of APEX2) is co-displayed via fusion to Aga1p. The yeast cell library is treated with biotin-phenol (BP) and H_2O_2 for 1 minute. Cells with active peroxidase self-biotinylate (red dots). Biotinylation sites are detected by staining with streptavidin-phycoerythrin (PE), and anti-myc antibody (followed by anti-Chicken-AlexaFluor647) is used to quantify AP-0 expression level. Using FACS, we enrich cells with a high PE/myc staining ratio. (B) FACS plots comparing the extent of biotinylation generated from three different libraries against the template, which is the monoclonal population of non-mutagenized original starting sequence. The libraries were constructed from error prone PCR. 2/10 library indicates utilizing 2 μ M of non-natural dNTPS for 10 PCR cycles. (C) FACS

plots showing progress of selection. Cells were prepared and treated as in (A). AP-0 is the original N-terminal sAPEX fragment template. Round 4 is the population of yeast cells that remain after 4 rounds of FACS selection and re-amplification. AP-1 is the best clone from the Round 4 pool, whose mutations are shown in Figure 2C. “APEX2” is the original full-length APEX2. Percentage on the upper right corner of each FACS plot indicates the number cells in quadrant 2 divided by the total number of cells that are expressing N-terminal fragment (quadrant 2 + quadrant 4). The experiment has three biological replicates in total.

yeast pool (measured by FACS) progressively increased (**Figure 3-4C**). Zymoprep analysis was performed on Rounds 2, 3, and 4 to assess the level of convergence within the remaining pool. We isolated 12 individual clones, characterized their activity on yeast using FACS as well as in the mammalian cytosol using microscopy. Mutations that appeared to be beneficial were combined and also tested in parallel (more information in chapter 4). The final clone from the first generation of evolution was “AP-1”, which contains 3 mutations relative to the original APEX2. In a side by side comparison to the original split APEX fragment pair (AP-0 + EX) in HEK 293T cells, AP-1 shows noticeably improved activity in both DAB and BP labeling assays (see chapter 4 for characterization).

Generation 2

To further improve the reconstituted activity of sAPEX, we continued to perform another generation of directed evolution. The yeast display selection platform was modified and improved to provide more control over the selection environment. Now, the C-terminal EX fragment was not co-displayed on Aga1p but instead added as a purified, soluble protein (**Figure 3-5A**). This configuration allows us to precisely control the concentration of EX added and the time of incubation, and to select for mutations that improve AP-1 expression and stability in the absence of EX translation. AP-1, serving as the template for generation 2 evolution, was expressed on the yeast surface as a fusion to Aga2p. EX fused to the N-terminus of GFP was cloned into the PYFJ16 vector, which is utilized for transformation into BL21 E. coli for protein production.

However, because the fragments only encounter each other on the yeast surface, where endogenous heme is not available, it was necessary to supply exogenous heme (see *heme section*) to the cells both during and prior to EX fragment addition. Through the four rounds of selections,

the concentration and time allotted for reconstitution became increasingly more stringent (**Figure 3-5B**). Heme concentration and incubation times also varied with a general trend towards increased stringency. Using this scheme produced the clone “AP-2” which has three mutations and about a two-fold improvement (in the conditions of **Figure 3-5C**) over AP-1.

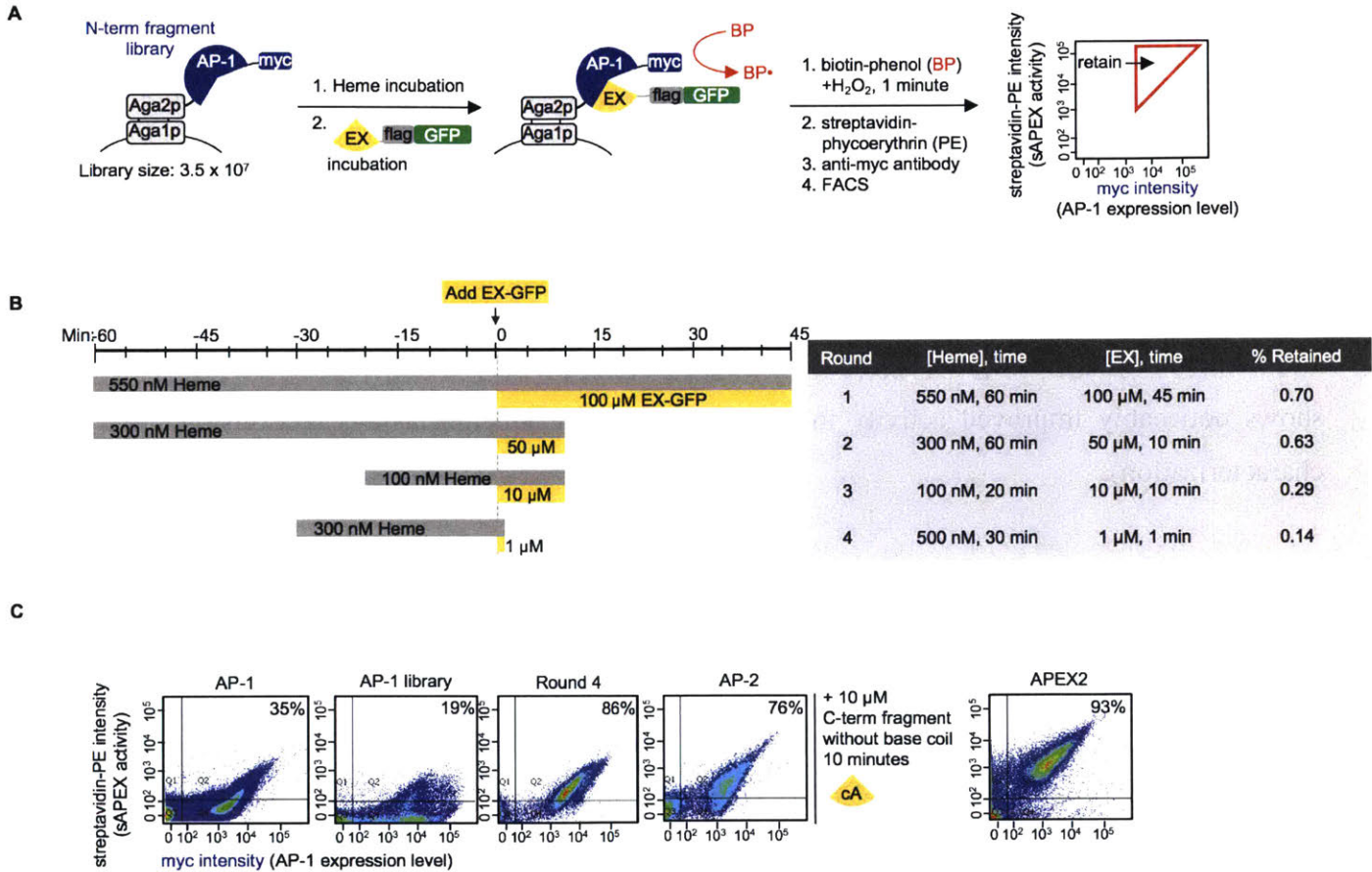


Figure 3-5. Generation 2 yeast display-based directed evolution. (A) Experimental setup. A library of N-terminal fragment AP-1 variants was displayed on the yeast surface via fusion to the mating protein Aga2p. The C-terminal fragment is added as a purified protein. BP labeling, streptavidin-phycoerythrin staining, and FACS are performed as in **Figures 4-4B**. (B) Specific labeling conditions used in each round of selection. (C) FACS analysis of indicated samples. Round 4 is the library population remaining after four rounds of selection. AP-2 is the best clone from 4 rounds of selection. Cells were incubated with 500 nM of heme for 30 minutes before addition of 10 uM EX-flag-GFP for 10 minutes followed by BP labeling, streptavidin staining, and FACS.

Generation 3

To drive the fragments together using a protein-protein interaction (PPI), in a third-generation effort, we again added EX as a soluble protein, but we used an artificially designed coiled-coil system, ACID-p1 and BASE-p1⁸⁴ to recruit EX to the proximity of the N-terminal fragment (**Figure 3-6A**). This mimics the sAPEX reconstitution that would occur in our target biological applications. After four rounds of selection, with gradual reduction of EX concentration and incubation time (**Figure 3-6B; 3-6C**), we produced clone “AP-3”, which incorporates one additional mutation relative to “AP-2” and had noticeably higher activity than preceding clones in both DAB and BP labeling assays, in HEK 293T cells, but this activity was not dependent on PPI-induced co-proximity (see chapter 4).

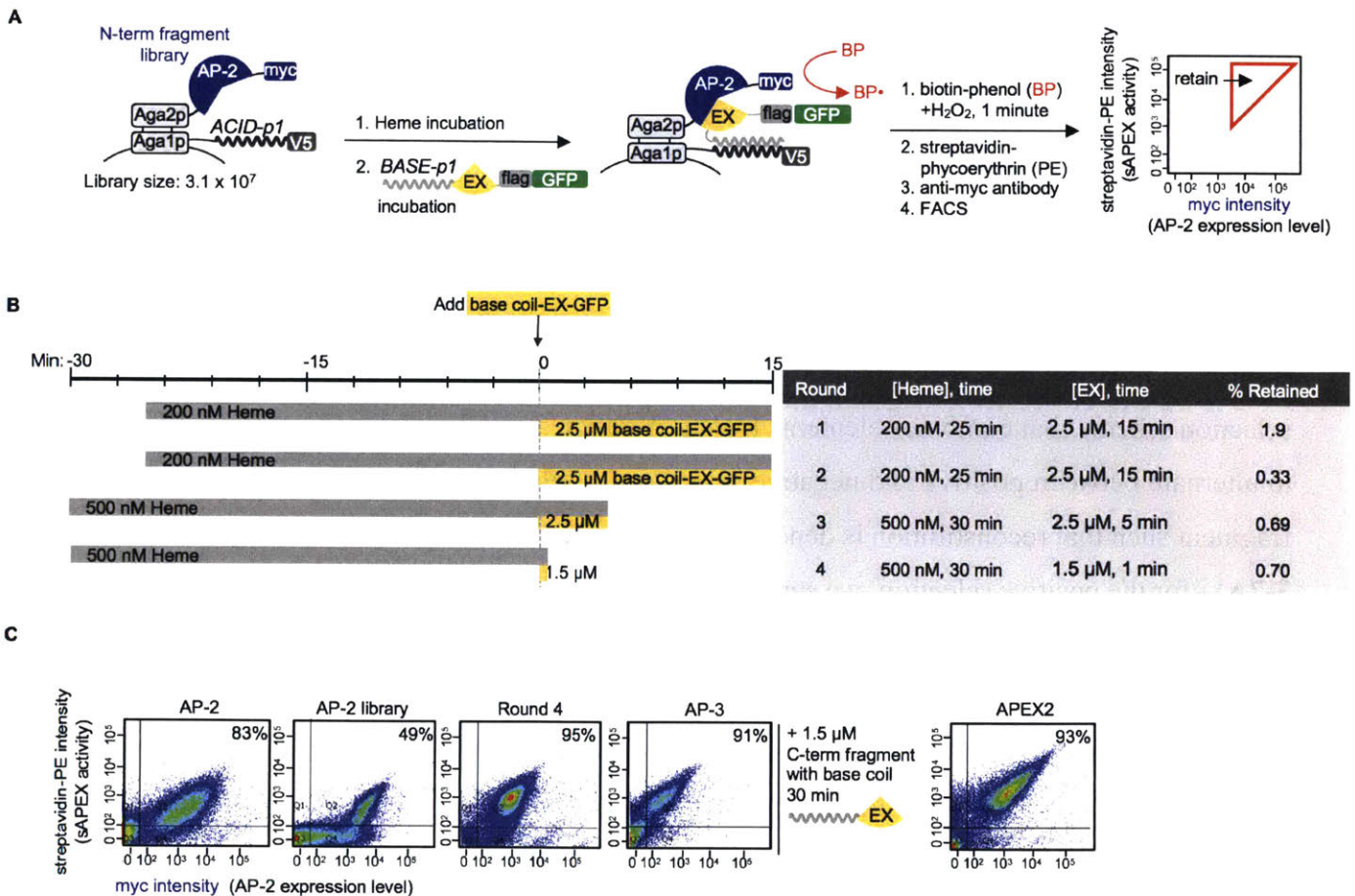


Figure 3-6. Generation 3 yeast display-based directed evolution. (A) Experimental setup. A library of N-terminal fragments (“AP-2”, finalized clone from generation 2 selections) was displayed on the yeast cell surface via fusion to the Aga2p mating protein. An acid coil was co-displayed, via fusion to Aga1p, to recruit base coil-fused C-terminal fragment (“EX”, amino acids 201-250 of APEX2). Base coil-EX-GFP was incubated with the AP-3 yeast library and reconstituted peroxidase activity was detected by treating the cell mixture with biotin-phenol (BP) and H₂O₂. Cells with high peroxidase activity label themselves with biotin to a high extent²⁵, enabling their enrichment via FACS after streptavidin-phycoerythrin (PE) staining. Myc staining on the x-axis provides a readout of AP expression level. (B) Specific labeling conditions used in each round of selection. (C) FACS analysis of indicated samples. Labeling condition was 500 nM of heme for 30 minutes, then 1.5 μM base coil-EX-GFP protein for an additional 30 minutes, before BP labeling, streptavidin staining, and FACS.

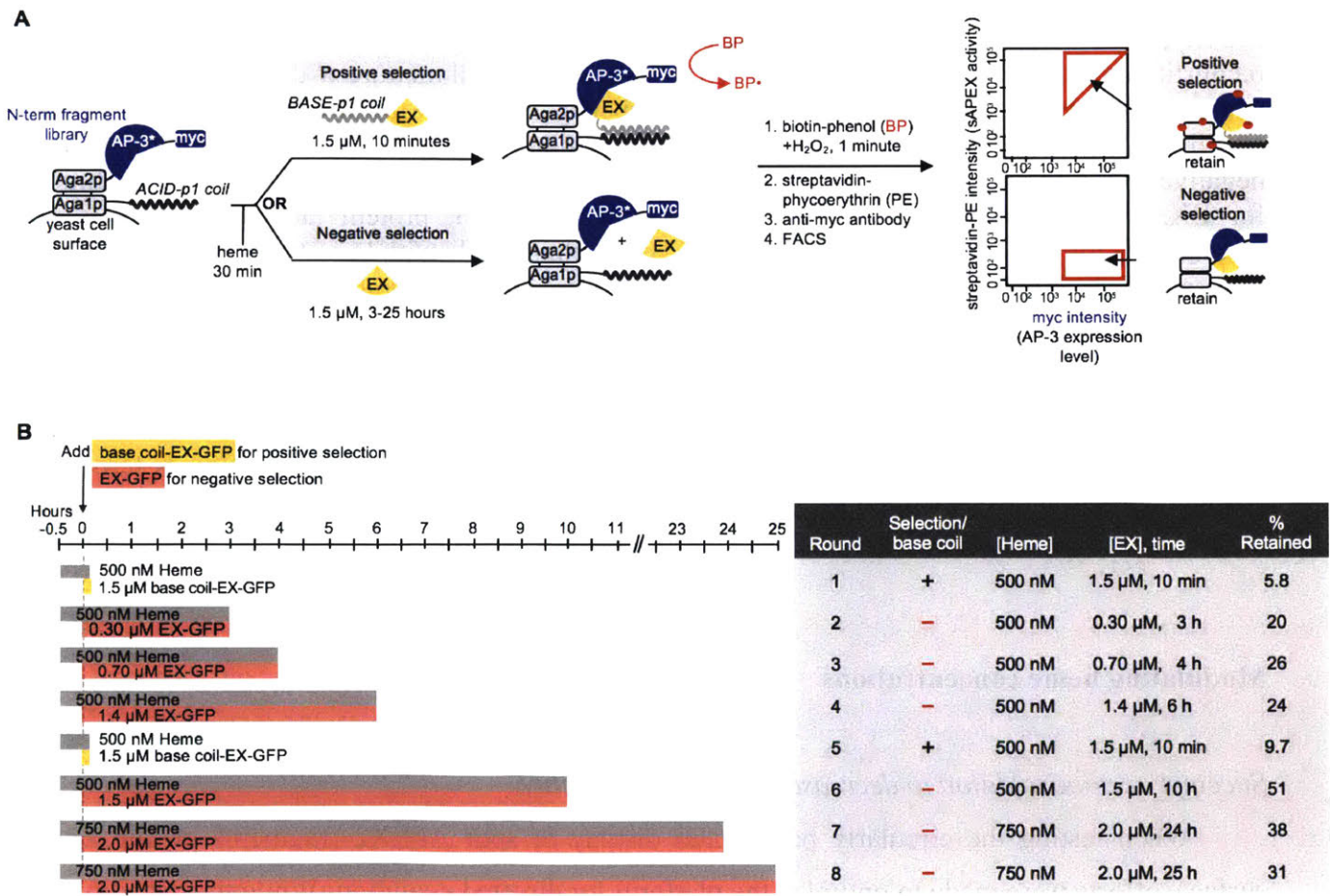
FACS-based negative selections for reduced fragment affinity

Generation 4

An ideal sAPEX fragment pair would have high catalytic activity when reconstituted, but low intrinsic binding affinity between the fragments, such that the fragments would only reconstitute when driven together by a PPI. The clone AP-3, obtained after three generations of directed evolution, has much greater reconstituted activity (with EX) than does the original template (AP-0); however, this activity was not dependent on PPI-induced co-proximity: using FRB and FKBP fusion proteins in HEK 293T cells, for example, we observed a considerable amount of DAB and BP signal in the absence of rapamycin (see chapter 4).

To preserve high reconstitution activity but decrease fragment affinity, we devised a new selection scheme that combines elements from generation 2 and 3 selection platforms. We wished to alternate between positive and negative selections to select for a low affinity variant of the AP fragment such that reconstitution is dependent on close co-proximity of both fragments (**Figure 3-7A**) - for the positive selection, we supplied the purified EX protein, fused to the BASE-p1 helix that facilitates recruitment to the Aga1p-ACID-p1. We performed BP labeling followed by streptavidin-phycoerythrin and anti-myc antibody staining. FACS was used to enrich cells with high SA/myc intensity ratios, as above. For the negative selection, we incubated the yeast with EX protein *lacking* BASE-p1 coil for extended periods of time (3 to 25 hours). AP-3 mutants that reconstituted with EX under these conditions were eliminated by FACS.

Starting with a library of 4.8×10^8 AP-3 variants, a permissive positive selection was first implemented to retain all mutants within the library that still reconstituted enzymatic activity. Subsequently, three rounds of negative selections were performed with greater demands for lower affinity properties by increasing the concentration and time of EX fragment incubation. In total, we performed a total of two rounds of positive selection interspersed by six rounds of negative selection (see **Figure 3-7B** for details). By round 8, the yeast population was significantly depleted in cells exhibiting fluorescence upon addition of EX lacking the BASE-p1 coil (**Figure 3-7C**). We isolated 4 unique clones, characterized them by FACS, combined beneficial mutations, and re-tested in the mammalian cytosol. These experiments resulted in AP, the final sAPEX N-terminal fragment, which contains 9 mutations compared to the original APEX2 sequence and two more mutations compared to the starting template, AP-3.



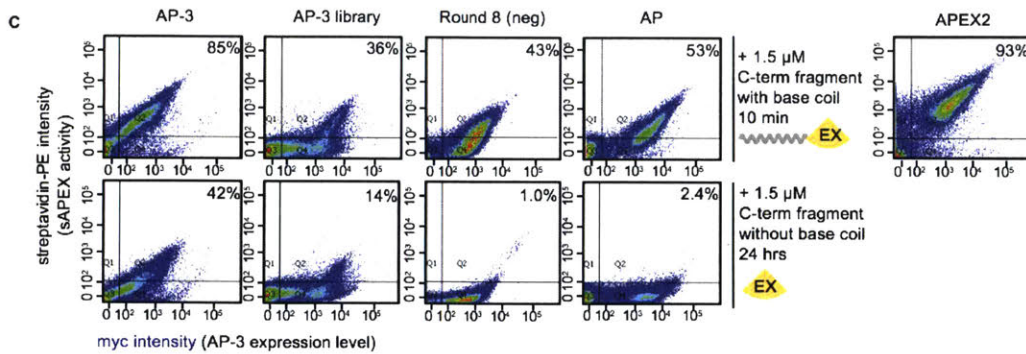


Figure 3-7. Generation 4 yeast display-based directed evolution. (A) Yeast display-based directed evolution scheme. A library of N-terminal fragments (“AP-3”, finalized clone from generation 3 selections) was displayed on the yeast cell surface via fusion to the Aga2p mating protein. An acid coil was co-displayed, via fusion to Aga1p, to recruit base coil-fused C-terminal fragment (“EX”, amino acids 201-250 of APEX2). In the positive selection for high sAPEX activity, base coil-EX-GFP was incubated with the AP-3 yeast library for 10 minutes, then reconstituted peroxidase activity was detected by treating the cell mixture with biotin-phenol (BP) and H₂O₂. Cells with high peroxidase activity label themselves with biotin to a high extent²⁵, enabling their enrichment via FACS after streptavidin-phycoerythrin (PE) staining. In the negative selection to deplete the sAPEX library of AP-3 fragments with excessively high affinity for EX, we incubated the AP-3 yeast library with EX-GFP protein lacking base coil for increasingly long time periods then performed BP labeling. Cells with low streptavidin-PE signal were retained via FACS. Myc staining on the x-axis provides a readout of AP expression level. (B) Specific labeling conditions used in each round of selection. (C) FACS analysis of indicated samples. Labeling condition was 500 nM of heme for 30 minutes, then 1.5 μ M EX-GFP, with or without base coil, for 10 min or 24 hours, respectively. Next, BP labeling, streptavidin staining, and FACS were performed. Percentage on the upper right corner of each FACS plot indicates the number cells in quadrant 2 divided by the total number of cells that are expressing N-terminal fragment (quadrant 2 + quadrant 4).

Modulating heme concentrations

Succinyl acetone addition to decrease heme bioavailability

When testing the circularly permuted display of split APEX2 fragments on the yeast surface, efforts were made to optimize the platform for directed evolution. While the detection of biotinylation with CP89 was overly robust, i.e. lacking sufficient range to select for an improved variant, the signal in it of itself was still a promising sign. As a method to increase the stringency

of the labeling experiment, 1 mM succinyl acetone was included in the protein induction step. The succinyl acetone is an inhibitor to the heme biosynthesis pathway. (ref) The reproducible results indicated that upon the addition of succinyl acetone, expression level and activity were drastically different between APEX2 and CP89. APEX2 maintained similar activity as without succinyl acetone. CP89 demonstrated much lower expression and very, very low activity (**Figure 3-8A**). This might indicate a lack of stability of the CP89 construct in the absence of heme and the addition of succinyl acetone. However, further testing of succinyl acetone addition resulted in inconsistencies of the extent of detectable biotinylation of the same yeast clones from experiment to experiment. This, in conjunction with the significant decrease in expression, shifted efforts towards optimizing the split display platform.

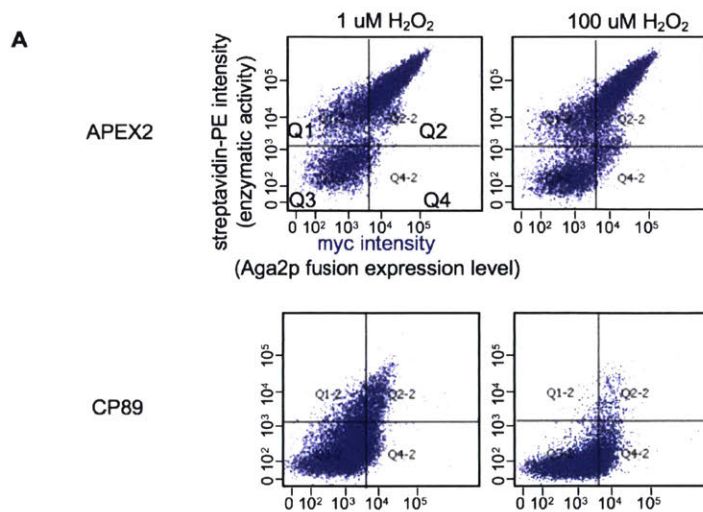


Figure 3-8. Using heme biosynthesis inhibitor, succinyl acetone, to reduce CP89 peroxidase activity. (A) Yeast cultures were induced with galactose containing media to express Aga2p fused to circularly permuted cut site 89 (CP89) and full length APEX2. Induction media also contained 1 mM succinyl acetone to inhibit heme biosynthesis. Biotin-phenol labeling followed by FACS analysis comparing APEX2 and CP89 reveals a significant difference between their enzymatic activity. See Figure 4-3 for comparison with and without succinyl acetone addition (along side with SD200).

Exogenous heme supplementation to yeast

Advancing the generation 1 split display platform to that of generation 2 was slightly challenging. Initially, FACS analysis of samples were rarely consistent; biotinylation from reconstituted activity was not even consistently detectable. This would occur even when utilizing the freshly purified EX-GFP proteins that had undergone no freeze thaw cycles and determined to be unpreolyzed by protein gel. Effective loading of the small fragment onto the yeast surface was confirmed by detection of both fluorescence of the GFP fusion and immunostaining for the flag epitope tag linker between EX and GFP. Furthermore, myc immunostaining indicates that the larger fragment AP-1 is also intact (**Figure 3-9B**). AP-1 co-expressed with EX on the yeast surface (Generation 1 platform) was used as a positive control and it showed robust, reproducible biotinylation-signal. This indicates that the inconsistencies were not due to degradation of the two split fragments and occurs prior to the labeling step. Additionally, the robust expression of both fragments suggest that geometry is likely not a constraint with the flexibility of 10 amino acid linkers.

As a result, we hypothesized there was a heme occupancy problem. With low heme bioavailability in yeast cells, sole expression of AP-1 might lead to poor heme-binding. As APEX is a heme-dependent peroxidase, binding of EX to AP-1 without heme would not result in any enzymatic activity. We hypothesized that the concentration of heme within yeast fluctuates with yeast health, age, growth, and washing conditions; these fluctuations heavily influence the amount of heme pre-bound to AP-1 on the yeast surface and resulted in inconsistencies between experiments. To address this issue, heme was supplemented exogenously prior to EX-GFP incubation (**Figure 3-9**).

From this heme supplementation assay, it reveals that heme supplementation does boost/restore peroxidase activity; thus, heme bioavailability is a critical aspect to consider (**Figure 3-9a**). Additionally, pre-binding of the heme is not a necessary step to reconstitute activity within the hour-long time scale (data not shown). The supplementation of heme boosts activity levels to that similar to generation 1 display platform, in which both fragments, AP-1 and EX, are co-expressed on the yeast surface. This positive control system represents a fully saturated system in which the EX fragment is expressed as a fusion of Aga1p and is constitutively on.

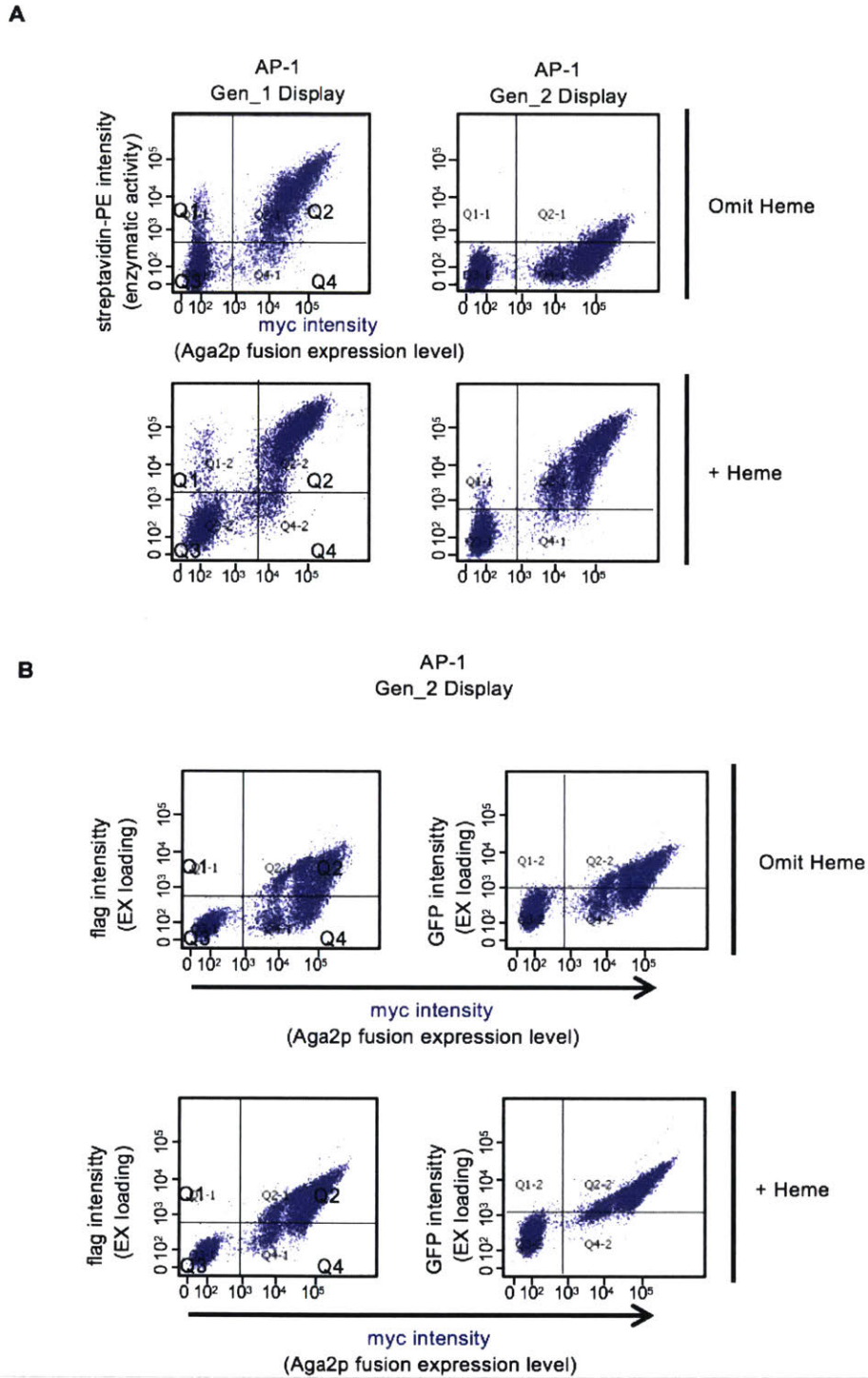


Figure 3-9. Testing exogenous heme supplementation prior to EX-GFP incubation. After 20-24 hrs of induction, heme was supplemented or omitted to yeast cultures for a final concentration of 2 μ M. After a 60 minute heme incubation period, purified EX-GFP protein was added to generation 2 display platform yeast cultures for a final concentration of 100 μ M. (A) Effect of

heme supplementation on biotinylation levels and AP-1 expression levels. FACS analysis of yeast with AP-1 co-expressed with EX (generation 1 platform) and sole AP-1 expression (generation 2 platform). **(B)** Effect of heme supplementation on EX-GFP recruitment to the yeast surface as measured by GFP intensity and flag intensity. EX-GFP contains a flag linker.

The supplementation of heme also slightly boosts the loading of EX-GFP fragment as determined by both GFP intensity as well as immunostaining of its flag epitope tag (**Figure 3-9b**). This indicates that with reduced heme levels, the AP-1 fragment is less efficient at the EX-GFP protein, indicating that the original heme binding was cooperative, in which the presence of the EX fragment aids the folding/stabilization of the AP-1 fragment, enhancing its heme recruitment/binding affinity. Most importantly, heme supplementation significantly improved the consistency between FACS analysis and provided a more constant enzymatic reconstitution environment, which was critical for both analysis and sorting of yeast libraries. Heme supplementation was implemented for generation 2 through generation 4.

Discussion

Having completed 20 rounds of selection using three different yeast display configurations, we characterized key clones side by side to assess the progress of our evolution. First, we prepared yeast displaying AP-0 (the starting template), AP (N-terminal fragment of final sAPEX), full-length APEX2, and AP-1 to AP-3 mutants, as fusions to Aga2p. Aga1p on these cells contained the ACID-p1 coil. We then supplied EX protein, either fused to (**Figure 3-10A top row**), or lacking (bottom row), the BASE-p1 coil for proximity-dependent reconstitution. **Figure 3-10A** shows that the streptavidin-phycoerythrin (PE) staining, reflecting reconstituted peroxidase activity, progressively increases from the template AP-0 to the finalized AP. However, the signal in the bottom row, reflecting proximity-independent reconstitution with EX (lacking the BASE-p1 coil) also increases, which is undesirable. After the implementation of negative selections in generation 4, however, the untagged EX signal decreases for the AP clone. The EX+BASE-p1 coil signal remains high for AP, although not quite as high as that seen for AP-3. These observations on yeast indicate that the selections worked as desired and that our final clone AP combines the

features of high reconstitution activity with low proximity-independent reconstitution (i.e., low fragment affinity).

After 4 generations of evolution, 9 mutations were acquired on the large split APEX2 fragment as compared to the original template. Mapping the positions of these 9 mutations (G50R, K61R, I165L from Generation 1; R24G, H62Y, N72S from Generation 2; K22R from Generation 3; and P125L, I185V from Generation 4) (summary in **Figure 3-10B**) onto the structure of wild-type ascorbate peroxidase, we observe that all lie in solvent-exposed regions, and none are at the interface between AP and EX (**Figure 3-10C**). Interestingly, half of the mutations are adjacent to cut sites we screened in chapter 2, and many of the mutations were clustered in specific regions of the protein structure (**Figure 3-10C**).

We hypothesized that these mutations might serve to adjust the rigidity of AP tertiary structure. For instance, mutation G50R might reduce the flexibility of the second loop while also increasing its hydrophilicity. Intriguingly, one area distant from both the heme binding site and the AP-EX interface accumulated a mutation in all but one generation of evolution (**Figure 3-10C**). The three mutations – P125L (Generation 4; located along the fourth loop amongst a group of prolines at aa 123-127 PEPPP to PEPLP) and K61R, H62Y (Generation 1 and Generation 3 respectively; located at a turn between alpha helices) might work in conjunction to modulate the tertiary structure of AP (like a hinge) for optimal reconstitution.

Other mutations are also clustered. The first alpha helix contains mutations K22R and R24G accumulated from Generation 2 and Generation 3, and I185V is adjacent to I165L in 3D space. The curious nature of clustered mutations may be explained by either bias in library generation, other regions do not tolerate mutation, or mutations to these regions are the most allosterically influential in engineering a low affinity sAPEX. Overall, it is difficult to accurately predict the actual effect of any mutation though, as one of the biggest boost of activity is from K22R.

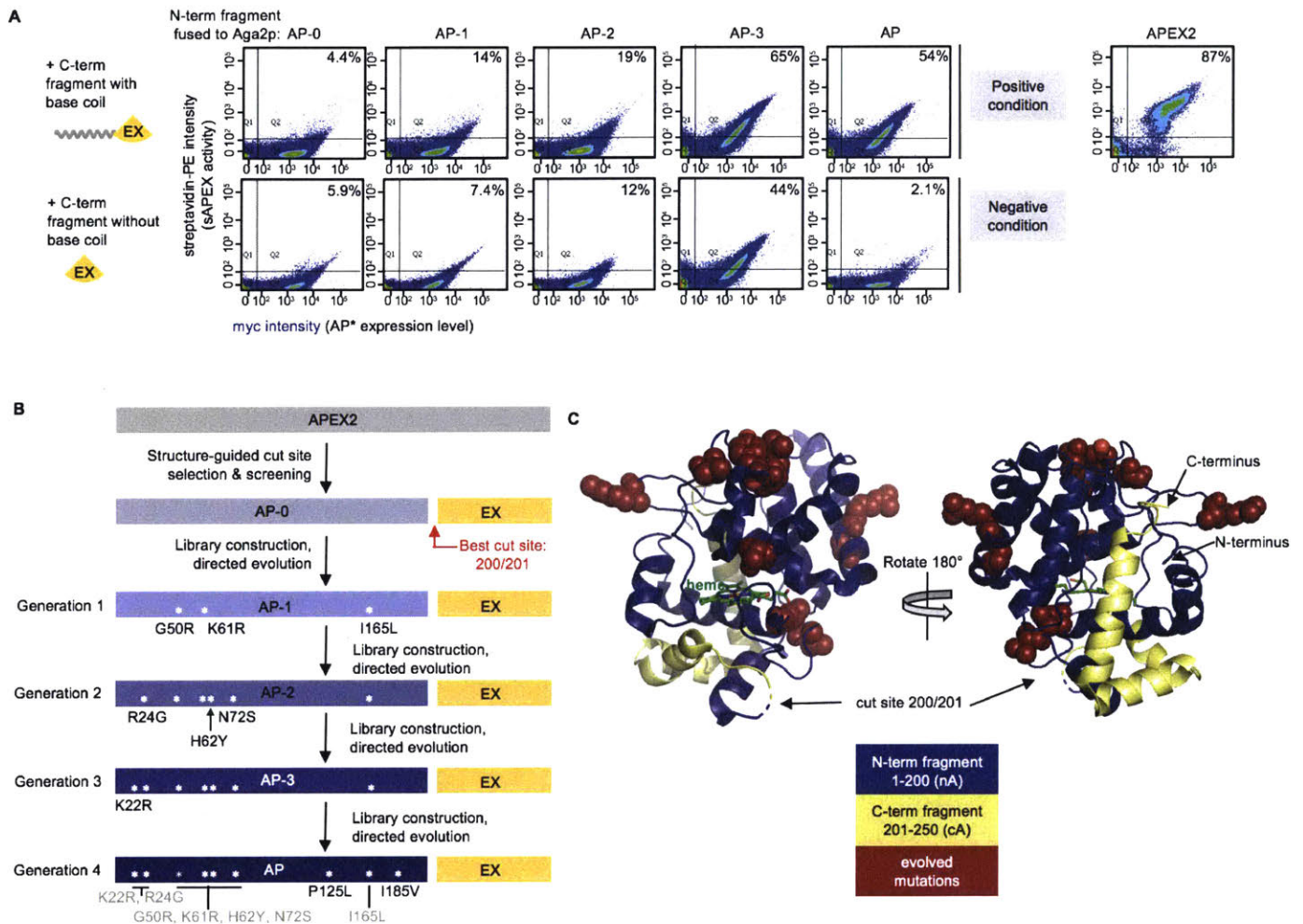


Figure 3-10 Yeast-display directed evolution and results. (A) Summary of improvement of sAPEX activity and PPI-dependence throughout generations of selections in yeast. Full length APEX2 fusion to Aga2p was used as a benchmark for desired activity range. Yeast were prepared as utilizing generation 4 platform, see **Figure 3-7A**, with the indicated N-terminal fragment of sAPEX expressed on the yeast surface as a fusion to Aga2p. Purified C-terminal fragment (EX), either with or without acid coil, was added to the cells for 10 minutes or 7 hours, respectively. Then BP labeling, streptavidin-PE staining, and FACS were performed as in **Figure 3-7A**. AP-1, AP-2, and AP-3 are the first, second, and third generation N-terminal fragment clones, whose mutations are shown in (B). The percentage of myc-positive cells in the top right quadrant (Q2) is indicated in the top right corner of each FACS plot. Showing data from one out of two biological replicates. (B) Summary of split APEX protein engineering. The name of the best N-terminal fragment clone to emerge from generation 1 selections is “AP-1”, and so on, as indicated. The best clone to emerge from the generation 4 selection is “AP.” Asterisks depict the locations of mutations within the protein sequence. (C) sAPEX split site and mutations in sAPEX relative to full-length APEX2¹¹. The split site is between E200 and G201, and the N-terminal fragment (“AP”) and C-

terminal fragment (“EX”) of sAPEX are colored blue and yellow, respectively. The nine mutations enriched via directed evolution are colored red and rendered in space-filling mode. Structures based on PDB ID: 1OAG⁵

Experimental Methods

Yeast strains, transformation, and culturing

See *Supplementary Table 3* for details on all plasmids and primers used in this study.

Aga2P-APEX2 (full length) yeast and Generation 2 selection yeast were generated by transformation of the yeast display plasmid pCTCON2 into the *Saccharomyces cerevisiae* strain EBY100. As previously reported^{61,85}, EBY100, propagated at 30 °C in yeast extract peptone dextrose (YPD) complete medium, was made competent and transformed using the Frozen E-Z Yeast Transformation II Kit (Zymoprep) according to the manufacturer’s protocols. Individual colonies of transformed cells containing the TRP1 gene were selected on synthetic dextrose plus casein amino acid (SDCAA) plates and propagated in SDCAA medium at 30 °C. Protein expression was induced by inoculating 4.5 mL of SGCAA (SDCAA medium with dextrose replaced by an equivalent amount of galactose) with 500 µL of overnight SDCAA yeast culture and incubating at 30 °C for 18–24 h. Successful incorporation was confirmed by sequencing of DNA isolated by Zymoprep (ZymoprepTM Yeast Plasmid Miniprep I, Zymo Research, see modified procedure below) and FACS analysis of immunostained yeast (see procedure below).

All other yeast used in this study were prepared using a two-step transformation of *S. cerevisiae* strain BJ5465, based on a previously described protocol^{61,85}. BJ5465 yeast, propagated at 30 °C in YPD complete medium, were transformed first with a linearized Yeast Integrating Plasmid (YIP) encoding Aga1p fusion constructs, plated and selected on SDCAA supplemented with 40 µg/mL tryptophan (lacking uracil). FACS analysis and DNA sequencing of PCR products amplified from extracted genomic DNA^{61,85} were used to confirm homologous recombination of the desired construct. Aga1p was constitutively expressed and immunostaining for the V5 epitope tag was present in both SDCAA and SGCAA induction media (see Figure 1E for example of yeast surface protein and epitope tags). Next, to introduce individual Aga2P constructs to generate non-library

yeast (such as starting templates and individual clones), Aga1p fusion expressing yeast were transformed with the applicable pCTCON2 plasmid and selected on SDCAA plates.

Generation of yeast libraries

For each generation of directed evolution, libraries of AP variants were generated using error-prone PCR according to published protocols^{25,61,81}. In brief, for Generation 1 library, 150 ng of the template sAPEX AP-0 gene in pCTCON2 vector was amplified for 20 rounds with 0.4 μ M forward and reverse primers, 2 mM MgCl₂, 5 units of Taq polymerase (NEB), and 2 μ M each of the mutagenic nucleotide analogs 8-oxo-2'-deoxyguanosine-5'triphosphate (8-oxo-dGTP) and 2'-deoxy-p-nucleoside-5'-triphosphate (dPTP). The PCR product was then gel-purified and amplified through three standard PCR reactions of 30 cycles with primers that introduce overlap to the vector, 100 μ M regular dNTPs (VWR) and 5 units of Taq polymerase in 50 μ L volume. In parallel, fresh electrocompetent BJ5465 yeast constitutively expressing EX-Aga1p were prepared. BJ5465 yeast expressing EX-Aga1p were passaged at least two times before this procedure to ensure that the cells were healthy. 3 mL of fresh saturated culture was used to inoculate 100 mL of fresh YPD media, and the cells were cultured with shaking at 220 rpm at 30 °C for 6–8 h until the OD₆₀₀ reached 1.5–1.8. The cells were then harvested by centrifugation for 3 min at 3000 rpm and resuspended in 50 mL of sterile 100 mM lithium acetate in water, by vigorous shaking. Fresh sterile DTT (1 M stock solution, made on the same day) was added to the yeast cells to a final concentration of 10 mM. The cells were incubated with shaking at 220 rpm for 12 min at 30 °C. Cells were then pelleted by centrifugation at 3000 rpm for 3 min at 4 °C and washed once with 25 mL ice-cold sterile water, pelleted again, and resuspended in 1 mL ice-cold sterile water. The amplified error-prone PCR inserts was electroporated with BamHI-NheI linearized pCTCON2 vector (10 μ g insert:1 μ g vector) backbone into electrocompetent yeast expressing EX-Aga1p. Electroporation was performed using a Bio-Rad Gene pulser XCell with the following settings: 500-V, 15-ms pulse duration, one pulse only, 2-mm cuvette. The electroporated cells were immediately rescued with 2 mL pre-warmed YPD media and then incubated at 30 °C for 1 h without shaking. 10 μ L of this solution was used to determine transformation efficiency, and the remainder was pelleted to remove the YDP media, and resuspended in 100 mL SDCAA media supplemented with 50 units/mL penicillin and 50 μ g/mL streptomycin. The culture was grown at 30 °C with shaking at 220 rpm for 1 d, before induction of protein expression for analysis by FACS

and Zymoprep. Transformation efficiency was 2.9×10^7 . DNA sequencing of 21 distinct colonies showed an average of 1.4 nucleotides changed per clone with a range of 0–4. For Generation 2 AP-1 library, the same procedure was employed except electroporation was performed on *S. cerevisiae* EBY100. Transformation efficiency was 3.5×10^7 . DNA sequencing of 24 distinct colonies showed an average of 1.2 nucleotides changed per clone with a range of 0–5. For Generation 3 AP-2 library, the same procedure was employed except that the rate of mutagenesis was higher; ten rounds of error-prone PCR were performed with 15 μ M of each of the mutagenic nucleotide analogs. Electroporation was performed on BJ5465 cells stably expressing acid coil-Aga1p. Transformation efficiency was 3.1×10^7 . DNA sequencing of 22 distinct colonies showed an average of 7.3 nucleotides changed per clone with a range of 0-12. For generation 4 selections, two libraries derived from AP-3 were combined: one generated by ten rounds of error-prone PCR performed with 10 μ M each of the mutagenic nucleotide analogs and another generated by 20 rounds of error-prone PCR performed with 4 μ M each of the mutagenic nucleotide analogs. Transformation efficiency of the respective libraries were 2.0×10^7 and 2.8×10^7 . DNA sequencing of 24 distinct colonies from each library showed an average of 5.6 and 2.9 nucleotides changed per clone, with a range of 0-9 and 0-4, respectively.

Yeast display selections

For each generation of selections, yeast libraries (as described above) were induced by transferring them to 1:9 SDCAA/SGCAA media and growing the cells for 18–24 h at 30 °C with shaking at 220 rpm. Overnight yeast cultures were pelleted by centrifugation (3000 rpm, room temperature, 3 min), washed with room temperature 0.1% BSA (w/v) DPBS (DPBS-B) twice, and resuspended in an equivalent volume of DPBS-B. The OD600 for a 1/100 dilution of the cell suspension was measured (1 OD600 \sim 1×10^7 yeast cells). An aliquot of yeast with at least a tenfold coverage above the diversity of the yeast pool was pelleted at $7000 \times g$ for 1 min. This oversampling was employed for all passaging and labeling steps except for the first-round sort of each generation, in which we aimed for threefold oversampling during labeling. This lower sampling in the first round was a result of the very large diversity size and the practical difficulties and expense of labeling large numbers of yeast cells. Prior to the first round of sorting, “diversity” was considered to be the number of transformants in the initial library. In subsequent rounds, the “diversity” was the number of yeast cells collected in the previous round of sorting. Yeast libraries from Generations

2-4 were first incubated with heme (see details above); to implement both the positive and negative selections, we added concentrated EX fragment (expressed and purified from *E. coli* as described below) to the heme-containing yeast DPBS-B culture. For more details regarding the presence of base coil, concentrations, and time of protein and heme incubations, see Figures S3B, S4B, S5A. Yeast were then washed twice with DPBS-B and resuspended in a room temperature solution consisting of 100 μ M biotin-phenol and 1 μ M H₂O₂ at a density of 2.5×10^6 cells/mL (7.5×10^6 to 1.0×10^7 cells/mL for first round sorts) by rapid vortexing. After 1 min, the yeast suspension was quenched by addition of an equal volume of 20 mM Trolox and 10 mM sodium ascorbate in DPBS-B, followed by rapid vortexing. Minimal intercellular labeling was observed under these conditions. The cells were pelleted at $14,000 \times g$ for 1 min and washed twice with DPBS-B. For two-dimensional FACS sorting, samples were then resuspended in 50 μ L of mouse anti-c-Myc (Calbiochem, 1:400) and incubated at 4 °C with rotation for 1 h. Samples were washed twice with DPBS-B and resuspended in 50 μ L of streptavidin-phycoerythrin (Jackson ImmunoResearch, Cat. No. 016-110-084, 1:250) and Alexa Fluor 647 goat anti-chicken IgG (Life Technologies, 1:300). After 1 h of incubation with rotation at 4 °C, samples were washed twice, and the labeled and fluorophore-stained yeast cells were resuspended in DPBS-B to a concentration of $4.5 \times 10^7 - 1.2 \times 10^8$ cells/mL and sorted by FACS into a collection tube containing 5 mL SDCAA with 50 units/mL penicillin and 50 μ g/mL streptomycin. Collected yeast cells were immediately placed back into a 30 °C incubator and cultured with shaking at 220 rpm. Yeast cells were grown until saturation and passaged once more in SDCAA; depending on the number of cells collected, it took 1-2 days to reach initial saturation. 500 μ L of fresh saturated yeast culture was added to 4.5 mL of SGCAA for induction, labeling, and subsequent rounds of sorting or analysis. For positive selections, trapezoidal gates were drawn to collect yeast cells positive for AF647 signal that also had high PE signal; these were the cells with a good activity to expression ratio. During generation 4 negative selections, generous rectangle gates were drawn to collect all AF647-positive yeast that were negative for PE signal, representative of cells with no activity despite robust expression.

We used a BD FACS Aria II cell sorter with the 647-nm laser and appropriate emission filters (657 low-pass (for AF 647), 580/30 for PE).

Sequence analysis of individual yeast clones

Plasmid DNA from yeast cultures was harvested using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) according to the manufacturer's instructions for liquid culture except for the following. 6 μ L Zymolase was added for a total incubation time of 1-3 h, rotating at 37 °C. After the addition of solution 2, the mixture was briefly vortexed. DNA was eluted with 10 μ L of warm water, and all eluted DNA were transformed into XL1Blue *E. coli* cells by heat-shock transformation and rescued for 1 h at 37 °C. The entire culture was then plated on an LB-ampicillin plate, and 18–36 colonies were picked for sequencing of individual clones.

FACS analysis

Yeast samples taken from picked colonies of transformed clones or from samples regrown after sorting were passaged in SDCAA overnight. Expression was induced by inoculating 4.5 mL of SGCAA with 500 μ L of fresh saturated SDCAA yeast culture and incubating at 30 °C with shaking at 220 rpm for 18–24 h. Labeling was performed as described above, and yeast populations and clones were analyzed by FACS (BD LSR II flow cytometer, BD Biosciences) with 488-nm and 640-nm lasers and 582/42 (for phycoerythrin) and 670/30 (for AlexaFluor647) emission filters.

Expression and purification of C-terminal sAPEX fragments (EX) from E. coli

For positive selections employed in Generation 2 and negative selections in Generation 4, we cloned N-terminal His6-tagged EX-flag-GFP into pYFJ16 vector for *Escherichia coli* (*E. coli*) expression. For positive selections employed in Generations 3 and 4, we added a base coil between the His6-tag and EX. Homemade competent BL21 cells, derived from BL21-CodonPlus(DE3)-RIPL *E. coli*, were transformed with pYFJ16 constructs and plated on LB-amp plates. Individual colonies were amplified in two separate 5mL cultures containing B (Luria-Bertani, Miller) media with 100 mg/L ampicillin at 37 °C overnight with shaking. The 10 mL saturated culture was then used to inoculate 1 liter LB with 100 mg/L ampicillin, which was grown at 25 °C with shaking until OD600 = 0.8 at 25 °C. Protein expression was induced with 1 mM isopropyl β -d-1-thiogalactopyranoside (IPTG, EMD Millipore), and cultures were grown with shaking at 220 rpm at 25 °C for another 20 h. Bacteria were pelleted by centrifugation at 6000 rpm for 10 min at room temperature and immediately lysed with B-PER buffer (Thermo Fisher Scientific) supplemented with 1 \times protease cocktail and 1 mM PMSF at 20 mL per 1 liter of original culture. All subsequent steps for protein purification were performed on ice or at 4°C. The lysate was clarified by

centrifugation at 10000 rpm for 15 min, and the supernatant was incubated with 1 mL of Ni-NTA agarose beads suspension (Qiagen) in binding buffer (50 mM Tris, 300 mM NaCl, pH = 7.8), in a total volume of 45 mL, for 20 min with rotation and then transferred to a gravity column. The beads were washed with 10 mL binding buffer followed by 30 mL washing buffer (30 mM imidazole, 50 mM Tris, 300 mM NaCl, 1 mM DTT, pH = 7.8), then the protein was eluted with 7 mL of elution buffer (200 mM imidazole, 50 mM Tris, 300 mM NaCl, pH = 7.8). Fractions were collected and transferred to a centrifugal filter Amicon Ultra-15 and exchanged 3 times into ice-cold DPBS (Dulbecco's Phosphate Buffer Saline) and concentrated. Protein samples were flash frozen in liquid N₂ and stored in aliquots at -80 °C. Purity was checked using SDS-PAGE. Concentrations of total protein content were determined using the bicinchoninic acid (BCA) Protein Assay Kit (Pierce).

Plasmid name	plasmid vector	promoter	Expression in	Features
p71	PCTCON2	Gal1	yeast	Aga2p HA-33 linker- full length APEX2
p72	PCTCON2	Gal1	Yeast	Aga2p- HA 33 aa linker aa52-250-5aa linker-aa2-51myc
p73	PCTCON2	Gal1	Yeast	Aga2p- HA 33 aa linker aa90-250-5aa linker-aa2-89myc
p74	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-AP-0-myc
p75	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-AP-1-myc
p76	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-AP-2-myc
p77	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-AP-3-myc
p78	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-AP-myc
p79	YIP	GPD	Yeast	acid coil-5aa linker-V5-11aa linker -Aga1p
p80	YIP	GPD	Yeast	V5-EX-10aa linker-Aga1p
p81	YIP	GPD	Yeast	V5-aa2-89-10aa linker-Aga1p
p82	YIP	GPD	Yeast	V5-aa90-250-10aa linker-Aga1p
p83	YIP	GPD	Yeast	V5-aa2-200-10aa linker-Aga1p
p84	YIP	GPD	Yeast	V5-aa2-89-18aa linker-Aga1p
p85	YIP	GPD	Yeast	V5-aa90-250-18aa linker-Aga1p
p86	PCTCON2	Gal1	Yeast	Aga2p-HA-33aa linker-90-250-myc
p87	PCTCON2	Gal2	Yeast	Aga2p-HA-33aa linker-2-89-myc
p88	PCTCON2	Gal3	Yeast	Aga2p-HA-33aa linker-201-250-myc
p89	PCTCON3	Gal4	Yeast	Aga2p-HA-45aa linker-90-250-myc

p90	PCTCON4	Gal5	Yeast	Aga2p-HA-45aa linker-2-89-myc
p91	PYFJ16	PMSF	E.coli	His6x-EX-linker-GFP
p92	PYFJ16	PMSF	E.coli	His6x-EX-basecoil containing linker-GFP

References

1. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).
2. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).
3. Morawski, B. *et al.* Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. *Protein Eng.* **13**, 377–384 (2000).
4. Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).
5. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
6. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
7. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* (2011). doi:10.1073/pnas.1101046108
8. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* (2002). doi:10.1073/pnas.082243699
9. O’Shea, E. K., Lumb, K. J. & Kim, P. S. Peptide ‘Velcro’: Design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658–667 (1993).
10. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science (80-.)*. **339**, 1328–1331 (2013).
11. Sharp, K. H., Mewies, M., Moody, P. C. E. & Raven, E. L. Crystal structure of the ascorbate peroxidase-ascorbate complex. *Nat. Struct. Biol.* **10**, 303–307 (2003).
12. Wang, W. *et al.* A light- and calcium-gated transcription factor for imaging and manipulating activated neurons. *Nat. Biotechnol.* **35**, 864–871 (2017).

Chapter 4: Testing different clones of split APEX in Mammalian Cells

Introduction

Due to the lack of robust activity of existing split APEX2 fragments (Chapter 2), yeast surface-display directed evolution was utilized to find favorable mutations to improve the activity of split APEX. Utilizing yeast display, our final goal was to find an AP mutant with low affinity for the EX fragment, yet still demonstrated high activity upon close-proximity reconstitution. The previous chapter discussed the evolution of the large fragment over four generations, as well as the evolution of our display platform.

For every generation of evolution, we utilized zymopreps to extract plasmid DNA from the yeast library pools of the last few rounds of sorts to isolate individual clones and to assess the level of convergence^{25,81,82}. To determine the best clone and the starting template for the subsequent generation of evolution, we tested enzymatic activity preliminarily on the cell surface of yeast by FACS analysis. We next tested the clones with beneficial mutations in the mammalian cell cytosol using confocal microscopy to see if improvement in activity was still upheld. This environment is quite different from the yeast cell surface, as it is 37 °C instead of 30 °C, and a reducing rather than oxidizing environment.

The clones of interest selected from the yeast display evolution were chosen based on how frequently they appeared, indicating possible convergence, as well as unique clones from later rounds of zymoprep analysis. Rare clones might still contain beneficial mutations that fulfilled the increasingly stringent selection conditions. Mutations that provide a boost in activity on the yeast surface may not necessarily translate within the cytosol of mammalian cells, thus it is advantageous to test keep a larger sample space. Beneficial mutations were combined and tested in parallel.

Determining the best clone from each generation of directed evolution

For each generation of evolution, we needed to determine the winning clone for mammalian applications. One end-point application we had in mind was to determine the proteome of the mito-ER junction (see chapter 5 for more information); thus, we had tested the different clones as targeted fusions to the mitochondria and the EX as a fusion to an ER targeting sequence (top two constructs in **Figure 4-1A**). We had also wished to test the different clones as soluble

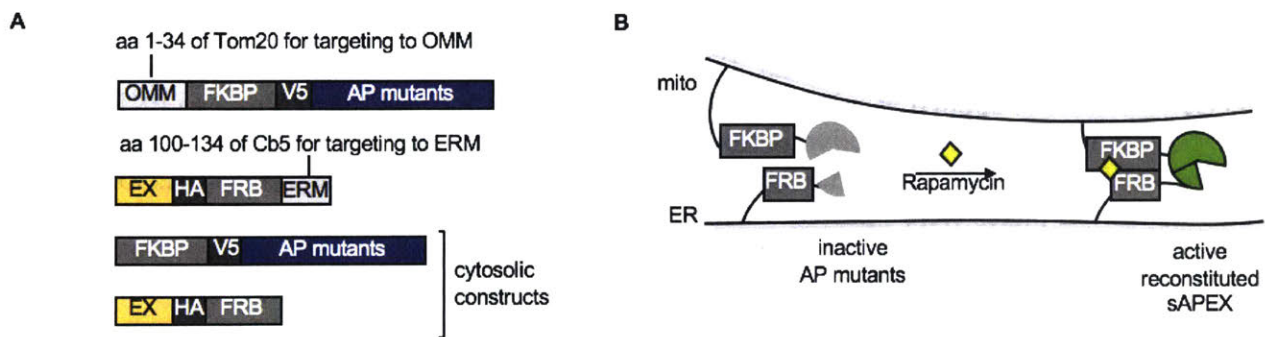
proteins rather than as membrane-anchored constructs with restricted geometry. Hence, we expressed the N- and C-terminal fragments as fusions to FKBP and FRB, respectively (bottom two constructs in **Figure 4-1A**).

Generation 1 - directed evolution of AP-0

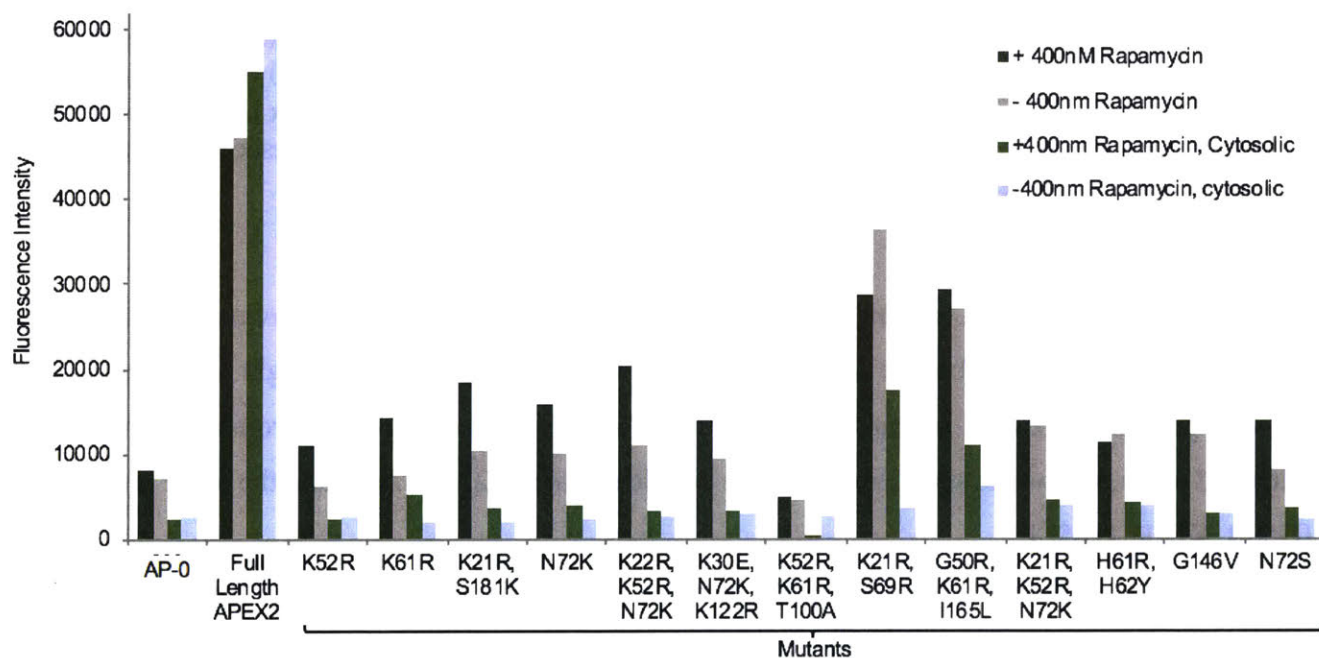
Mutagenized AP-0 clones yielded from the first generation of yeast display were cloned into an optimized mito-ER targeting scaffold (**Figure 4-1A; 4-1B**, see Chapter 5 for more details) to test how their mutations contribute to enzymatic activity as compared to the original template, AP-0. We compared peroxidase activity with or without rapamycin, using both our Amplex UltraRed and biotin-phenol labeling assays, thus enabling us to both rapidly and rigorously compare the activity of our AP-0 mutants.

Cytosolically-targeted constructs were also cloned (**Figure 4-1A**) in which the mito and ER targeting sequences were omitted. Cytosolically localized fragments provide a clearer characterization of fragment affinity; if these constructs had high affinity for each other, fluorescence intensity would not be rapamycin dependent. When targeted to the mito and the ER, the local concentration is much higher, potentially obscuring low affinity properties. Though we ultimately desired a mutant that had the right properties for an application such as ones at the mito-ER junction space, we needed to identify a promising starting point first.

First, to quickly obtain unbiased information on how the mutations affected enzymatic activity on a large population of cells, fluorescence intensity from the Amplex UltraRed assay was measured using a plate reader instead of confocal microscopy, which only visualizes a few cells per field of view. (**Figure 4-1C**) compares the fluorescence intensities from 20-minute Amplex UltraRed labeling of mito-ER targeted split APEX2 cut site 200 variants with that of the 30-minute Amplex UltraRed labeling of cytosolically-targeted when transfected with equal stoichiometries of plasmid DNA to HEK 293T cells plated with the same density. While not every clone provided a boost in activity, it is clear that many did outperformed AP-0, the original starting template. Although none of the mutants produced enough resorufin product to matched the fluorescence intensity of full length APEX2 in 20 minutes, mutants [K21R, S69R] and [G50R, K61R, I165L] were competitive. When AP-0 mutants were cytosolically localized, the overall detected enzymatic activity was considerably lowered, but general trends in enzymatic activity between the mito-ER and cytosolic targeted remained consistent. However, a few key differences were spotted regarding



C Amplex UltraRed Assay of AP-0 mutants



D Amplex UltraRed assay of top two AP-0 mutants

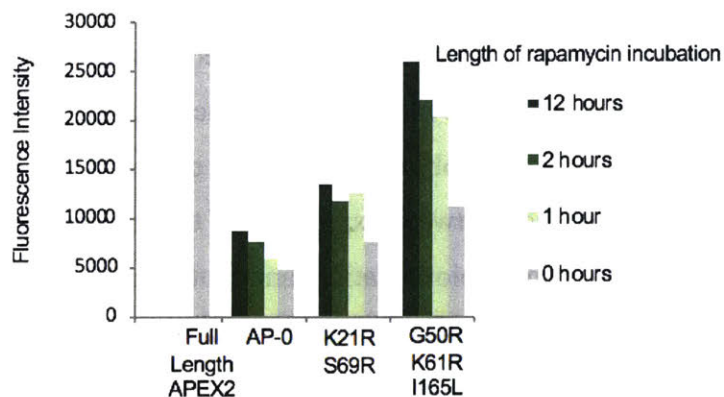


Figure 4-1. Amplex UltraRed Assay of AP-0 mutants derived from first generation of yeast display evolution. (A) Constructs used for targeting AP and EX fragments of sAPEX to the outer mitochondrial membrane (OMM) and ER membrane (ERM), respectively. (B) Schematic overview of sAPEX applied at mito-ER contacts. Inactive fragments (grey) fused to FKBP and FRB can reconstitute to give active peroxidase (green) when driven together by an inducible protein-protein interaction (PPI) with the addition of rapamycin (red triangle). (C) HEK293T cells plated in 48-well plates were lipofected with split APEX fragment pairs targeted to the mito-ER or the cytoplasm (A). The large fragment was either the original template (AP-0) or a mutant found from yeast display. Full-length APEX2 targeted to the cytosol was used as a positive control. After 23 hours of protein production, cells were incubated with +/- 400 nM rapamycin for 30 minutes. Amplex UltraRed assay was performed by incubation with 100uM Amplex UltraRed for 20 minutes on ice for Mito-ER targeted and 30 minutes for cytosolically-targeted. Fluorescence intensity was collected by plate reader. (D) HEK 293T cells were transfected to express fragments targeted to the mito-ER (top two constructs in A) to examine the effect of increasing rapamycin incubation time, effectively also increasing reconstitution time. AP-0 template was compared to the two top-performing mutants: [K21R, S69R] and [G50R, K61R, I165L]. 400 nM rapamycin was added during a 22-hour protein production period for 12, 2, 1, 0 hours. Amplex UltraRed assay was performed by incubation of Amplex UltraRed for 30 minutes on ice. Fluorescence intensity was collected by plate reader.

the rapamycin dependency. Mutant [K21R, S69R] appeared to have high rapamycin dependency in the cytosol for instance, but the decreased enzymatic activity in the cytosol might be a result of reduced protein expression or maybe protein instability/degradation. Overall, from the Amplex UltraRed assay, the two best mutants were the double mutant [K21R, S29R] and the triple mutant [G50R, K61R, I165L].

After preliminary screening by the Amplex UltraRed assay, mutants that demonstrated more robust reconstitution were selected for testing via biotin-phenol labeling. Confocal microscopy was performed on transfected HEK 293T cells to assess the level of biotinylation (**Figure 4-2A**), as well as to get a preliminary idea of how the expression of AP and EX may affect organelle morphology (**Figure 4-2B**). In general, while it did appear that the targeting sequence dictated the morphological pattern, there were many instances in which one construct adopted the morphology of its complimentary fragment. This typically was observed when one construct was extremely over-expressed. We also noted a weak correlation between expression of the fragments and detected biotinylation. There were cells with strong biotinylation and both constructs expressed at low levels; conversely, there were also instances of cells with moderate expression levels of both construct and no detected biotinylation. With hindsight, a major contributing factor to the variance might have been insufficiencies in heme bioavailability in the cell.

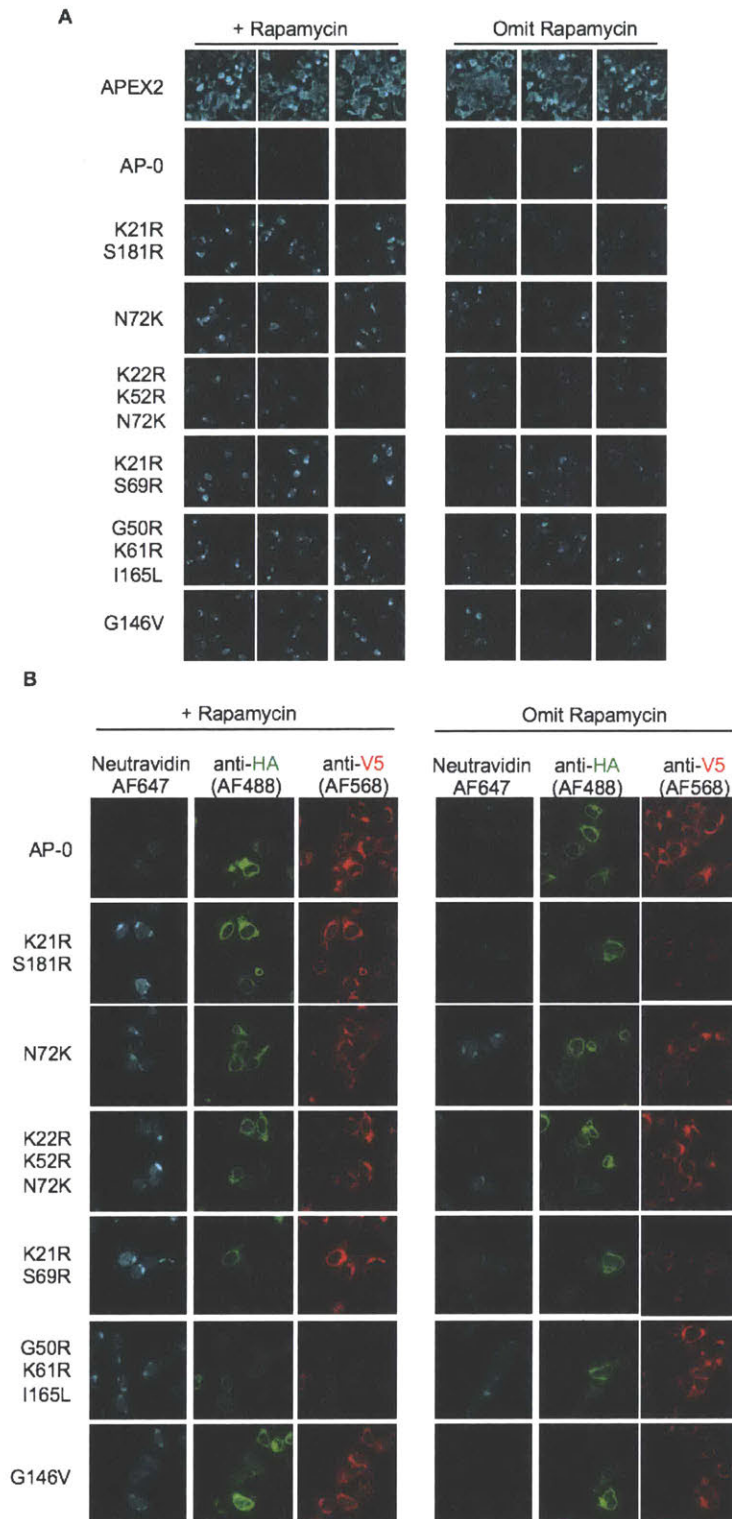


Figure 4-2. Biotin-phenol labeling on live HEK 293T cells comparing the top mutants of Generation 1 evolution against APEX2, and starting template AP-0 (A) HEK293T cells plated in 48-well plates were transfected with both the mito and ER targeted constructs from (**Figure 4-**

1A) in which the AP-0 fragment could be that of the original template, or a top mutant from the Amplex UltraRed assay from (**Figure 4-1C**). Full-length APEX2 targeted to the cytosol was used as a benchmark. After 24 hours of protein production, cells were incubated with +/- 400 nM rapamycin for 30 minutes and subjected to biotin-phenol labeling with standard conditions, followed by immunostaining and staining with nAv647. 3 representative fields of view of nAv647 staining are shown from 43x magnification. Confirmation that constructs expressed at similar level across these fields was done by immunostaining for the epitope tags, V5 and HA. (**B**) Biological replicate of the same experiment, with confocal microscopy performed at 63x magnification. Single representative field of view is shown for each condition.

Consistently in both assays, the enzymatic activity was not very significantly dependent on the 30-minute incubation with 400 nM rapamycin. This indicated that these constructs might be either high affinity and/or require longer incubation times with rapamycin due to slower kinetics in enzymatic reconstitution. To follow up on the possibility of lengthier enzymatic reconstitution kinetics, the same Amplex UltraRed assay was performed on the two clones that exhibited the highest enzymatic activity in both assays: [K21R, S69R] and [G50R, K61R, I165L]. However, 400nM rapamycin incubation was extended to 1, 2, and 12 hours (**Figure 4-1D**). Longer incubation times with rapamycin did increase enzymatic activity. The effect was very dramatic for AP-0 mutant [G50R, K61R, I165L], which when incubated with 400 nM rapamycin for 12 hours, was able to generate comparable levels of fluorescence intensity as full length APEX2. The increase in intensity can potentially be attributed to longer rapamycin incubation times allowing for both more protein-protein interactions to form, stabilizing fragments that would otherwise be degraded, and increasing fragment reconstitution time.

Utilizing these two different assays in two different cellular contexts, two clones stood out in particular due to their robust reconstitution in HEK293T cells: [K21R, S69R] and [G50R, K61R, I165L]. However, when we tested these clones at the mito-ER junction via transfection in Cos7 cells, we were unable to detect any biotinylation. Ultimately, we chose triple mutant [G50R, K61R, I165L] as our finalized first-generation clone, named AP-1. We selected this over the comparable double mutant due to observations of slightly enhanced expression/transfection efficiencies via immunostaining post biotin-phenol labeling. Unfortunately, the enzymatic activity observed in HEK 293T cells was not very PPI-dependent – likely a reflection of the selection platform. These results indicated that further evolution was required on both the AP-1 fragment and the selection platform.

Generation 2 - directed evolution of AP-1

To further improve the reconstituted activity of sAPEX, we continued to perform a second generation of directed evolution. The display platform was modified to provide increased control over the selection environment. C-terminal EX fragment was exogenously added as a purified, soluble protein to allow for precise control of the concentration of EX added and the time of incubation. We hoped to select for mutations that improve AP-1 expression and stability in the absence of EX translation. After the fourth round of selection, a zymoprep was performed to isolate the plasmid DNA of the remaining yeast pool. The isolated DNA was transformed into *E. coli* and 35 clones were sequenced. Of the 35 clones, 18 contained a mutation on the N72 residue: 10 N72S, 7N72K, and 1 N72H (15 were unique sequence, i.e. not a duplicate/convergent sequence. The mutations K21R, R24G, K52R, and H62Y also came up at least 3 times each in unique sequences.

Amplex UltraRed assay was utilized again as a preliminary screen of all the different mutants derived from the second generation of evolution. HEK 293T cells were transfected to express AP-1 variants targeted to the mito and EX targeted to the ER (**Figure 4-1A**). After overnight protein synthesis, cells were treated with 400 nM of rapamycin for two hours, or treated with vehicle. While the Amplex UltraRed assay is great for detecting peroxidase activity, the high sensitivity of the assay provided less discernable differences in peroxidase activity between the template, AP-1, and the different mutants (**Figure 4-3A**). Reproducibly better mutants from the Amplex UltraRed assay were tested by biotin-phenol assay. Hek293T cells were transfected to express top AP-1 variants targeted to the mito and EX targeted to the ER (**Figure 4-1A**). Initial attempts to use live biotin phenol labeling to assay for activity included a lot of variability in detectable biotinylation from cell to cell. Variability amongst cells might have been a result of varying expression level of both fragments or heme concentrations. To analyze the collected imaging data, the neutravidin-647 intensity, reflecting extent of biotinylation, was averaged over 10 fields of view (**Figure 4-3A**). Consistently through both assays, triple mutants [H52R, H62Y, N72S] and [R24G, H62Y, N72S] demonstrated robust reconstituted peroxidase activity.

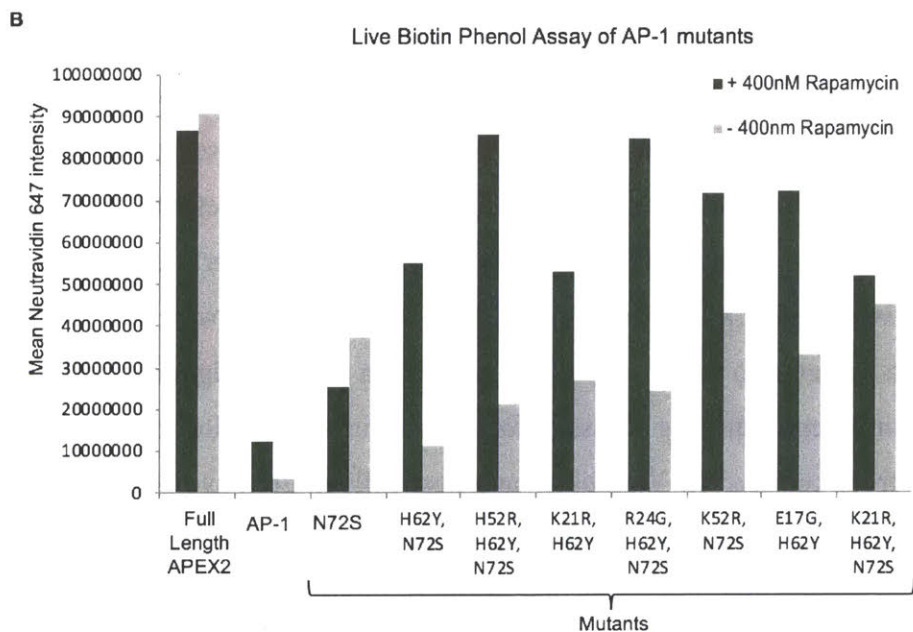
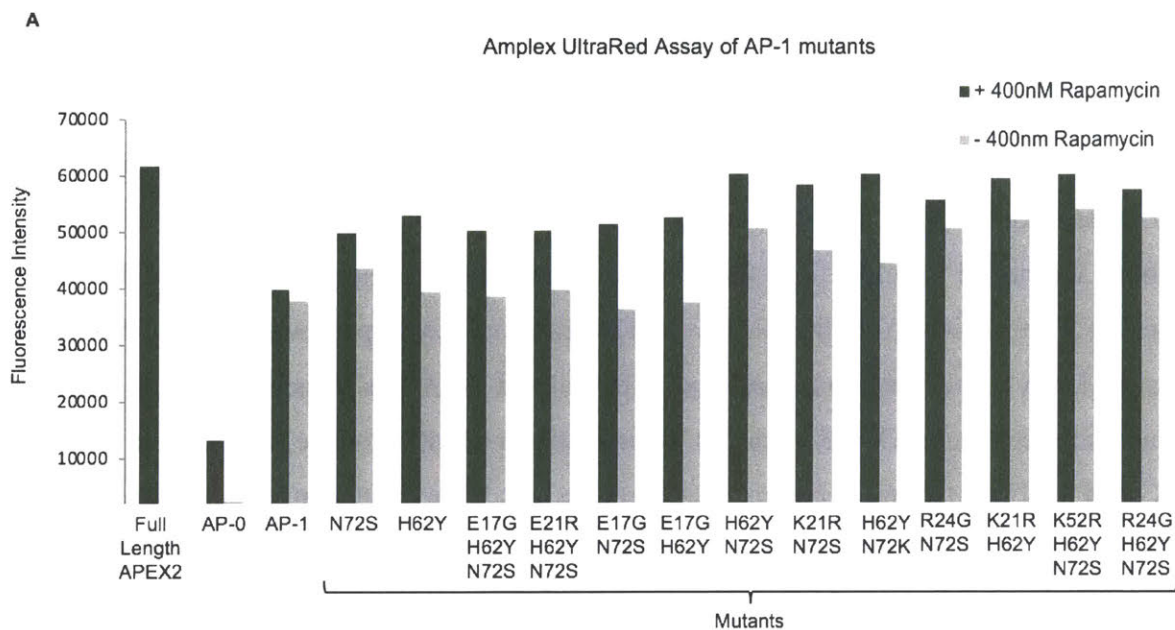


Figure 4-3. Amplex UltraRed and Biotin-phenol labeling of AP-1 mutants derived from second generation of yeast display evolution. (A) HEK293T cells plated in 48-well plates were lipofected with split APEX fragment pairs targeted to the mito-ER (top two constructs of **4-1A**). The large fragment was either the original template (AP-1) or a mutant found from yeast display. Full-length APEX2 targeted to the cytosol was used as a positive control. After 22 hours of protein production, cells were incubated with +/- 400 nM rapamycin for 2 hours. Amplex UltraRed assay was performed by incubation with 100uM Amplex UltraRed for 30 minutes on ice. Fluorescence intensity was collected by plate reader. (B) Top performers from (A) were tested by live biotin-

phenol labeling assay. HEK293T cells plated in 48-well plates were transfected with both the mito and ER targeted constructs from (Figure 4-4A) in which the mito fragment could be that of the original template AP-1, or a top mutant from the Amplex UltraRed assay. Full-length APEX2 targeted to the cytosol was used as a benchmark. After 24 hours of protein production, cells were incubated with +/- 400 nM rapamycin for 2 hours and subjected to biotin-phenol labeling with standard conditions, followed by immunostaining and staining with neutravidin-647. Average neutravidin-647 intensity for 10 fields of view are shown for each condition. Confirmation that constructs expressed at similar level across these fields was done by immunostaining for the epitope tags, V5 and HA.

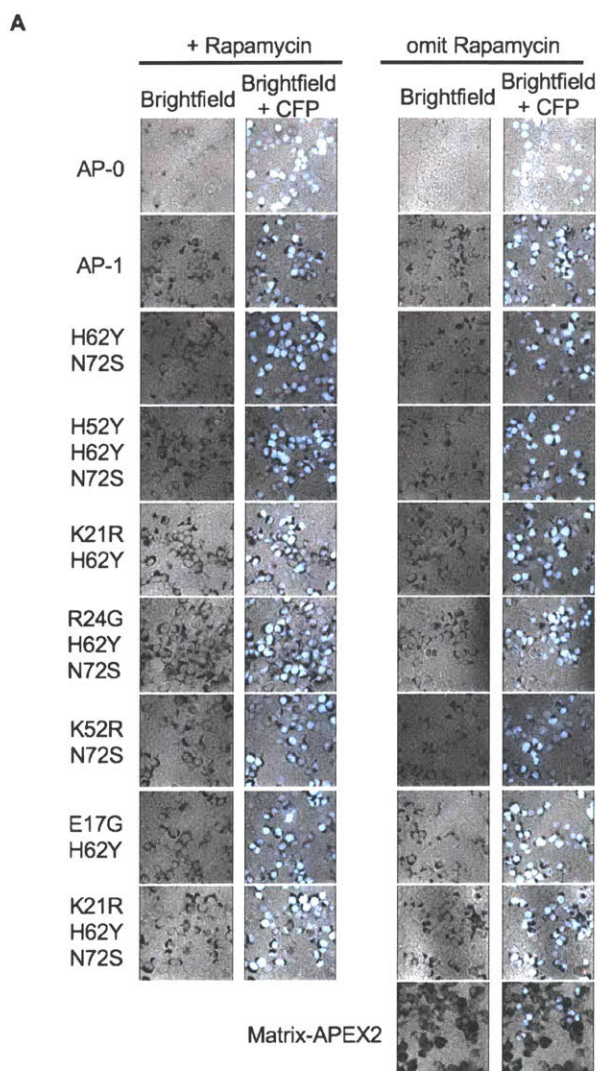


Figure 4-4. DAB labeling of AP-1 mutants derived from second generation of yeast display evolution. (A) HEK293T cells plated in 48-well plates were lipofected with split APEX fragment pairs targeted to the mito-ER (top two constructs of 4-1A). After overnight protein synthesis, cells

were treated with or without 400 nM of rapamycin for one hour. Cells were then fixed and labeled with DAB for 45 minutes. DAB polymer formation is a product of peroxidase activity and generates a dark staining pattern that can be visualized by confocal microscopy, here taken at 40x. After DAB labeling, immunostaining is not possible so CFP-NLS was transfected in tandem as a transfection marker to indicate which cells were likely transfected with split APEX variants as well.

With improved activity, mutants were also screened using the DAB assay, which has lower sensitivity and requires more robust peroxidase activity. Previously, we did not utilize this assay because it lacked sensitivity. HEK 293T cells transfected to express original split APEX2 with fragments AP-0 and EX targeted to the mito and ER, respectively, did not give very robust/detectable signal (**Figure 4-4A**). When comparing AP-1 against the top performing mutants, it did indeed appear that the second-generation selection platform did seem to improve not only activity, but also PPI-dependency. There was a reproducible and increased rapamycin dependency for reconstituted peroxidase activity across all three assays. Mutants overall had comparable or improved activity.

Triple mutant AP-1 clone [R24G, H62Y, N72S] reproducibly demonstrated robust activity in all three assays, contained very frequent and beneficial mutations, and was selected as the finalized mutant and named AP-2. Furthermore, during the second-generation evolution in yeast, efforts were made to generate HEK 293T and COS7 cells that stably expressed the small fragment targeted either to the ER or just in the cytosol for screening purposes (more details in Chapter 5). With cells stably expressing one of the fragments, screening experiments performed would only require the introduction of one fragment. Furthermore, stable cells express fragments in all cells and at lower expression levels that may be less physiologically perturbing. We also tested viral introduction of the large fragment for even more reduced fragment concentrations.

Within the lowered expression level regime, it was exciting to get detectable biotinylation activity with AP-2 when it was below detection for previous generations (**Figure 4-5A**). While only one FOV is shown in the figure, there were also noteworthy examples of cells that had detectable biotinylation without any rapamycin addition even at these reduced expression level regimes. Overall activity at these levels also is not comparable to full length APEX2. And comparing the number of cells in total that express the both fragments and the number of cells that give detectable signal, it is clear that there were still many inactive cells. This could again be a

result of heme bioavailability, but also further emphasizes the need to further improve activity and reduce background.

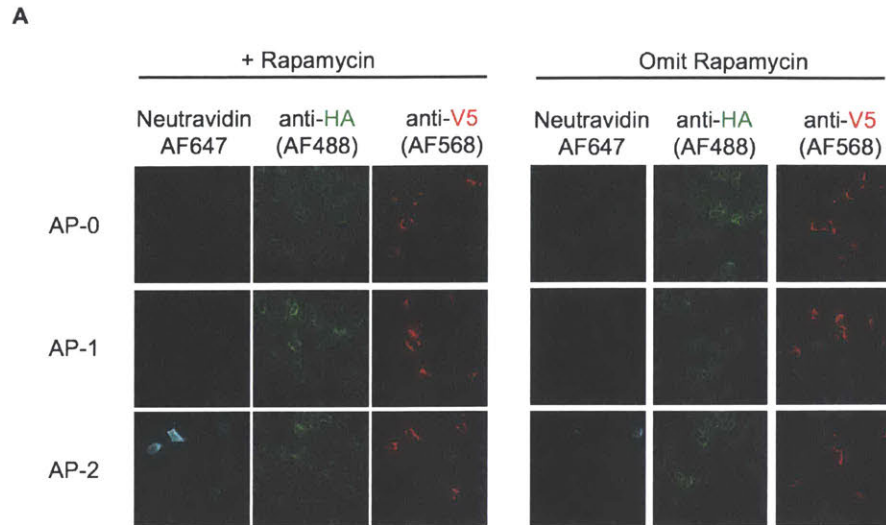
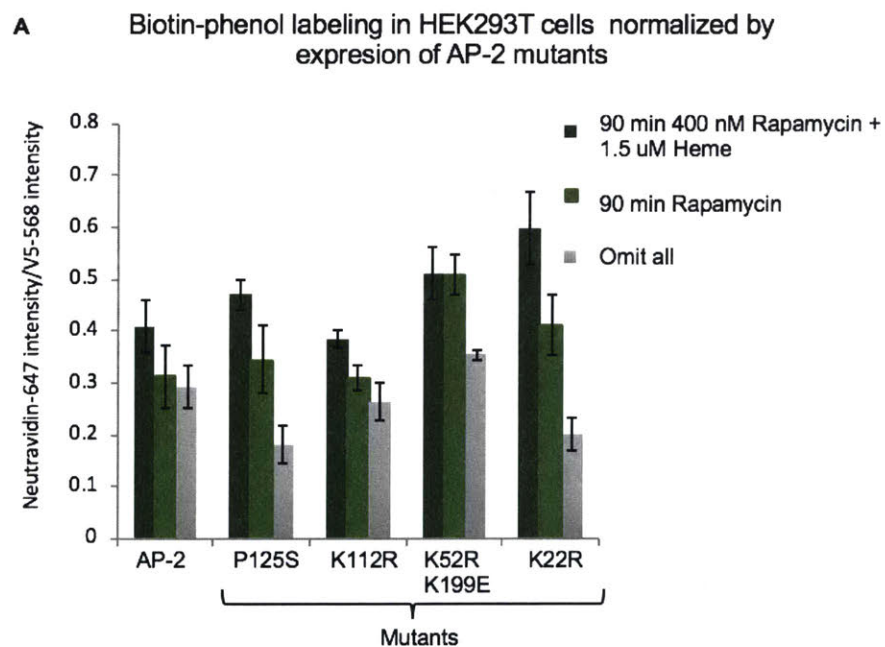


Figure 4-5. Biotin-phenol labeling comparing AP-0, AP-1 AP-2 in HEK 293T cells stably expressing EX-HA-FRB-ERM (A) HEK 293T that stably express EX-HA-FRB-ERM were infected with lentivirus that contains to express Mito-FKBP-AP-*, where* = 0 (unevolved), 1 (best mutant from first-generation of evolution), or 2 (best mutant from second generation of evolution). After 48 hours of expression, cells were treated with or without 400 nM or rapamycin for one hour before incubation with biotin-phenol for 30 minutes. The experiment was done without heme supplementation. Cells were then biotin-phenol labeled with standard conditions, and were stained for visualization by confocal microscopy.

Generation 3 - directed evolution of AP-2

In generation 3, we aimed to design a yeast-display selection platform that mimics the sAPEX reconstitution that would occur in our target biological applications (see chapter 5). To drive the fragments together using a protein-protein interaction (PPI) we again added EX as a soluble protein, but we used an artificially designed coiled-coil system, ACID-p1 and BASE-p1⁸⁴ to recruit EX to the proximity of the N-terminal fragment (see chapter 3 **Figure 3-6A**). Similar strategies were employed for generation 3. Zymopreps were performed to isolate winning mutants and tested by the more stringent biotin-phenol (**Figure 4-6**). and DAB labeling assays in mammalian cells (**Figure 4-7**).

4 clones were isolated from zymopreps of round 3 and round 4 of selection. Number of unique clones from the generation 3 were few in number but did contribute to an increase in activity. Double mutant [K52R, K199E] and single mutant K22R performed well in all assays. Including heme supplementation did rescue some activity for almost all mutants, and starting template (**Figure 4-6A**). Retrospectively, the heme may have rescued the activity for the minus rapamycin condition as well, and future testing did end up taking that into consideration. The decreased heme availability in COS7, and perhaps also the greater spacing between organelle, helped contribute to really distinguishing mutants from the starting template (**Figure 4-6B**).



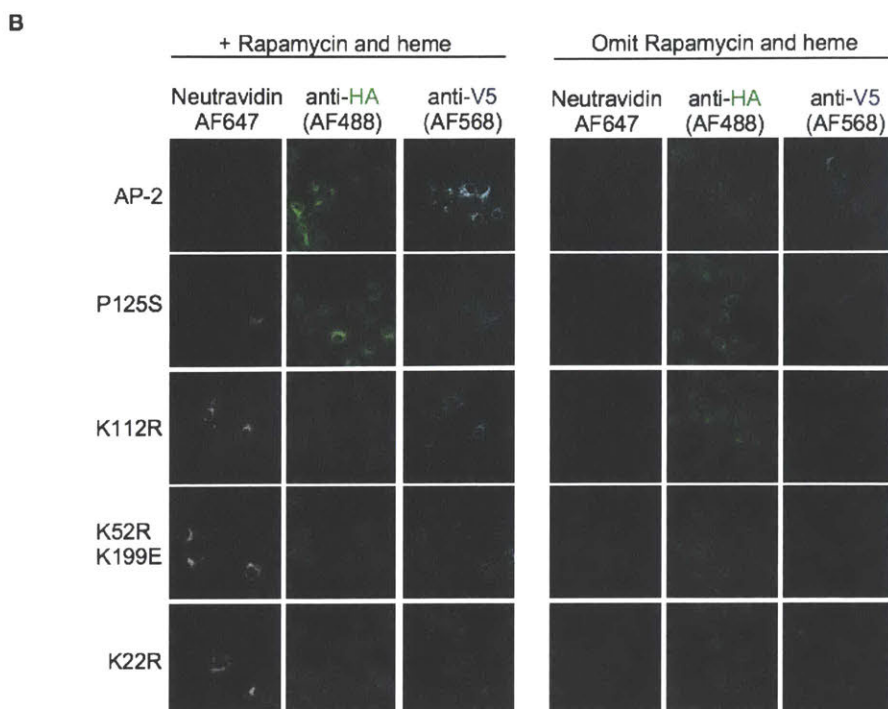


Figure 4-6. Biotin-phenol labeling of AP-2 mutants derived from third generation of yeast display evolution in stable EX-FRB-ERM cells. (A) Monoclonal HEK 293T cells stably expressing the small fragment, EX-FRB-ERM, at a low level were plated in 48-well plates and infected with the complementary mito-targeted AP-2 variants. After 48 hours of expression, cells were treated with or without 400 nM rapamycin for 90 minutes before incubation with biotin-phenol for 30 minutes. 1.5 μ M heme was supplemented in tandem with rapamycin for one of the conditions before being subjected to biotin-phenol labeling with standard conditions, immunostaining and staining with neutravidin-647. Average neutravidin-647 intensity for 10 fields of view are shown for each condition. Error bar shown is standard error. Confirmation that constructs expressed at similar level across these fields was done by immunostaining for the epitope tags, V5 and HA. (B) Same as (A), except cells utilized were polyclonal COS7 cells stably expressing EX-ERB-ERM. Cells were incubated with 400 nM rapamycin and 1 μ M heme for one hour before labeling, immunostaining, and imaging by confocal microscopy.

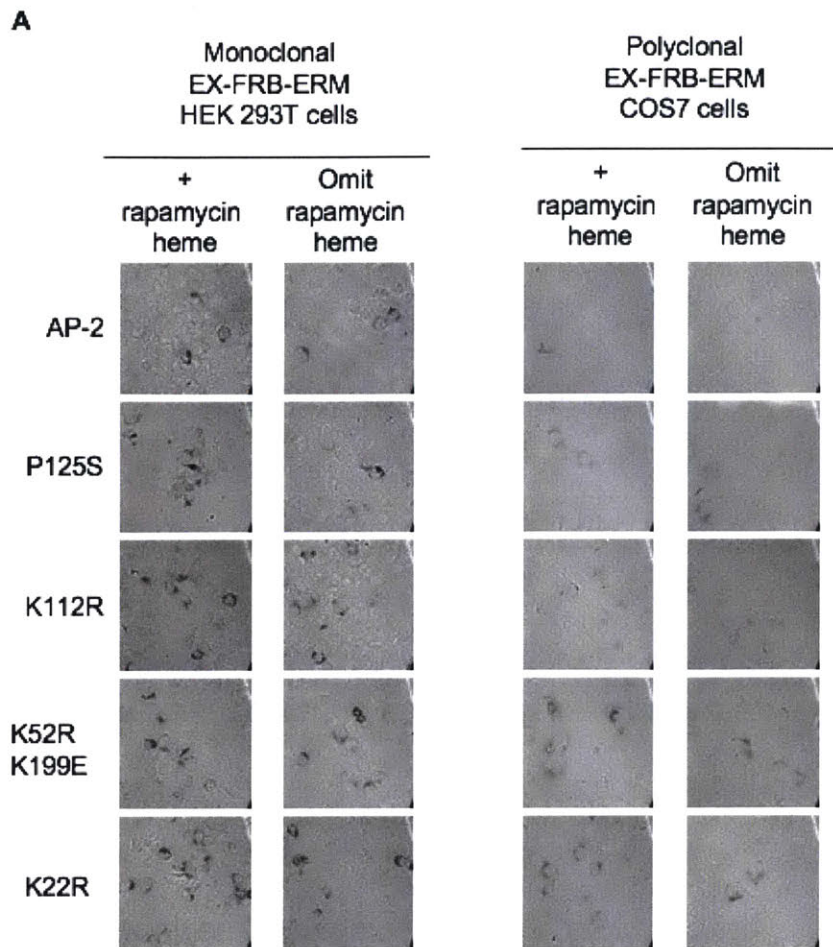


Figure 4-7. Diaminobenzidine (DAB) labeling of AP-2 mutants derived from third generation of yeast display evolution in stable EX-FRB-ERM cells. (A) Monoclonal HEK 293T and polyclonal COS7 cells that stably express EX-HA-FRB-ERM were infected with the complementary mito-targeted AP-2 variants. After 48 hours of expression, cells were treated with or without 400 nM rapamycin and 1 μ M heme for 90 minutes fixation and DAB labeling on ice for 45 minutes. DAB polymerization results in a dark product that was visualized by brightfield imaging at 40x.

We selected single mutant [K22R] as the final mutant, AP-3, which had noticeably higher activity than preceding clones in both DAB and BP labeling assays, but without incorporation of a negative selection, this activity was still not significantly dependent on PPI-induced co-proximated reconstitution. This was apparent in **Figure 4-7** in which even with reduced expression level techniques utilizing infection in monoclonal HEK 293T cells, resulted in significant

background without the addition of rapamycin. The increased PPI dependency in COS 7 cells may be a result of the aforementioned lack of heme but not necessarily the lack of reconstituted protein, i.e. still reconstituted peroxidase but without the heme forms an apoprotein.

The improvement in activity with AP-3 was promising but results suggested a very important need to incorporate a negative selection.

Generation 4 - directed evolution of AP-3

To select for high reconstitution activity but decreased fragment affinity, we combined elements from generation 2 and 3 selection platforms to alternate between positive and negative selections to select for a low affinity variant of the AP fragment such that reconstitution is only dependent on close co-proximity of both fragments (see chapter 3 **Figure 3-7A**). After a total of 8 rounds of selection, the yeast population was significantly depleted in cells exhibiting fluorescence upon addition of EX lacking the BASE-p1 coil for PPI. For the final 3 rounds of selection, zymoprep isolation of plasmid DNA, followed by transformation into E. coli cells, failed to yield many colonies. The few colonies that did grow were of silent mutations and template, despite the final yeast populations behaving drastically different from the original AP-3 template. Furthermore, the later the round of selection, the fewer the colonies that grew for parallel zymopreps. We hypothesized that perhaps the winning clones were not in plasmid DNA but got homologously incorporated into the yeast genome instead. Thus, we performed yeast genome extraction of the final rounds to identify the mutations of the winning clones. Sequence analysis revealed 6 unique clones, with convergence upon two main mutants: single mutant [I185V] and double mutant [P125, E171D].

During generation 4 evolution, monoclonal EX-FRB-ERM in COS7 cells were isolated. Since one major application of interest is at the mito-ER junction, these COS7 cells provided a stringent low heme environment that allowed for easily imaged mito and ER organelles. The round cellular morphology of HEK 293T cells made it difficult to determine if organelles were perturbed by the expression and reconstitution of split fragments. Additionally, the monoclonal nature of these cells made it easy to ensure that the small fragment expression level remains consistent, and any variations observed would be a property of the addition of the large fragment.

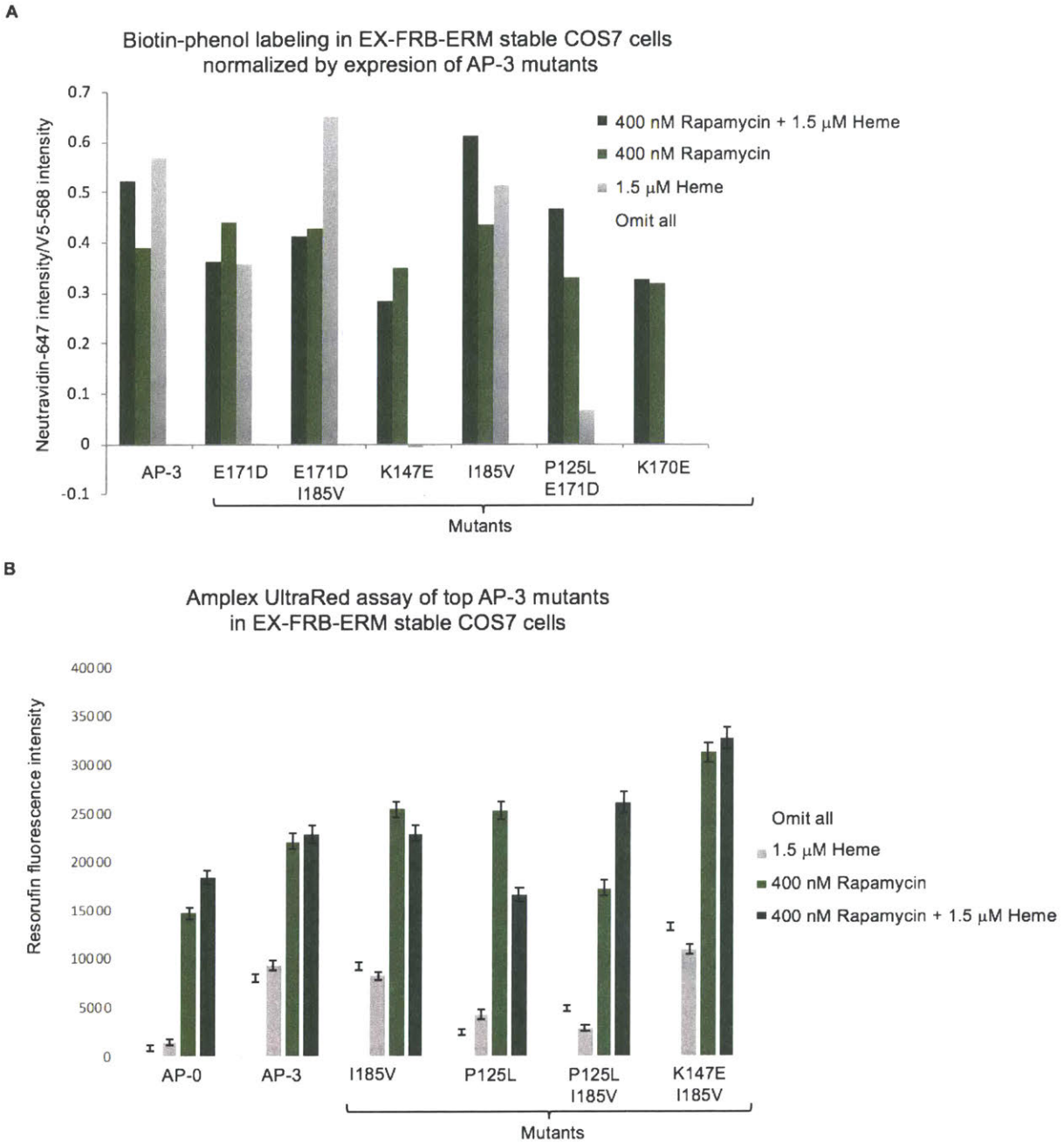


Figure 4-8. Biotin-phenol and Amplex UltraRed labeling of AP-3 mutants derived from fourth generation of yeast display evolution in monoclonal stable EX-FRB-ERM COS7 cells. (A) Monoclonal COS7 293T cells stably expressing the small fragment, EX-FRB-ERM, at a low level were plated in 48-well plates and transfected with the complementary mito-targeted AP-3 variants. After overnight protein expression, cells were treated with or without 400 nM rapamycin, with or without 1.5 μM heme supplementation for 30 minutes, before incubation with biotin-phenol for an additional 30 minutes. Cells were biotin-phenol labeled with standard conditions,

immunostaining and staining with neutravidin-647. Average neutravidin-647 intensity for 10 fields of view are shown for each condition, as normalized by the expression of the AP-3 mutant. Confirmation that constructs expressed at similar level across these fields was done by immunostaining for the epitope tags, V5 and HA. **(B)** Monoclonal COS7 293T cells stably expressing the small fragment, EX-FRB-ERM, at a low level were plated in 48-well plates and transfected with the complementary cytosolic-targeted top AP-3 variants. After overnight protein expression, cells were treated with or without 400 nM rapamycin, with or without 1.5 μ M heme supplementation for 30 minutes, before 30 minute incubation with Amplex UltraRed on ice. Resorufin intensity from 5 different regions of each well was measured by plate reader, standard error shown.

To first test the mutants with a more demanding assay, monoclonal COS7 cells were Biotin-phenol labeling assay was performed with the 6 isolated clones (**Figure 4-8A**). The E171D mutation on its own seemed detrimental for PPI-dependent reconstitution. Common mutations I185V, P125L appeared to be beneficial for enhancing PPI-dependency, without significantly compromising activity. Combinations of mutations that were beneficial and tolerated were made and tested.

While the mito-ER junction is a very appealing application, we wished to test the more promising mutants in a more general setting. The ER-tethering of the less stable small fragment, EX, was very beneficial for maintaining robust levels of EX expression. We continued to utilize these stable monoclonal COS7, and instead transfected cytosolically targeted AP-3 mutants. While this assay is less demanding (the large fragment is spatially more diffuse in the cytosol than when targeted to the mitochondria) it provided a facile way to visualize the localization of both fragments and ensure that the large fragment remains cytosolic and only the addition of rapamycin will recruit the free-floating large fragment to adopt the morphology of EX, which is tethered. Amplex UltraRed was initially utilized to quickly screen activity in this new assay (**Figure 4-8B**).

After multiple comparative assays, it appeared that the quikchange-generated double mutant [P125L, I185V] consistently generated robust PPI-dependent enzymatic activity against the AP-3 and other mutants; it is also the combination of the two most frequently seen mutations from the genome extraction. Ultimately, we named this clone AP, the final sAPEX N-terminal fragment, which contains 9 mutations in total compared to the original APEX2 sequence.

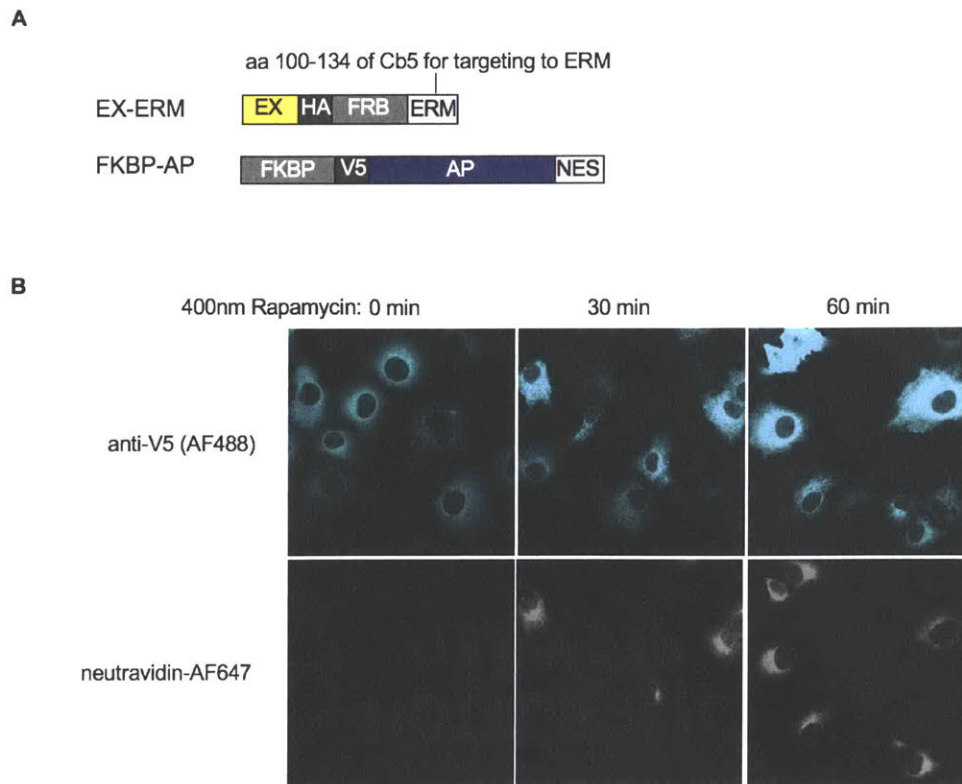


Figure 4-9. Biotin-phenol labeling of finalized split APEX in monoclonal stable EX-FRB-ERM COS7 cells infected with FKBP-AP. (A) Split APEX constructs utilized. EX-ERM is stably expressed in monoclonal COS7 cells. The AP construct is cytosolic: FKBP-AP, in which a NES signal sequence was added during cloning to the C-terminus as compared to previous cytosolic constructs tested. (B) Monoclonal COS7 293T cells stably expressing the small fragment, EX-FRB-ERM were plated in 48-well plates and infected at 70% confluency with FKBP-AP. After 31 hours of protein expression, all cells were treated 1 μ M heme supplementation for 60 minutes, during which 400 nM rapamycin was added for 60, 30, and 0 minutes prior to biotin-phenol incubation for an additional 30 minutes. Cells were biotin-phenol labeled with standard conditions, immunostaining and staining with neutravidin-647 for imaging by confocal microscopy.

We utilized the hybrid ER-cytosolic assay again to ensure that AP demonstrates low affinity properties for tethered EX. COS7 stable EX-ERM cells were infected with FKBP-AP; the expressed AP fragment is cytosolically targeted with a C-terminal NES (nuclear export signal), if low affinity, we would expect to see that the localization of the AP to remain diffuse and cytosolic (0 minutes, **Figure 4-9B**). However, with just a 30-minute incubation of 400 nM Rapamycin, the morphology of AP became more ER-like as it is recruited to the EX fragment, which is tethered.

Overall signal suffered at such low expression levels, and would benefit from optimization of expression, heme levels, rapamycin incubation times – which could all vary from one cellular context to the next.

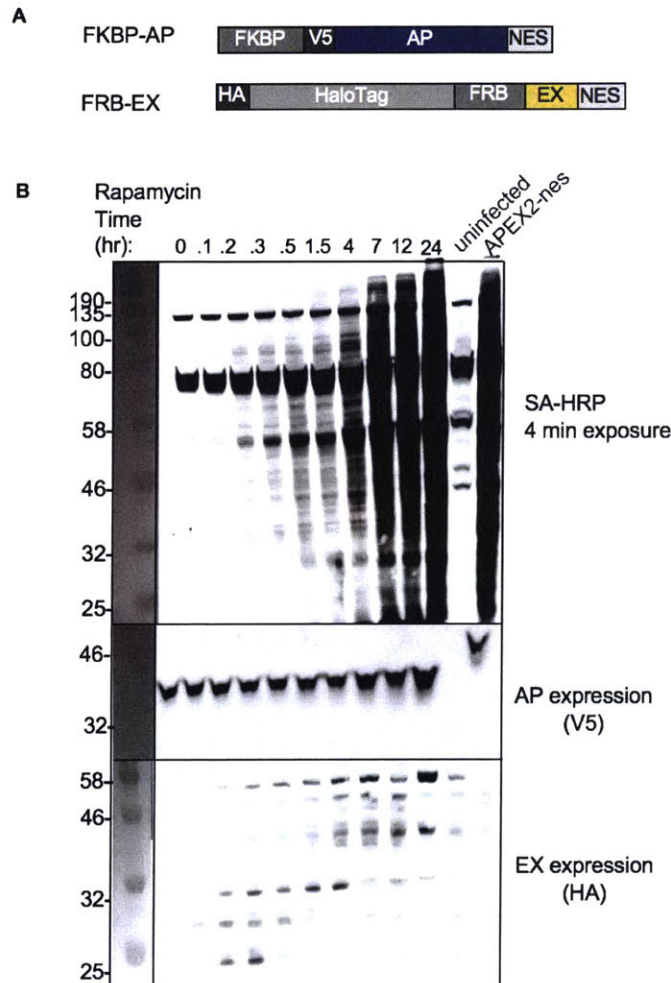


Figure 4-10. Western blot analysis of biotin-phenol labeling of finalized split APEX in stable FRB-EX HEK 293T. (A) Comparison of sAPEX in the mammalian cytosol; stable FRB-EX HEK 293T cells were plated in 6 well plates. FKBP-AP was introduced by lentiviral infection and after 24 hours, virus was removed for all conditions and 400 nM rapamycin was added for the designated time. All conditions underwent a 3 μ M heme supplementation for 90 minutes followed by live biotin-phenol labeling before analysis by western blot. Top blot is analysis of extent of biotinylation with streptavidin blot readout. Anti-V5 and anti-HA blots detect expression of N-terminal and C-terminal fragments, middle and bottom blots, respectively. Full length APEX2 contains a V5 epitope tag.

Preliminary efforts were also made to characterize the extent of reconstituted activity overtime with free-floating, geometrically unconstrained cytosolic constructs (**Figure 4-10a**). HEK 293T cells were infected and selected to generate stable cells that express FRB-EX, which is HA-HaloTagTM-EX-NES. The unstable nature of the EX fragment required much troubleshooting and the fusion to a soluble and stable protein, HaloTag. Without the fusion, infections to express just the unoptimized EX fragment cytosolically resulted in undetectable immunostaining for the HA epitope tag. With the EX construct fused to HaloTag, EX was stabilized enough such that it was detectable by immunostaining, but western blot analysis revealed that it was still a highly unstable construct. In **Figure 4-10b**, western blot analysis of EX expression level by immunostaining for the HA epitope tag, revealed that with decreased rapamycin incubation the free-floating EX fragment is highly prone to degradation (full MW at 58 kDA). **Figure 4-10b** also analyzed the extent of biotinylation using Streptavidin-HRP (SA-HRP) western blot analysis. Unsurprisingly, increasing the rapamycin incubation time increased the extent of biotinylation as detected by the SA-HRP. The 24-hour addition of rapamycin most stabilized the EX fragment, but also the AP fragment. The stoichiometric excess of rapamycin was able to induce PPI as new AP and EX proteins were still being synthesized, such that they were reconstituted and stabilized thus generating the strongest biotinylation signal. Promisingly, immunostaining for the V5 epitope tag revealed that the AP fragment is much more stable and not prone degradation even in the absence of EX fragment.

Comparing finalized split APEX (AP +EX) against earlier generations

After a cumulative 20 rounds of selection using three different yeast display configurations, we characterized final generation clones side-by-side to assess the progress of our evolution. In yeast (Discussion section of Chapter 3), reconstituted peroxidase activity, progressively increases from the template AP-0 to the finalized AP. After the implementation of negative selections in generation 4, PPI-independent peroxidase activity is significantly reduced from AP-3 to AP. Observations on yeast indicate that the selections worked as desired and that our final clone AP combines the features of high reconstitution activity with low proximity-independent reconstitution (i.e., low fragment affinity).

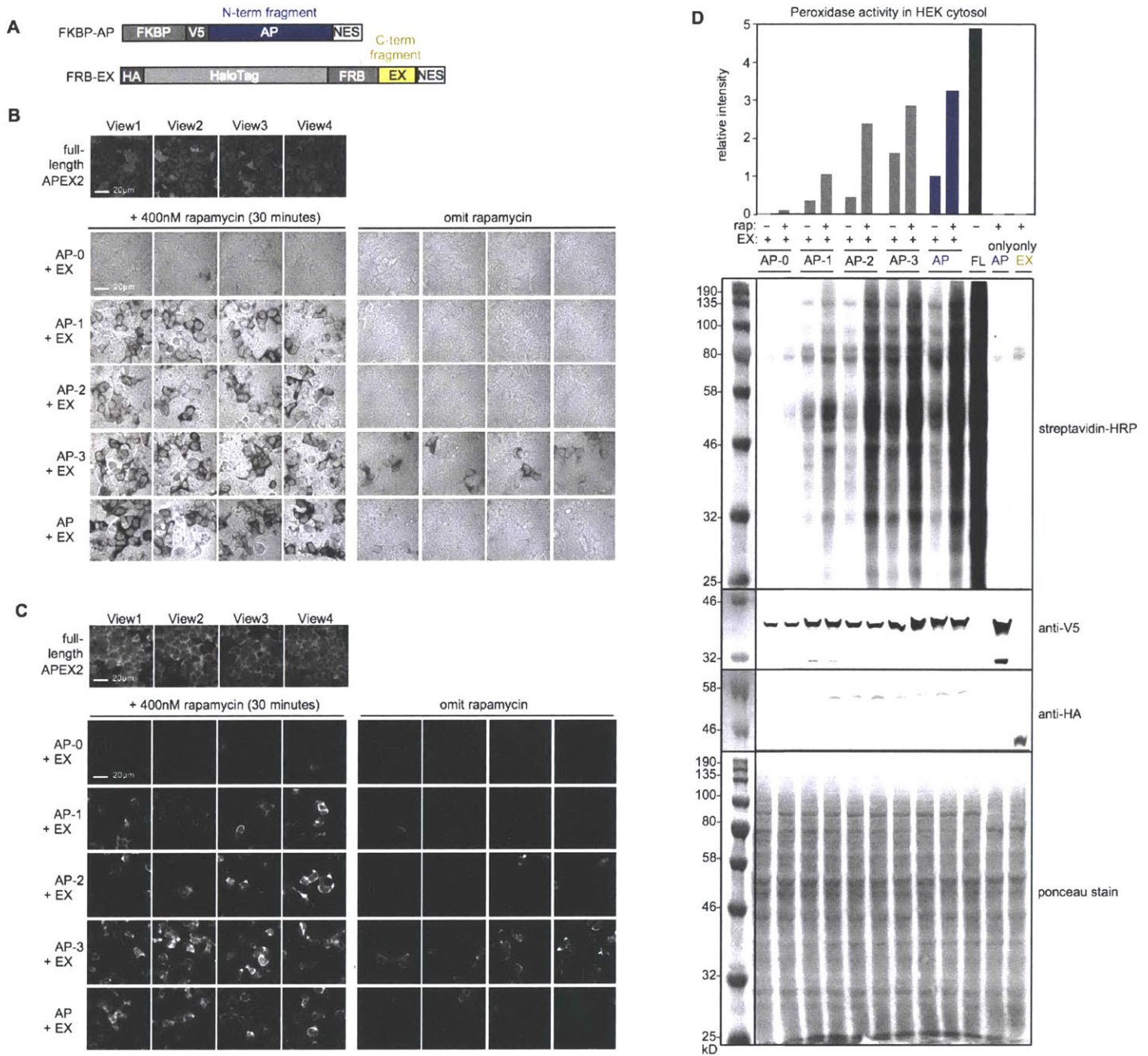


Figure 4-11. Comparing the different generations of evolved split APEX clones in mammalian cells. (A) Depiction of protein sequences for FKBP and FRB fusions to sAPEX for analysis in mammalian cells. While generating the HEK 293T cells with stable expression of EX (used for matched comparisons of the different variants of AP), at the lower protein expression level, we discovered that FRB-EX is rapidly turned over. N-terminal HaloTag fusion improved the protein stability. Conversely upon swapping the sAPEX fragments, FRB-AP does not have the same degradation problems. We postulate that this is due to the instability of the EX fragment, although this can be remedied, this property may also serve as a possible advantage to further

reduce non-specific background. **(B)** Comparison of sAPEX variants in the mammalian cytosol, with DAB (diaminobenzidine) polymerization as the readout of peroxidase activity. The indicated N-terminal variants of sAPEX were introduced by transient transfection into HEK 293T cells stably expressing FRB-EX, which were incubated with rapamycin for 30 minutes (left) or left untreated (right). We utilized HEK 293T cells stably expressing APEX2-NES as a benchmark; but because transfection and lentiviral transduction infection efficiencies are imperfect, it results in a reduced number of comparable HEK 293T cells that would express both fragments. Cells were fixed and incubated with DAB and H₂O₂ for 15 minutes, as previously described²⁵, to allow peroxidase-catalyzed polymerization of DAB. In the bright field images, dark regions indicate peroxidase activity. Four separate fields of view are shown per condition. Scale bar, 20 μ m. Two biological replicates were performed. **(C)** Same assay as in **(B)** except FKBP-AP was introduced by lentiviral infection, and live biotin-phenol labeling was used to detect peroxidase activity. Induced HEK 293T cells were treated with biotin-phenol in the presence of H₂O₂ for 1 minute, then fixed and stained with neutravidin-AlexaFluor647 to visualize peroxidase-catalyzed promiscuous biotinylation¹¹. Scale bar 20 μ m. Four biological replicates performed. **(D)** Same assay as in **(C)** but with streptavidin blot readout. Two biological replicates performed. Quantitation of signal in each lane shown in bar graph at top. Anti-V5 and anti-HA blots detect expression of N-terminal and C-terminal fragments, respectively.

We next tested whether these trends would hold in the mammalian cell cytosol. This environment is quite different from the yeast cell surface, as it is 37 °C instead of 30 °C, and a reducing rather than oxidizing environment. We also wished to test the sAPEX clones as soluble proteins rather than as membrane-anchored constructs with restricted geometry. Hence, we expressed the N- and C-terminal fragments from each stage of directed evolution as fusions to FKBP and FRB, respectively (**Figure 4-11a**), where AP is the large fragment and could be a predecessor clone.

We first compared peroxidase activity with or without rapamycin, using our DAB polymer assay (relevant for EM applications^{2,25}), which is much less sensitive than both Amplex Red and BP assays, thus enabling us to rigorously compare the activity of our sAPEX fragment pairs (**Figure 4-11b**). The original sAPEX template, AP-0, gave barely detectable DAB staining, while AP-1 was dramatically improved, and AP-3 gave the strongest staining of all. However, as observed in yeast, AP-3 also gave significant signal in the absence of rapamycin, suggesting that the fragments have high affinity. In contrast, the final AP + EX clones displayed high DAB staining in the presence of rapamycin, but nearly undetectable activity in the absence of rapamycin. The DAB assay utilized HEK 293T cells stably expressing FRB-EX (**Figure 4-11a**) transfected with equal amounts of DNA to express the different AP variants.

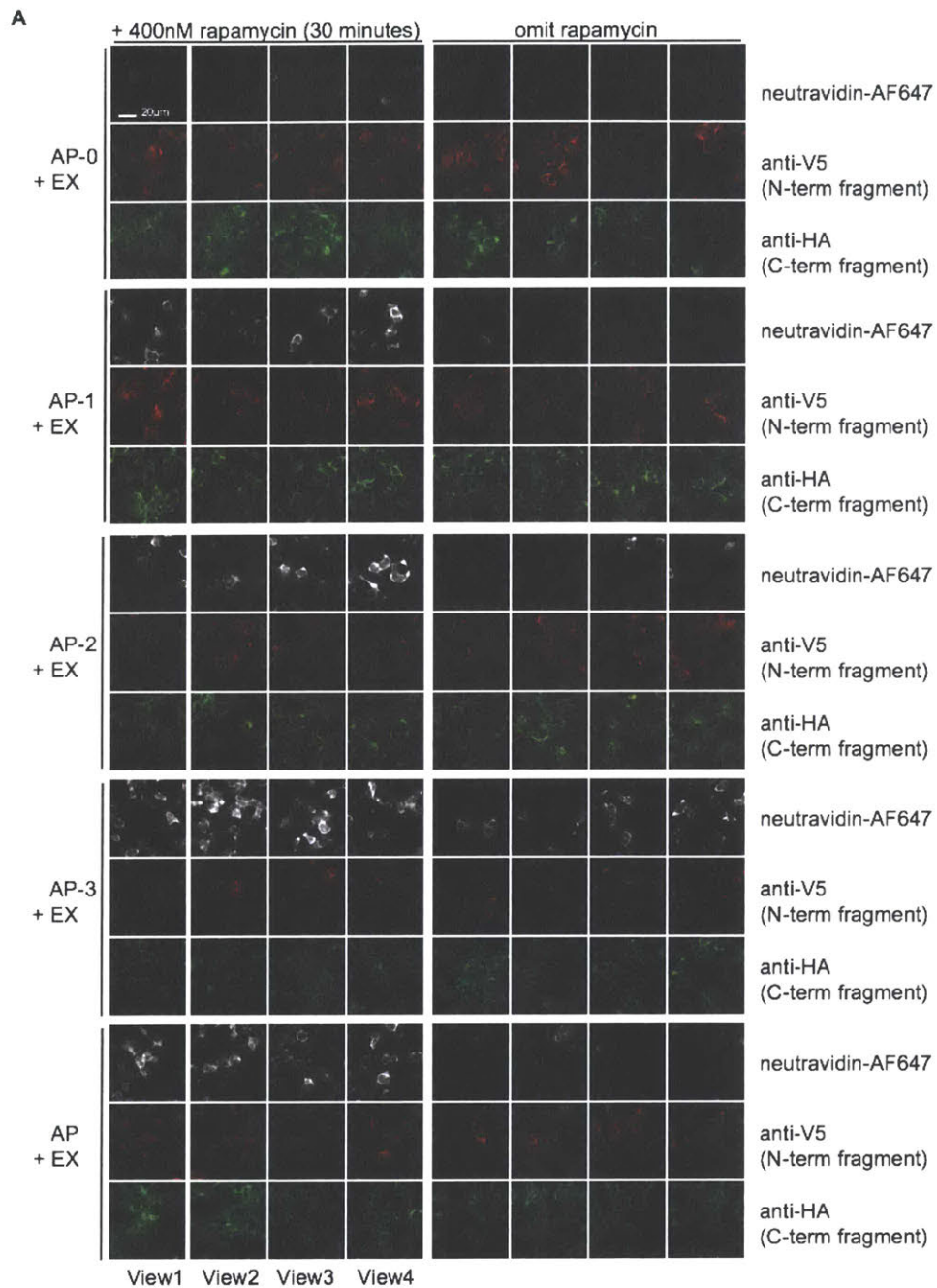


Figure 4-12. Comparison of sAPEX variants in the mammalian cytosol, with biotin-phenol labeling as readout of peroxidase activity. (A) Images are from **Figure 5-11C**, but here anti-V5 and anti-HA channels are shown as well. Anti-V5 staining quantifies expression of the N-terminal fragment, AP, and anti-HA staining quantifies expression of the C-terminal fragment, EX. Scale bar, 20 μ m.

For the more sensitive biotin-phenol labeling assay, the AP variants were introduced via lentiviral infection to reduce background reconstitution from overexpression. The experiments using the biotin-phenol assay—relevant to spatial proteomics^{11,15} and transcriptomics^{52,53}—showed a similar trend (**Figure 4-11C** see **Figure 4-12A** for matched fields of view of expression level of EX and AP variants). While AP-3 showed high activity following rapamycin treatment, it also revealed background labeling in the –rapamycin condition. In contrast, the final sAPEX pair, AP + EX, displayed +rapamycin activity close to that of AP-3, but with minimal –rapamycin activity. Besides confocal microscopy, we also characterized BP labeling by lysing the cells and blotting the cell lysate with streptavidin-HRP (**Figure 4-11D**). The mammalian assays recapitulated the trends observed on the yeast-surface with regards to activity trends. The PPI-dependence of sAPEX makes it a more versatile tool for cellular applications such as those outlined in the introduction.

A

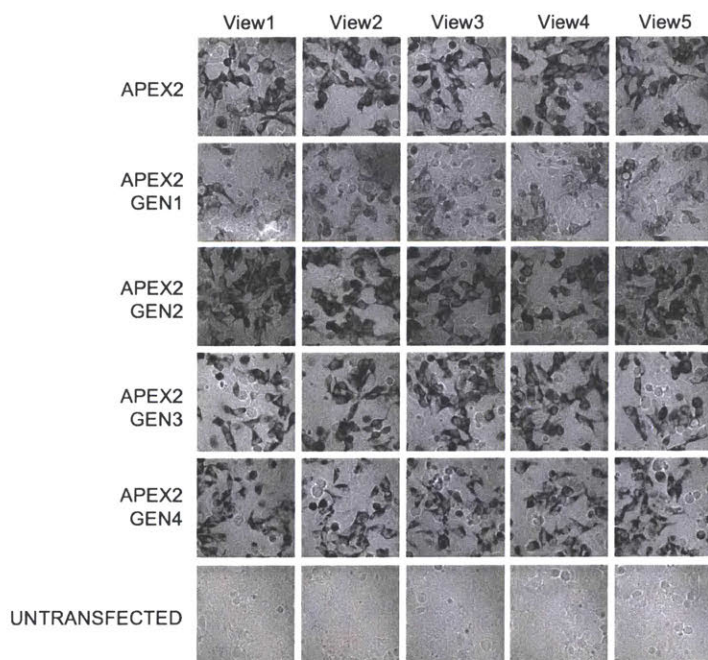


Figure 4-13. Examining accumulated AP mutations in full length APEX2. (A) HEK 293T cells were transfected with variants of full length APEX2 containing a C-terminal NES (nuclear export signal). Mutations accumulated from each generation of yeast-display directed evolution were incorporated and tested. After overnight protein expression, cells were fixed for 30 minutes on ice and DAB labeled on ice for 5 minutes. The resultant DAB polymer is a dark product that is visualized by brightfield microscopy. 5 fields of view shown for each condition

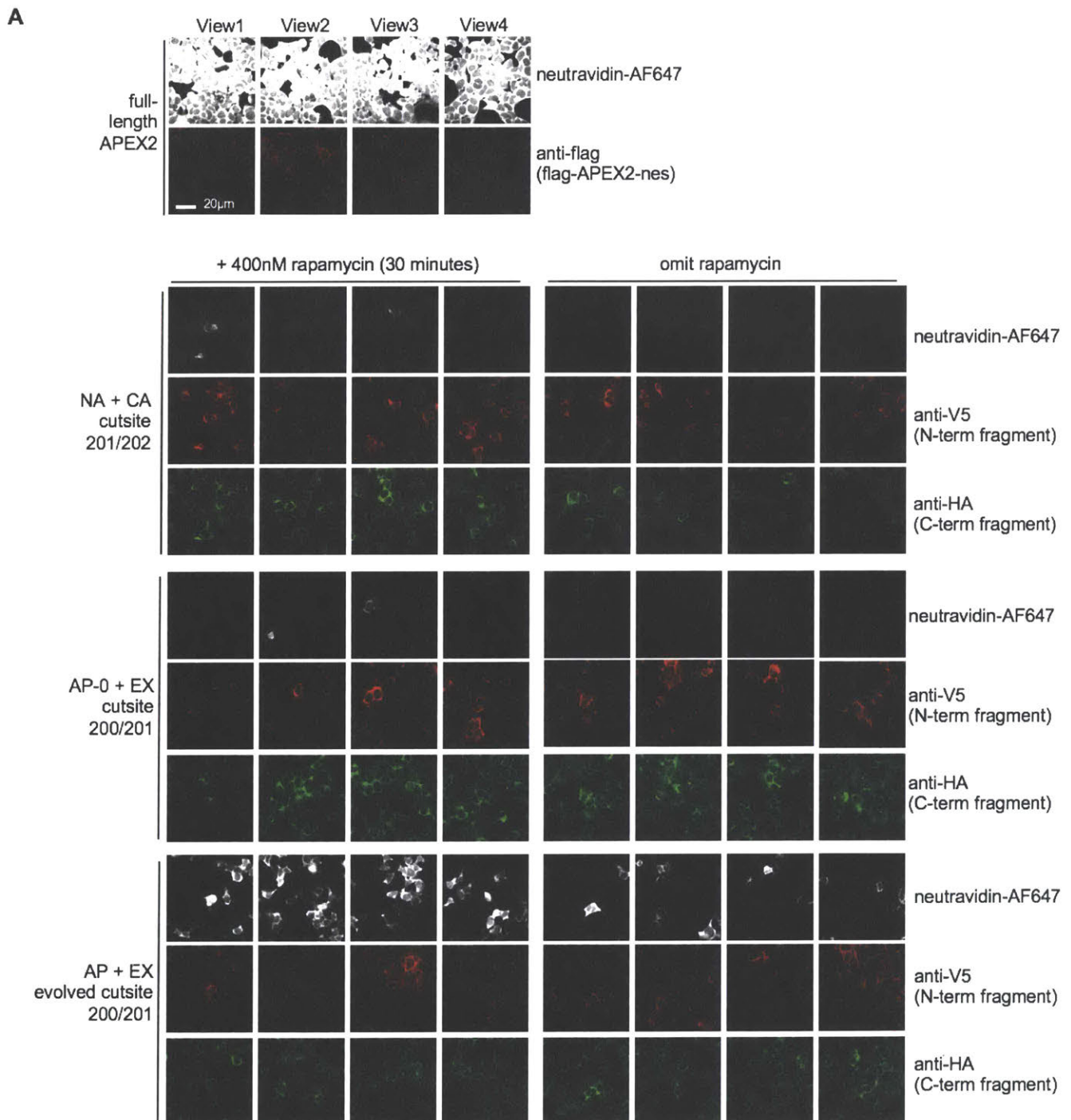
Curious as to what the mutations accumulated from each generation of evolution did to full length APEX2, mutations from each generation were cloned to generate APEX2-NES variants and transfected into HEK293T cells. For example, APEX2 GEN1 has 3 mutations compared to original APEX2; APEX2 GEN3 has 7 mutations compared to original APEX2. After 23 hours of protein expression, HEK 293T cells were fixed for 30 minutes on ice, DAB labeled on ice for five minutes, and imaged by bright field. Interestingly the mutations from generation 1 evolution decreased overall enzymatic activity, whereas the next three mutations accumulated from generation 2 evolution more than recovered the lost activity. The two mutations acquired from generation 4, which included 6 rounds of negative selections, did not significantly decrease the overall enzymatic when incorporated into the full-length APEX2 protein. However, while the DAB labeling pattern of all variants were diffuse and cytosolic in pattern, the APEX2 GEN 4 DAB polymer appeared more aggregate and clumpier in morphology, suggesting that APEX2 GEN4 is more prone to aggregation.

Comparing a published variant of split APEX2¹, our original split APEX2, and evolved split APEX2.

During the development of our finalized split APEX2 (AP + EX), another lab published a split APEX2 tool (NA + CA)¹. This alternative split APEX2 had been fragmented at aa 201/202, just one amino acid from our final selected cut site, aa 200/201. The published group performed rational design screening but did not perform any evolution on either fragment. Utilizing both biotin-phenol and DAB labeling assays, we wished to compare their split tool, NA + CA, with our original starting split APEX2 template (AP-0 + EX) and our evolved finalized split APEX2 (AP + EX).

The published alternative split APEX2, NA + CA, was very similar in activity to our starting template (AP-0, EX) by both biotin-phenol labeling (**Figure 4-14A**) and DAB labeling (**Figure 4-14B**). For the biotin-phenol labeling assay, both fragments of split APEX2 variants were introduced by lentiviral infection. While it does appear that at these expression levels the detectable biotinylation was PPI-dependent, the amount of biotinylation is very low and many cells

that had robust expression of both fragments lacked detectable neutravidin signal. In the same assays, AP-EX had robust biotinylation in the presence of a PPI. Although there is background labeling, previous experiments have demonstrated that reducing fragment expression levels can help reduce background reconstitution resulting from over-expression or just strong expression.



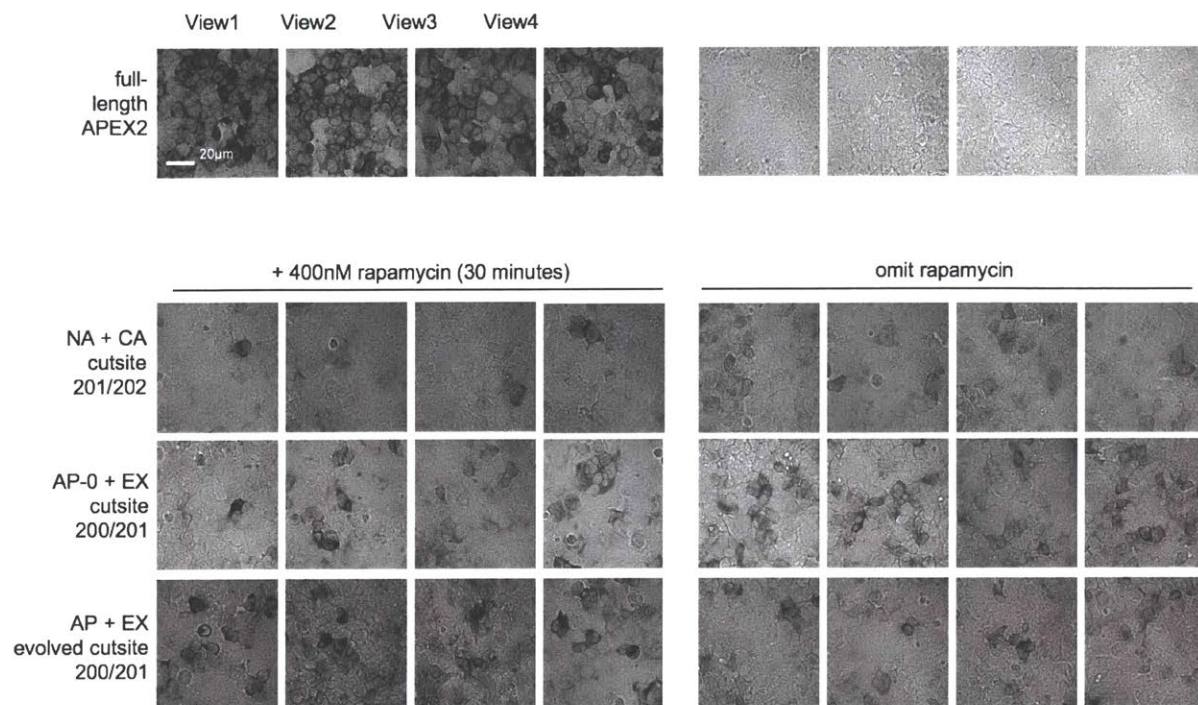
B

Figure 4-14. Comparing a published split APEX2¹, our original split APEX2, and evolved split APEX2. (A) Comparison of 3 different split APEX2 tools in the mammalian cytosol, with extent of biotinylation as the readout of peroxidase activity. See **Figure 4-11A** for depiction of split APEX2 protein fusions. Cut site 201/202 (NA and CA)¹ was tested in tandem with our chosen cut site 200/201 (AP-0 + EX) as well as our final evolved version of AP and EX. Both the indicated N- and C-terminal variants of sAPEX were introduced by lentiviral infection into HEK 293T cells. Full-length APEX2 was infected in parallel as a positive control. After 40 hours of protein production and heme supplementation for 90 minutes at 5 μ M, cells were incubated with biotin-phenol for 30 minutes, in which 400 nM rapamycin was added to drive reconstitution. Cells were then labeled in the presence of H₂O₂ for 1 minute, fixed and stained with neutravidin-AlexaFluor647 to visualize peroxidase-catalyzed promiscuous biotinylation¹¹. Four separate fields of view are shown per condition. Scale bar 20 μ m. (B) HEK 293T cells were transfected to express both fragments of the three variants of split APEX2. After 24 hours of protein expression, cells were treated with heme supplementation for a total of 90 minutes at 5 μ M. During the final 30 minutes of heme incubation, 400 nM rapamycin was added to chemically-induce a protein-protein interaction. Afterwards, HEK 293T cells were fixed for 30 minutes on ice and DAB labeled on ice for 45 minutes. The resultant DAB polymer is a dark product that is visualized by brightfield microscopy. 5 fields of view shown for each condition

For the DAB labeling assay (**Figure 4-14B**), both fragments of split APEX2 variants were introduced by transfection, with a CMV promoter; with these high expression levels and 45 minutes of DAB labeling, there is no discernible PPI-dependent increase in DAB polymer product formation for the unevolved large fragments: AP-0 and NA. Evolved pair, AP + EX, demonstrated both the highest level of PPI-dependent reconstituted enzymatic activity and largest signal to background noise ratio. Furthermore, the background labeling can be reduced by reducing fragment expression levels.

Discussion

Through 4 generations of evolution, 20 rounds of selection, three different yeast display configurations, positive and negative selections, zymopreps and genome extractions, Amplex UltraRed assays, live biotin-phenol labeling assays, and fixed DAB labeling assays, quikchange mutagenesis, stable cell generation, virus generation, we finalized our split APEX2 – sAPEX consisting of the two fragments AP and EX. AP acquired 9 mutations compared to the original template. When comparing the final evolved AP against its predecessors, AP-3, AP-2, AP-1, AP-0, it was clear that there was an improvement in PPI-dependent reconstituted enzymatic activity. Moreover, comparing AP against a previously published split APEX2 variant (NA + CA), our evolved sAPEX (AP + EX) demonstrated both improved enzymatic activity and PPI-dependency, which are properties we deliberately designed our yeast-display platform to select for. These assays really demonstrated the utility of yeast-display directed evolution.

Additionally, through these various assays in both yeast and mammalian cells, the importance of heme bioavailability for the heme-dependent peroxidase was repeatedly emphasized. Efforts were made to select for an AP variant that would express and bind heme robustly independent of the presence of EX. Starting from the second generation of evolution, the large fragment was expressed on the yeast surface without co-expression of EX, which was exogenously added. Heme preincubation, with increasingly lowered heme concentrations were implemented in these selections to search for mutants with increased heme-binding properties. Due to the low bioavailability of free-floating heme in the cell, especially in the cytosol and COS7 cells in general, we found it necessary to still supply a heme incubation in mammalian cells.

Methods

Sequence analysis of individual yeast clones

Plasmid DNA from yeast cultures was harvested using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research) according to the manufacturer's instructions for liquid culture except for the following. 6 μ L Zymolase was added for a total incubation time of 1-3 h, rotating at 37 °C. After the addition of solution 2, the mixture was briefly vortexed. DNA was eluted with 10 μ L of warm water, and all eluted DNA were transformed into XL1Blue E. coli cells by heat-shock transformation and rescued for 1 h at 37 °C. The entire culture was then plated on an LB-ampicillin plate, and 18–36 colonies were picked for sequencing of individual clones.

Lentivirus generation and mammalian cell infection

To generate viruses, HEK 293T cells plated in a T25 flask (Corning) were transfected at ~75-90% confluency with 5 mL of serum-free MEM containing the lentiviral plasmid encoding the gene of interest (2500 ng), the lentiviral packaging plasmids dR8.91 (2250 ng) and pVSV-G (250 ng), and 30 μ L of 1 mg/mL Polyethylenimine "Max" (Polysciences) pH 7.1 for 3-4 h, followed by media change with complete MEM. 48 h after transfection, the cell medium containing lentivirus was harvested, filtered through a 0.45- μ m filter, and added to other fresh cell cultures to induce expression of the gene of interest. Typically, for both HEK 293T and COS7 cultures plated in wells of a 48-well plate (0.95 cm² per well), cells were plated the night before in 1:1 complete growth media. HEK 293T and COS7 cells, at ~70% and ~85% confluency, respectively, were then infected by exchanging the overnight medium with 150-200 μ L of complete MEM and 50-100 μ L of viral medium. Cells were then incubated for 48 h at 37 °C with 5% CO₂ prior to heme incubation, labeling, and imaging.

Generation of stable cell lines

To generate stable cell lines, constructs of interest were cloned into the lentiviral vector pLX304 to make virus, as described above. Low passage HEK 293T cells plated in 6-well plates (Corning) were infected with 800 μ L-1000 μ L of lentivirus for FRB-EX, APEX2-NES, MCP-AP, or MCP-APEX2. After 2 days at 37°C, 5% CO₂, each well was trypsinized and replated into two T25 flasks in full growth media containing 8 mg/ml blasticidin S HCl (Calbiochem, CAS no. 3513_03_9).

Media was changed to fresh 8 mg/ml blasticidin S HCl media every 24 h. Cells were trypsinized and replated before reaching full confluency within the first 7 days. After 7–10 days of selection, surviving cells can be used for experiments. To make cells stocks for long-term storage at -80 °C, cells from a ~90% confluent T75 flask are harvested, pelleted, resuspended 3x concentration in complete media with 5% DMSO, and frozen at a controlled rate to -80 °C overnight using a Mr. Frosty Freezing Container (Thermo Scientific™). COS7 cells stably expressing EX-FRB-ERM were prepared similarly, with a selection in media containing 10 mg/mL blasticidin S HCl.

Cloning

See *Table 4* for a list of genetic constructs used in this study, with annotated epitope tag, promoters, resistance, vector, linkers, etc. For cloning the constructs, PCR fragments were amplified using Q5 polymerase (New England BioLabs (NEB)) or PfuUltra II Fusion HS DNA polymerase (Agilent Technologies). The vectors were double-digested and ligated to gel-purified PCR products by T4 ligation or Gibson assembly. Ligated plasmid products were introduced by heat shock transformation into competent XL1-Blue bacteria. The APEX2 gene used for initial cut site screening was amplified from vimentin-APEX2 with codons optimized for mammalian expression. Mutants of AP were generated either using QuikChange mutagenesis (Stratagene) or isolated from individual yeast clones and transferred to simammalian expression vectors using standard cloning techniques.

Heme supplementation

Where indicated in both yeast and mammalian culture, heme was supplemented by dilution from a thawed 300 μM heme stock solution, which was prepared by dissolving hemin-Cl (Sigma) in 10 mM NaOH with sonication for 3 min, aliquoted, flash frozen in liquid N₂, and stored at -80 °C. Heme stocks were used within 1 h after thawing and never reused. Heme stock is diluted to 5 μM and 6 μM in complete growth media and added to HEK 293T and COS7 cells, respectively, for 90 min prior to bp incubation (without heme) or DAB labeling. For yeast heme supplementation in Generations 2-4, washed yeast cells were pelleted at 7000 × g for 1 min and resuspended to a concentration of 1.5 × 10⁷ yeast cells/mL in .1% BSA in DPBS-B at the designated heme concentrations (Figure S3B, S4B, S5A) just prior to EX incubation.

Live biotin-phenol labeling

Genes were introduced into HEK 293T or COS7 cells through either transient transfection with Lipofectamine 2000 or lentiviral infection. After 18–24 h (transfection) or 48 h (lentivirus), biotin-phenol (bp) labeling was initiated by changing the medium to 200 μ L of fresh growth medium containing 500 μ M bp that was sonicated for at least 5 minutes to ensure bp was fully dissolved. Cells were incubated at 37 °C under 5% CO₂ for 30 min according to previously published protocols. To initiate labeling, 2 μ L of 100 mM H₂O₂ was spiked into each well, for a final concentration of 1 mM H₂O₂, and the plate was immediately gently agitated. To quench the reaction after 1 minute, the bp solution is aspirated, and cells were immediately fixed with 4% formaldehyde with 5 mM Trolox in PBS at room temperature for 5 minutes before continuing fixation on ice for an additional 25 minutes. Cells were then washed with chilled DPBS three times and permeabilized with pre-chilled methanol at –20 °C for 10 min. Cells were washed again three times with DPBS and blocked with 3% BSA in DPBS for 1 h to overnight with rocking at 4 °C. To detect the expression of sAPEX fusions, cells were incubated with primary antibodies mouse anti-V5 (Life Technologies, Cat. No. R96025, 1:1500 dilution) and rabbit anti-HA (Cell Signaling, 1:1000 dilution) in 1% BSA in DPBS for 1 h to overnight at 4 °C followed by 4 \times 5 min washes with chilled DPBS. Cells were then incubated with a 1% BSA in DPBS solution containing secondary Alexa Fluor 488-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11001, 1:1000 dilution), Alexa Fluor 568-goat anti-mouse IgG (Life Technologies, Cat. No. A-11004, 1:1000 dilution), and homemade streptavidin–Alexa Fluor 647 (1:1000) for 25-45min at 4 °C with rocking. Cells were then washed 4 times for 5 min with chilled DPBS and imaged by confocal microscopy.

Fixed Diaminobenzidine (DAB) labeling

Transfected HEK 293T cells plated on glass coverslips were fixed with 4% formaldehyde in PBS for 20 min at room temperature. Cells were subsequently washed twice with room temperature DPBS, incubated for 10 min in buffer containing 20 mM glycine to quench unreacted formaldehyde, and washed 3 more times with DPBS. A fresh solution of 0.5 mg/mL (1.4 mM) DAB (5.4 mg/mL of 3,3'-diaminobenzidine, Sigma, dissolved in 0.1 M HCl) was combined with 0.03% (v/v) (10 mM) H₂O₂ in DPBS for 15 min. To quench polymerization, the DAB solution

was removed and cells were placed on ice and washed four times for 5 min with chilled DPBS, before imaging.

Microscopy

Confocal imaging was performed on a Zeiss AxioObserver inverted confocal microscope with 10× air and 40× oil-immersion objectives, outfitted with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), 561 (DPSS) and 640-nm (diode) lasers (all 50 mW). All images were acquired and processed using Slidebook 5.0 or 6.0 software (Intelligent Imaging Innovations), through a 48× or 63× oil-immersion objective CFP(405 laser excitation; 635/85 emission), Alexa Fluor 488 (491 laser excitation; 528/38 emission), Alexa Fluor 568 (561 laser excitation; 617/73 emission), Resorufin (491 laser excitation; 550/585), Alexa Fluor 647 (647 excitation; 680/30 emission), and differential interference contrast (DIC). Acquisition times ranged from 100 to 1000 ms. DAB labeled cells were imaged by bright field; acquisition time was 200 ms. Imaging conditions and intensity scales were matched for each data set presented together.

Streptavidin blotting

For streptavidin blotting, cells were grown on polystyrene 6-well plates (Corning) and labeled under the same conditions described above (see “Biotin-phenol labeling”). After 1 min of labeling, the cells were washed three times with room temperature quencher solution (10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in DPBS) and then scraped and pelleted by centrifugation at 3000g for 10 min. The cell pellet was then lysed by gentle resuspension into peroxidase-quencher containing RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1× protease cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF (phenylmethylsulfonyl fluoride), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox). After 2 minutes on ice, the lysates were clarified by centrifugation at 15000 × g for 10 min at 4 °C before separation on 10% SDS-PAGE gel. For blotting analysis, gels were transferred to nitrocellulose membrane, stained by Ponceau S (10 min in 0.1% (w/v) Ponceau S in 5% acetic acid/water), and blocked with “blocking buffer” (3% (w/v) BSA and 0.1% Tween-20 in Tris-buffered saline) for 1 h at room temp or 4 °C overnight. The blots were immersed in

streptavidin-HRP in blocking buffer (Thermo Scientific, cat. no. 21126, 1:3000 dilution) at room temperature for 1 h and then rinsed with blocking buffer 4 times for 5 min. For assessing comparative fragment expression level, identical gels and blots were prepared in parallel and immersed in blocking buffer containing either mouse anti-V5 (Life Technologies, R96025, 1:3000 dilution) or mouse anti-HA (Santa Cruz Biotechnology, G1817, 1:3000 dilution) for 1 h to overnight at 4°C. The blots were rinsed with blocking buffer 4 times for 5 minutes before being immersed in anti-Mouse-HRP (Bio-Rad, 1:3000) in blocking buffer at room temp for 1 h, rinsed with blocking buffer 4 times for 5 min again. All blots were developed using Clarity reagent (Bio-Rad) for 5 min and imaged using an Alpha Innotech gel imaging system.

Plasmid name	plasmid vector	promoter	Expression in	Features
p93	PLX304	CMV	Lentivirus	EX-HA-FRB
p94	pcDNA3.1	CMV	mammalian	FKBP-V5-AP-0 and AP-1 variants
p95	pcDNA3.1	CMV	mammalian	OMM-FKBP-V5-all AP viarants
p96	pcDNA3.1	CMV	mammalian	EX-HA-FRB-ERM
p97	pcDNA3.1	CMV	mammalian	EX-HA-FRM-NES
p98	pcDNA3.1	CMV	mammalian	FKBP-V5-AP-2 and AP-3 variants-nes
p99	PLX304	CMV	Lentivirus	HA-HaloTag-10aa linker-FRB-12aa linker-EX-NES
P100	pcDNA3.1	CMV	mammalian	FKBP-10aa linker-V5-12aa linker-AP-0-NES
P101	pcDNA3.1	CMV	mammalian	FKBP-10aa linker-V5-12aa linker-AP-1-NES
p102	pcDNA3.1	CMV	mammalian	FKBP-10aa linker-V5-12aa linker-AP-2-NES
p103	pcDNA3.1	CMV	mammalian	FKBP-10aa linker-V5-12aa linker-AP-3-NES
p104	pcDNA3.1	CMV	mammalian	FKBP-10aa linker-V5-12aa linker-AP-NES
p105	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-AP-0-NES
p106	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-AP-1-NES
p107	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-AP-2-NES
p108	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-AP-3-NES
p109	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-AP-NES
p110	PLX304	CMV	lentivirus	APEX-NES
p111	PLX304	CMV	lentivirus	APEX2-gen1 variant-NES
p112	PLX304	CMV	lentivirus	APEX2-gen2variant-NES
p113	PLX304	CMV	lentivirus	APEX2-gen3variant-NES
p114	PLX304	CMV	lentivirus	APEX2-gen4variant-NES
p115	PLX304	CMV	lentivirus	HA-HaloTag-10aa linker-FRB-12aa linker-CA-NES
p116	PLX304	CMV	lentivirus	FKBP-10aa linker-V5-12aa linker-NA-NES

References

1. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).
2. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* (2011). doi:10.1073/pnas.1101046108
3. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
4. O’Shea, E. K., Lumb, K. J. & Kim, P. S. Peptide ‘Velcro’: Design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658–667 (1993).
5. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* (80-.). **339**, 1328–1331 (2013).
6. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).
7. Hung, V. *et al.* Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol. Cell* **55**, 332–341 (2014).
8. Garzia, A., Morozov, P., Sajek, M., Meyer, C. & Tuschl, T. PAR-CLIP for discovering target sites of RNA-binding proteins. in *Methods in Molecular Biology* **1720**, 55–75 (2018).
9. Sibley, C. R. Individual nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) to determine protein–RNA interactions. in *Methods in Molecular Biology* **1649**, 427–454 (2018).
10. Xue, M. *et al.* Optimizing the fragment complementation of APEX2 for detection of specific protein-protein interactions in live cells. *Sci. Rep.* **7**, (2017).

Chapter 5. Applying sAPEX to the mito-ER junction for EM imaging and proteomic mapping

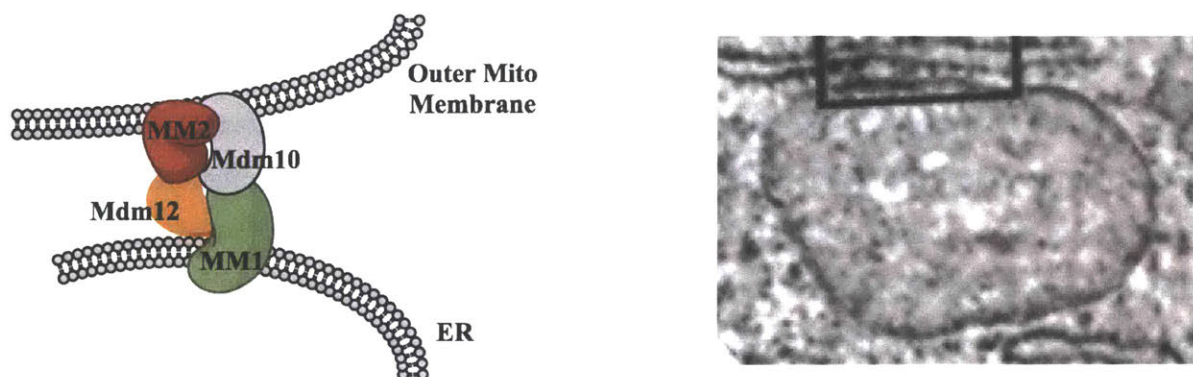
Introduction to the mito-ER junction

In general, the use of APEX tool begins with fusing it to a protein or peptide in order to target it to a subcellular region or macromolecular complex of interest. For instance, the mitochondrial matrix¹¹ and intermembrane space (IMS)¹⁵ proteomes were mapped using first-generation APEX, which lacks the A134P mutation that was later found to boost activity. The high specificity of the matrix proteome can be attributed to the region being inherently membrane bound and the membrane-impermeable nature of the biotin phenoxyl radical. The IMS is the space between the inner and outer mitochondrial membrane. Due to the porous nature of the outer mitochondrial membrane, a ratiometric statistic was used to increase specificity. This approach compared the extent of biotinylation of proteins by IMS-targeted APEX to cytosolically-targeted APEX and distinguished the proteins that are in the IMS from those proximal to the IMS¹⁵. Furthermore, we have targeted APEX2 to the outer mitochondrial membrane (OMM) and the endoplasmic reticulum membrane (ER membrane, or ERM) of mammalian cells by fusing the APEX2 gene to the C-terminal 31 amino acids of the native OMM mitochondrial antiviral signaling protein (MAVS), or to the N-terminal 29 amino acids of the native ERM protein P450 oxidase 2C1, respectively^{16,25}. These constructs were used for both EM²⁵ and proteomic analysis¹⁶ of the OMM and ERM environment.

While this strategy has provided access to many cellular regions and organelles, there are numerous compartments and structures that cannot be cleanly accessed by such an approach. For example, there is great interest in the biology of organelle-organelle contact sites, such as the junctions between mitochondria and ER, which participate in calcium signaling^{36,37}, lipid synthesis³⁸⁻⁴¹, and mitochondrial fission^{42,43}. Yet all APEX2 fusion constructs we have evaluated or considered, such as to the proteins Drp1⁴², Mfn2⁴⁵⁻⁴⁷, SYNJ2BP1¹⁶, and PDZD8⁴⁸ would also target the peroxidase to protein pools outside of mito-ER contacts, such as into the cytosol⁴⁹, along the cytoskeleton⁵⁰, and over the entire OMM¹⁶. For instance, Mitofusin 2 (Mfn2) is an outer-membrane mitochondrial protein essential for mitochondrial fusion with a small population of Mfn2 on the ER membrane, bridging the mitochondria and the ER⁴⁵, but Mfn2 knock out fibroblasts only showed reduced mito-ER contact, indicating Mfn2 is not a unique tether⁸⁶. Furthermore, this claim is controversial, as other Mfn2 knock out studies actually observed an increase in mito-ER contacts^{46,47}. Furthermore, proteins such as dynamin-related protein (Drp1)

and mitochondrial fission factor (Mff), are known to transiently localize to mito-ER contact sites during processes such as mitochondrial fission, but even during these events, significant pools of these proteins exist elsewhere within the cell⁸⁷. Hence, APEX2 fusions to these proteins would result in tagging of many irrelevant proteins.

In mammalian cells, an estimated 5-20% of the mitochondrial outer membrane makes intimate contact (<70 nm gap) with the membrane of the endoplasmic reticulum (ER)³⁷. These mito-ER contacts are thought to be important for a variety of functions and signaling processes^{36,37,38-41,42,43}. Recent work has identified proteins that reside at mito-ER contacts and may play a role in tethering the membranes together^{16,48}.



The ER-mitochondria encounter structure (ERMES) complex is known in yeast³⁹ (**Figure 5-1**), but the identity and function of the proteins that induce and are recruited to such contact sites are still unknown in mammalian cells⁸⁶. ERMES serves as a tether between the two organelles and is comprised of four proteins that do not have metazoan homologs⁸⁸. A major goal of this field is to comprehensively characterize the molecular composition of these contact sites within mammalian cells to better understand how they mediate important cellular processes. The current protocol for isolating mito-ER junctions is the mitochondria-associated membrane (MAM) prep⁸⁹, which purifies this region by first isolating a crude mitochondria sample, and then further fractionating this sample to pure mitochondria and mitochondria-associated membrane sample. However, as this method is performed on cell lysates, spatiotemporal information is lost.

Additionally, because the region of interest is not membrane-bound, loss of sample and insufficient fractionation results in poor coverage and false positives, respectively.

Given the limitations of the MAM prep for identifying mito-ER resident proteins and tethers, with no tethering protein has been unequivocally established in mammalian cells, and to advance the utility of the APEX2 approach to obtain proteomes of regions that are not membrane bound and cannot be targeted by single protein fusions an alternative strategy must be developed. Thus, we utilized a combination of rational design and yeast surface display directed evolution to develop a split APEX that could be utilized to map mito-ER contact sites in mammalian cells.

Applying sYFP and sGFP at the mito ER junction

While evolving sAPEX, efforts were made to test other protein complementation assays (PCAs) at the mito-ER junction to better understand the behaviors and limitations of these kinds of assays at these organelle-organelle contact sites. How important is low do these fragment affinities need to be? How much perturbation can occur upon reconstitution? We chose split GFP and split YFP because of their differential fragment affinities for one another.

Evaluating high affinity split GFP at the mito-ER junction

The split GFP PCA tool chosen is from the Joshua Sanes lab, *Harvard University*⁹⁰. The Franck Polleux lab, *Columbia University*, targeted the two fragments to the mito-ER and claimed to observe GFP reconstitute across the mito-ER junction without resulting in morphological perturbations: We obtained the following mito-ER sGFP constructs from the Polleux lab:

1. CAG promoter_CanX-10 amino acid linker- sGFP11 (ER targeted)
2. CAG promoter_sGFP1-10 amino acid linker- 10-ActA (mito targeted)

Initial testing was inconclusive in our hands; co-transfection of both constructs into HeLa Cekk did generate robust GFP signal and gave a pattern that resembled the mitochondria morphology. However, there were also many examples of overexpression and potential resultant artifacts like blobs, aggregates, etc. At lower expression levels, the GFP signal did occasional demonstrate a more punctate pattern. Transfecting the Polleux constructs into U2OS and Cos7 cells resulted in similar observations.

Ultimately, without any epitope tags, no conclusions could be made about the effect of sGFP targeted across the mito-ER junction on the morphology of the two organelles. The Polleux lab utilized mito trackers and separate fluorescent organelle marker that would not necessarily also

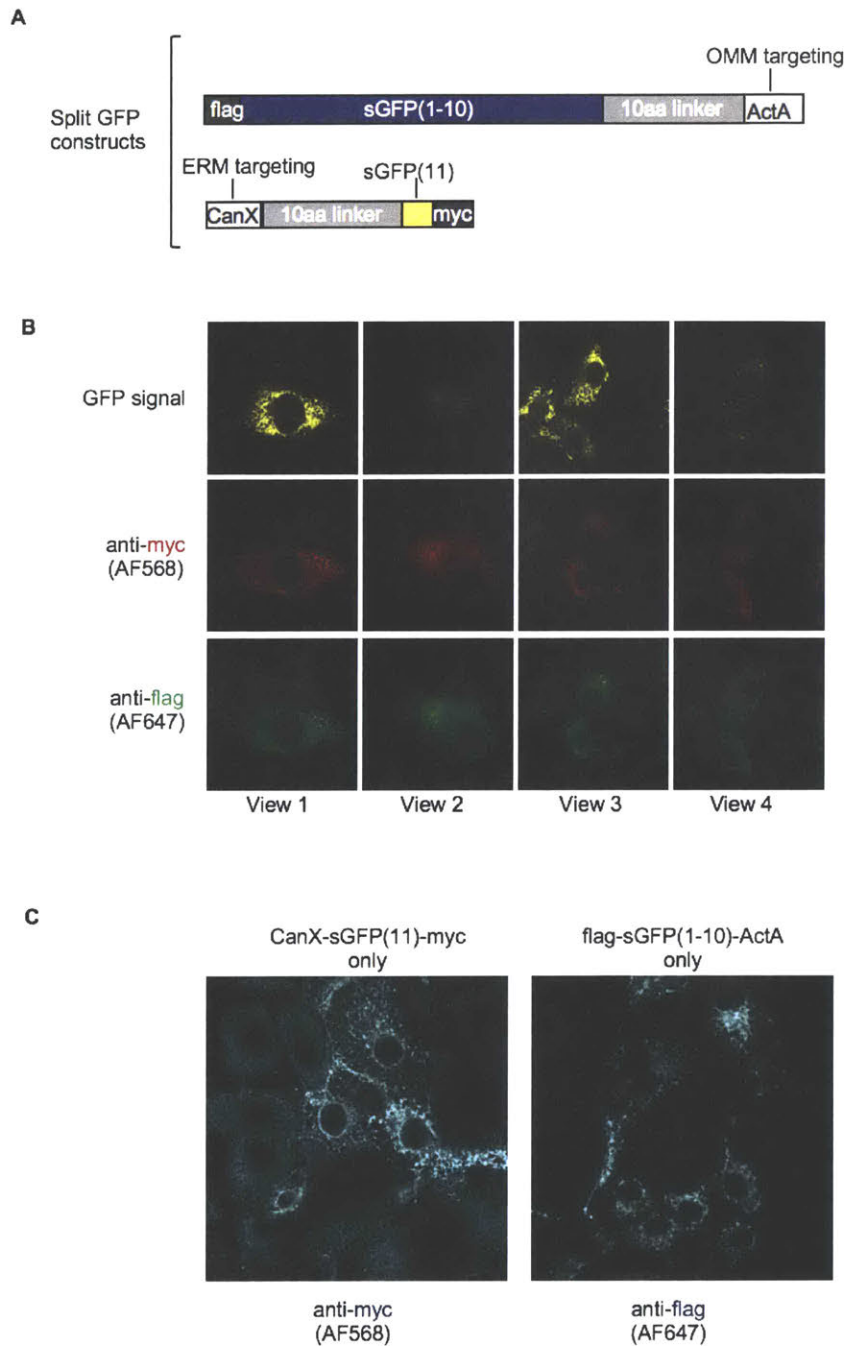


Figure 5-2. Testing split GFP at the mito-ER junction. (A) sGFP constructs. The highly asymmetrically fragmented split GFP was targeted to mito and ER. sGFP(1-10) had a N-terminal flag epitope tag and a C-terminal OMM targeting sequence derived Actin. sGFP(11) has a N-

Terminal ERM targeting sequence derived from Calnexin. **(B)** Cos7 cells were infected by lentivirus to express the constructs **(A)** for 48 hours at 37°C prior to fixation, immunostaining, and imaging by confocal microscopy at 40x. 4 fields of view are shown. **(C)** single sGFP fragment control. Cos7 cells were infected to express either sGFP(1-10) or sGFP(11). No GFP signal was detected. After 48 hours of protein synthesis, cells were fixed and sGFP protein localization was confirmed via immunostaining for the epitope tags and imaging by confocal microscopy.

coincide with the localization of the expressed construct. These constructs were recloned into a viral vector to include epitope tags (**Figure 5-2A**). The viral vector allows for lentivirus generation to lower the expression of these high affinity split proteins via plasmid induction instead of transfection and provides an epitope tag to visualize the localization.

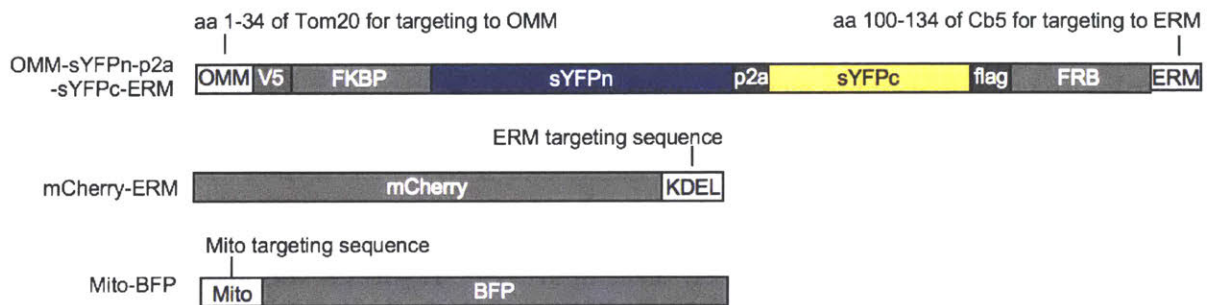
When COS7 cells were infected to express a single sGFP fragment, no reconstituted GFP signal was observed; immunostaining for the fragments revealed morphology that was reflective of the intended target (**Figure 5-2C**). When infecting both constructs into COS7 cells, reconstituted GFP signal was detected in cells with robust expression of both fragments. However, from many fields of view, it often appeared that one organelle did indeed adopt the morphology of the other, or at least the immunostaining for the epitope tags fused to the sGFP fragments have. For instance, closely examining the cells with reconstituted sGFP signal from fields of view 1 through 4 revealed that the cells with signal also had perturbed the localization of sGFP(1-10)-ActA. Imaging revealed that those cells had both more robust expression but also morphology more similar to that of the OMM (**Figure 5-2B**). This alteration in localization could be an artifact of the high affinity between the two fragments as well as the accuracy and robustness of the targeting sequence. Ultimately for the mito-ER application, a high affinity split APEX2 reporter would not be useful. This is further motivation to generate a low affinity split APEX2.

Evaluating lower affinity split YFP at the mito-ER junction

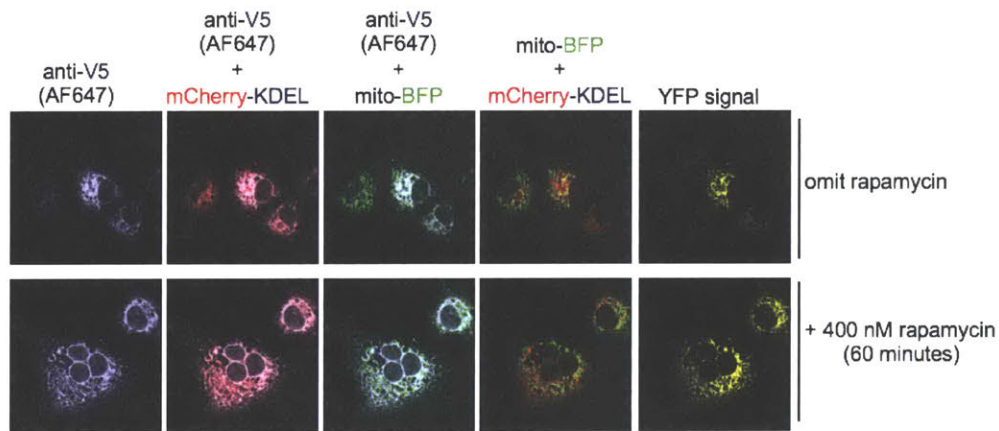
Another PCA we tested was the lower affinity sYFP. The less asymmetrical fragmentation was designed such that the reconstitution is more co-proximity dependent. However, it is important to note that the sYFP PCA requires the protein translation to occur at 30°C, which decreases the translation of proteins, causing expression levels to be even lower. Parallel experiments at 37°C did not result in any detectable YFP signal. With multiple attempts in altering

transfection and translation conditions, and permutations in which one of the organelle markers was eliminated to better understand the effects of these split YFP constructs, the reconstituted YFP signal remained very weak; expression level as determined by immunostaining was also extremely weak. Any detectable YFP fluorescence corresponded to areas of high expression.

A



B



C

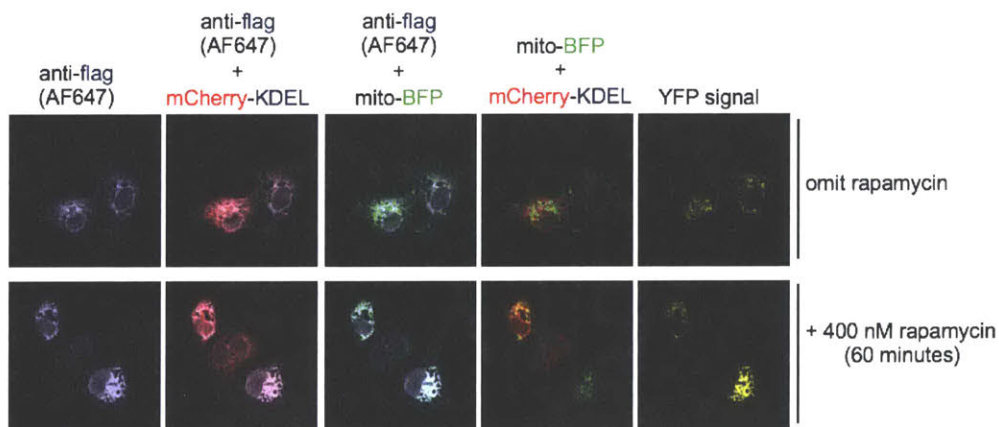


Figure 5-3. Testing split YFP at the mito-ER junction. (A) sYFP p2a construct. The highly asymmetrically fragmented split GFP was targeted to mito and ER. Mito and ER targeting sequence on the N and C-termini are the same utilized for final mito-ER applications. The p2a sequence links both constructs into one plasmid, thus reducing variability between transfection levels from cell to cell. The lower affinity sYFP fragments were also fused to the FKBP/FRB dimerization system. mCherry-KDEL and mito-BFP were utilized as fluorescent organelle markers for the ER and mito, respectively. (B) Cos7 cells were transfected to express the three constructs in (A). After media change to remove transfection media, cells were moved to 30°C and protein synthesis occurred overnight. Cells were then treated with or without 400 nM of rapamycin for one hour prior to fixation, immunostaining for the V5 epitope tag, and imaging by confocal microscopy at 40x. The resultant localization of OMM targeted sYFPn construct was compared to the morphology of the mito and the ER via overlays with mito-BFP and mCherry-KDEL. (C) Identical to (B) and prepared in parallel, immunostaining was also performed for the flag epitope tag.

One of the major issues imaging these split YFP fragments was the low transfection efficiency of COS7 cells. Additionally, we wished to visualize the morphology of the mito and ER organelles, and attempts to introduce fluorescent protein with targeting sequence would then require searching and imaging for COS7 cells that were transfected to express all 4 constructs. As these cells are rare, any imaging data would not necessarily be statistically relevant and often times with such varying expression levels of each construct between cell to cell, it was incredibly difficult to draw major conclusions. One method of alleviating this problem was fusing the two sYFP constructs into one plasmid with a 3x P2A sequence, which results in a cleavage between two constructs during translation (**Figure 5-3A**). This allowed for the transfection of 1 construct that would yield two desired proteins. Furthermore, we envisioned that one major benefit of a p2a construct was that upon introduction into a COS7 cell, both constructs would be ideally expressed and cleaved to generate a 1:1 ratio of both sYFP fragments.

COS7 cells were transfected with 3 constructs (**Figure 5-3A**): OMM-sYFPn-p2a-sYFPc-ERM, mito-BFP organelle marker, and mCherry-KDEL (ER) marker. Our confocal imaging experiments are limited to 4 colors. Two colors – BFP, mCherry were reserved for showing the organelle morphology. The reconstitution of sYFP is the YFP/GFP channel. Ultimately, only one color remains for immunostaining the p2a construct. Thus, two experiments were plated in parallel to immunostain for the V5 and flag epitope tag separately (**Figure 5-3B; 5-3C**). In both experiments, the effects of induced fragment co-proximity was studied utilizing the addition of

400 nM rapamycin addition. Reproducibly, and testing decreasing protein expression time, reconstituted sYFP fluorescence was not significantly rapamycin dependent.

Imaging results showed that upon immunostaining, the OMM-sYFPn-p2a-sYFPc-ERM did not fully cleave. While both fragments expressed and were visualized by immunostaining, the localization of the OMM fragment often appeared to be ER-like and did not overlay with the mito marker, mito-BFP. Additionally, immunostaining of an alternatively cloned p2a construct with the FKBP and FRB domains swapped resulted in no detectable expression of the V5 epitope tag located on the ER targeted construct. These results further motivated us to generate a low affinity split APEX2.

Optimizing sAPEX (AP + EX) targeting and expression levels for mito-ER applications

In addition to evolving an improved split APEX2 pair, we worked on optimizing other variables such as optimizing targeting sequences, heme addition, rapamycin addition, orientation and geometry of split APEX2 fragments, expression levels, and so forth. Major effort was made towards finding the optimal mito and ER targeting sequences. The most common ER targeting sequences used within the Ting lab is the N-terminal 29 amino acids of the native ERM protein P450 oxidase 2C1, C(1-29). The C-terminal 4 amino acid KDEL retention sequence was not tested as it also is known to localize in the golgi (cite). Two common mitochondria targeting sequences are the C-terminal 31 amino acids of the native OMM mitochondrial antiviral signaling protein (MAVSTM) and the AKAP mitochondrial targeting motif, which is localized between residues 49 and 63 of DAKAP1 (variant 1a and 1c)⁹¹.

To test the fidelity of these targeting sequences, FRB and FKBP fused to these targeting sequences were cloned (see **Figure 5-4A; 5-4C** for examples of tested constructs). Initially, mito and ER split YFP constructs were cloned to test the sYFP PCA in tandem. In **Figure 5-4B**, COS7 cells were transfected with constructs from **Figure 5-4A**; these constructs yielded very low expression by immunostaining. The AKAP construct had incredibly low expression and both displayed perturbed morphologies in the absence of rapamycin. It was highly difficult to find cells that had expression of both constructs. Furthermore of the few fields of view that contained both mito and ER fusion proteins, even without the presence of rapamycin, which chemically induces

a PPI, the morphology of the C(1-29) was altered significantly. Much if it had the morphology similar to the mitochondria (**Figure 5-4B**). This is a result of the split fragment causing aggregation.

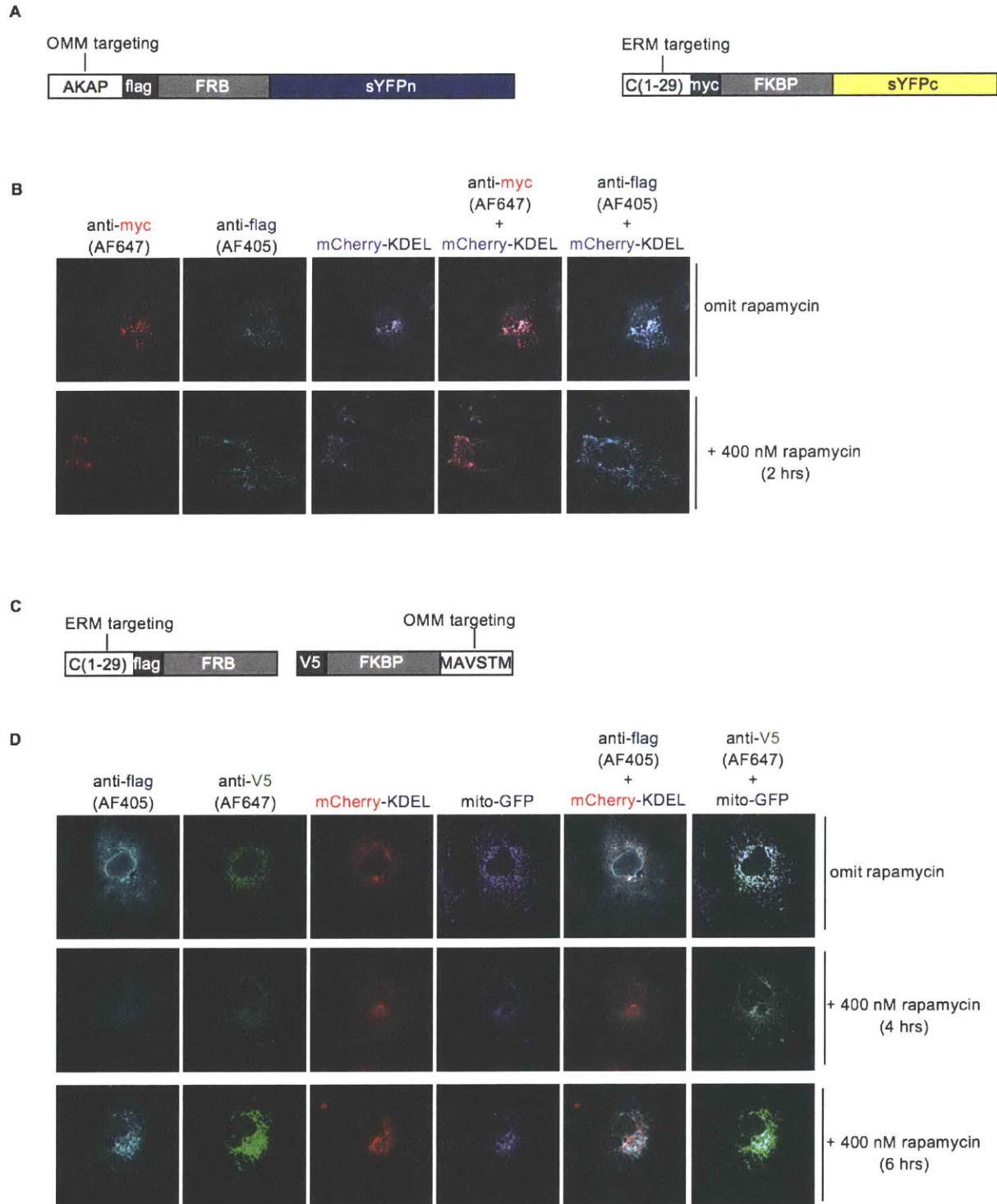


Figure 5-4. Testing mito and ER targeting sequences fused to the FKBP/FRB scaffold. (A) Testing OMM targeting with an N-terminal AKAP mitochondrial targeting sequence and ERM targeting with N-terminal C(1-29) ERM targeting sequence. **(B)** Cos7 cells were transfected to express the constructs from (A) with mCherry-KDEL as a ER organelle marker. After 24 hours of protein expression, cells were treated with 2 hours of rapamycin to evaluate of the effect of a protein-protein interaction on the localization of the FKBP and FRB fusions. **(C)** Testing OMM targeting with C-terminal MAVSTM and ERM targeting with N-terminal C(1-29) ERM targeting sequence in a more simplified scaffold. Eliminating the sYFP fragments allows for a four color experiment. **(D)** Cos7 cells were co- transfected to express the constructs from (C), mito-GFP, and mCherry-KDEL which were utilized as organelle markers. After 24 hours of protein expression, cells were treated with 0, 4, 6 hours of rapamycin to evaluate of the effect of a protein-protein interaction on the mito and - ER localization of the FKBP and FRB fusions, respectively.

However, both constructs did correctly localize when overlaid with mito and ER markers. Upon extended period of rapamycin incubation (21 hrs), mito and ER morphologies became extremely perturbed, unsurprising and data not shown.

Replacing the mitochondria targeting sequence with the MAVSTM targeting sequence in proved to be only slightly higher in expression (**Figure 5-4C; 5-4D**). Interestingly, utilizing the MAVSTM and C(1-29) targeting sequences in tandem revealed that the C(1-29) ER targeting sequence might cause “rippage”. During incubation with rapamycin, the morphology of the ER construct, as immunostained by the flag epitope tag, revealed strong mitochondrial morphology. However, the ER marker, mCherry-KDEL, appeared to remain ER-like in morphology. Overlap between these two channels did not demonstrate as much overlap as with anti-flag (for the C(1-29) construct and the mito marker, mito-GFP).

These results encouraged us to seek alternative mito and ER targeting sequences. The Takanari Inoue lab, *Johns Hopkins University*, has utilized the FKBP/FRB system to chemically induce organelle tethering of the mito and ER⁹². Fluorescence confocal microscopy confirmed morphology changes within 5 minutes of incubation with the chemical dimerization agent iRap⁹². We received constructs from them with ER membrane targeting sequence from Cytochrome b5 (Cb5) and outer mitochondrial membrane targeting sequence from the Tom20 protein. Overall, a remarkable difference was noted in expression levels. Both of these targeting sequences^{40,41} localized cleanly to their intended organelle and improved expression level (by immunostaining).

The constructs received from the Inoue lab lacked epitope tags. I wished to be able to

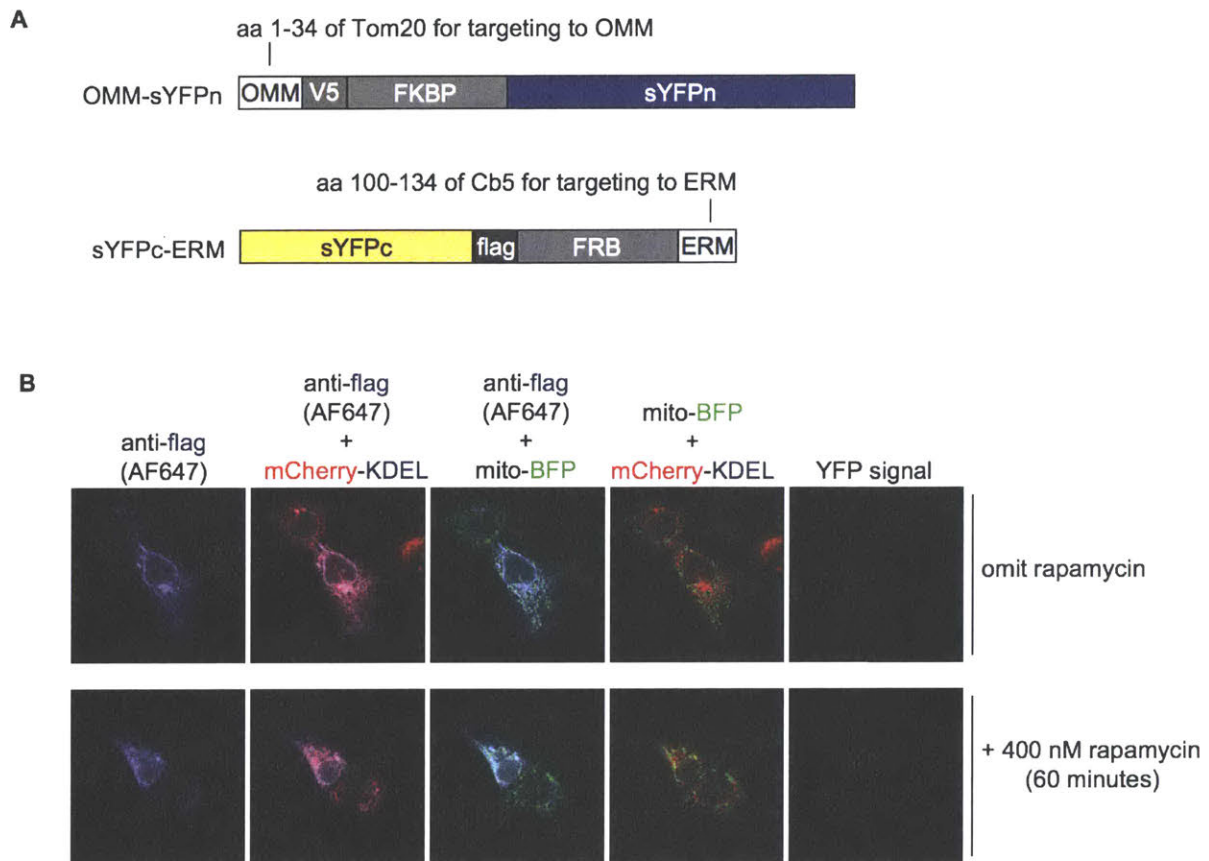


Figure 5-5. Testing mito and ER targeting sequences fused to the FKBP/FRB scaffold. (A) Testing OMM targeting with an N-terminal AKAP mitochondrial targeting sequence and ERM targeting with N-terminal C(1-29) ERM targeting sequence. **(B)** Cos7 cells were transfected to express the constructs from (A) with mCherry-KDEL as a ER organelle marker. After 24 hours of protein expression, cells were treated with 2 hours of rapamycin to evaluate of the effect of a protein-protein interaction on the localization of the FKBP and FRB fusions.

transfect Cos7 cells with both mito and ER split YFP constructs as well as mito and ER markers. With fluorescent microscopy, immunostaining and 4-color imaging could reveal all the localization of all 4 constructs. There remains the difficulty in that split YFP reconstitution generates YFP, which requires a fifth color channel. However, it has been previously determined that YFP reconstitution only occurs at incubation at 30 °C so to avoid background YFP signal, transfected cells were cultured at 37 °C

In order to perform these 4-color imaging experiments, I first cloned constructs that targeted the mito and the ER with improved targeting sequences, Tom20 and Cb5 (**Figure 5-5A**). I had also tested swapping the targeting of FKBP and FRB. The Inoue lab had shared that in their experience, the fusion of FRB-Cb5 with Tom20-FKBP orientation was the best, and our immunostaining had also revealed FRB targeted to the ER instead of the mito provides a slight enhancement in expression with clean localizations of both constructs; the N-terminal fragment of sYFP was consistently targeted to the mitochondria (which as a N-terminal targeting sequence), as to keep the native C-terminal end of the fragment at the C-terminus. The same logic was applied in keeping the C-terminal fragment of sYFP on the ER.

Transfecting these constructs into Cos7 cells revealed that no construct “ripping” was seen, which was the case with C(1-29), in which previously, it was suspected that upon rapamycin addition, the protein protein interaction was stronger than the ER tether, and immunostaining of C(1-29) constructs overlapped better with mito markers instead (**Figure 5-5B**). Using these targeting sequences, FKBP and FRB scaffold fusion proteins were correctly localized, and protein-protein interaction was inducible on a short time scale and appeared to be promising candidates for future sAPEX2 constructs.

Further construct designs utilized an HA epitope tag in place of the flag epitope tag, due to high background by immunostaining with antibodies tested. Additionally, the N- and C- termini of APEX2 are close in space, so to geometrically facilitate reconstitution, constructs were designed such that the newly generated N- and C- termini are also the overall constructs’ N and C terminus. With these improvements in construct scaffold design, the higher expression levels also led to higher levels of biotin-phenol labeling as detected by nAv647 staining. Further testing revealed that just 30-minute incubation of rapamycin was enough to see mito-ER overlap. Longer periods of incubation with rapamycin, such as over 2 hours, resulted in significant organelle morphological perturbances, emphasizing the need for a split APEX2 tool that would reconstitute rapidly. The overlay with organelle markers remained consistent.

As expression levels of both fragments can significantly affect the extent of background reconstitution, efforts were also made to optimize expression levels. As mentioned in Chapter 4, the small fragment EX suffered from protein degradation and lacked the stability of the AP fragment. Optimization parameters to reduce the protein expression of AP included changing the initial CMV promoter to that of a lower promoter: human Ubiquitin Promoter (hUBC).

Furthermore, introduction of the AP fragment was transitioned from transfection to infection to further reduce the expression level and ensure a higher percentage of cells that express both fragments. To ensure that every cell would express the EX fragment, we generated stable cell lines targeting the EX fragment to the ERM or the cytosol. Due to the instability of the EX fragment, it remained under the strong CMV promoter. Furthermore, to enhance stability when EX is cytosolic, it was fused to a HaloTag™ protein.

Preliminary testing of sAPEX at the mito-ER junction for proteomics in Cos7 cells

Mito-ER contacts are delicate and easily perturbed structures. Overexpression of various proteins, such as SYNJ2BP or even green fluorescent protein, can lead to dramatic distortion of one or both organellar membranes^{16,93}. We recognized that a high affinity sAPEX system such as AP-0 + EX or AP-3 + EX would likely produce severe disruption of native mito-ER contacts. It would be essential to use our finalized lower-affinity, proximity-dependent sAPEX fragment pair: AP + EX.

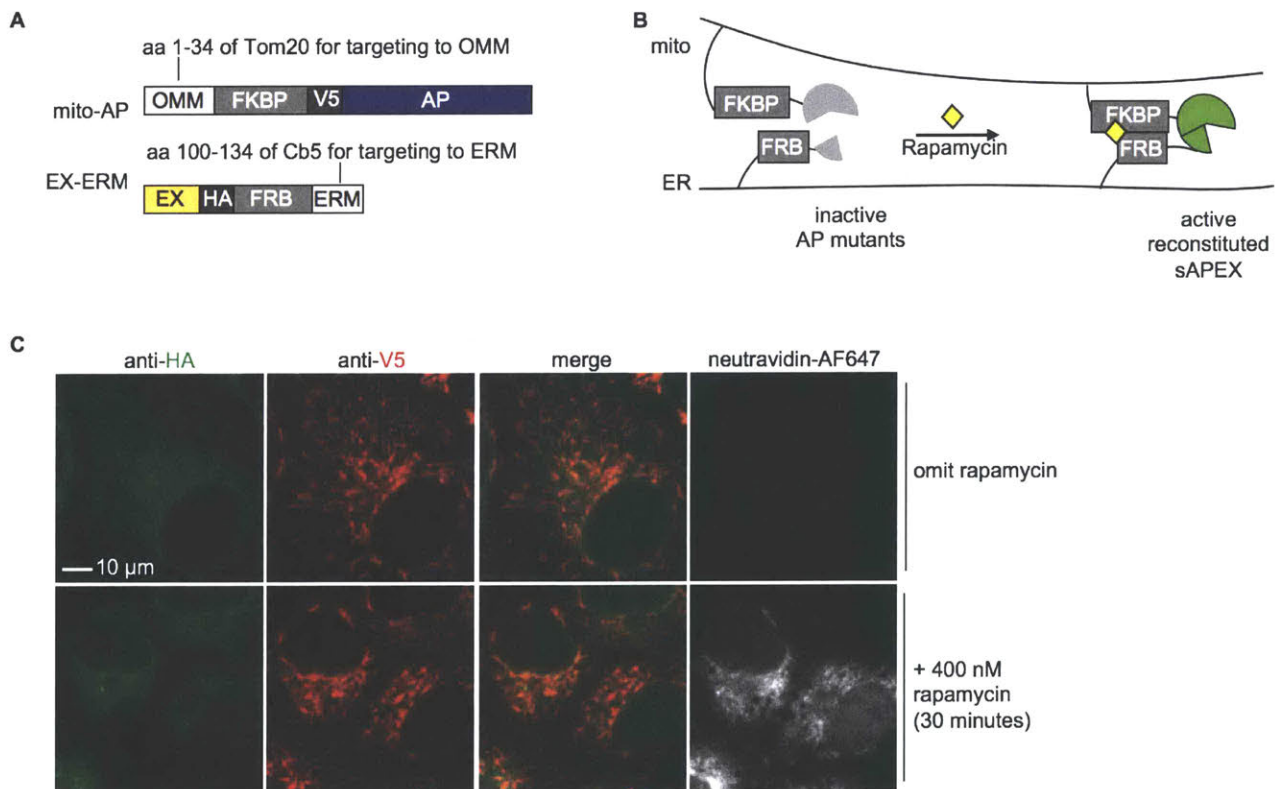
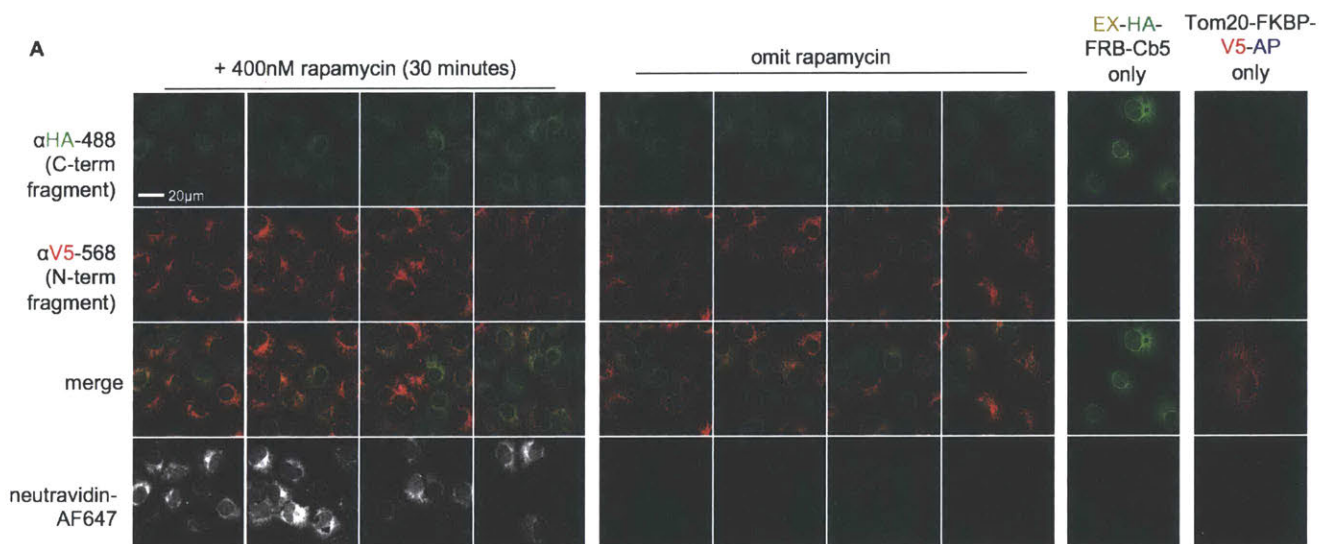
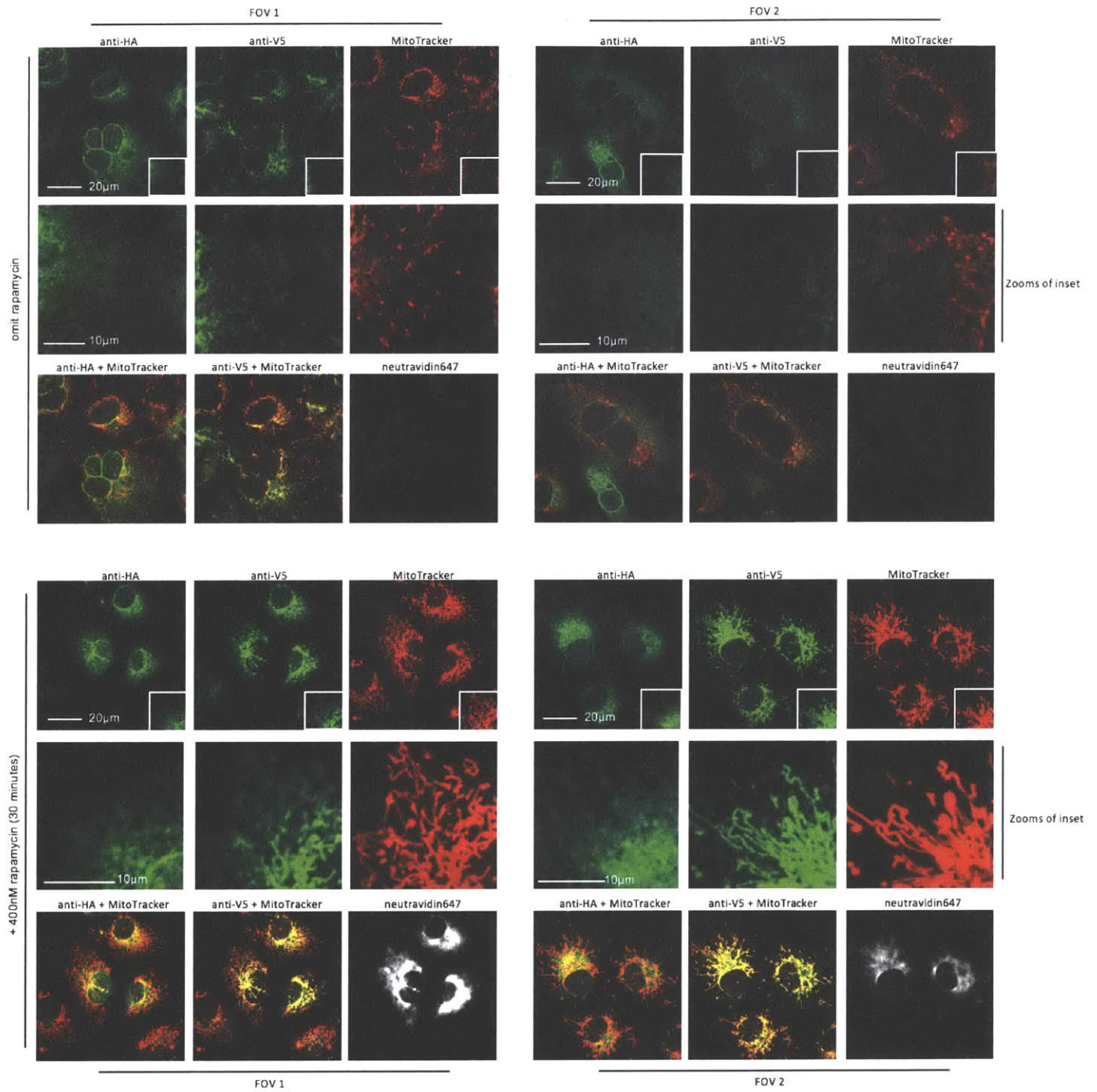


Figure 5-6. Testing sAPEX at mito-ER contact sites. (A) Constructs used for targeting AP and EX fragments of sAPEX to the outer mitochondrial membrane (OMM) and ER membrane (ERM), respectively. (B) Schematic overview of sAPEX applied at mito-ER contacts. Inactive fragments (grey) fused to FKBP and FRB can reconstitute to give active peroxidase (green) when driven together by an inducible protein-protein interaction (PPI) with the addition of rapamycin (yellow diamond). (C) sAPEX reconstitution in COS7 mammalian cells. COS7 cells stably expressing EX-ERM from (A) were infected with lentivirus containing mito-AP. 46 hours later, cells were incubated with heme for 90 minutes prior to the 30-minute biotin-phenol incubation in heme-free media in which rapamycin is added or omitted. Biotin-phenol labeling was performed for 1 minute, and cells were fixed and stained with neutravidin-AlexaFluor647 to visualize reconstituted sAPEX activity. HA and V5 staining show the localizations of total EX and total AP, respectively. The top row is a negative control with rapamycin omitted. Scale bar, 10 μ m. Additional fields of view are shown in **Figure 5-7**. This experiment has four biological replicates.

Finalized evolved fragment derived from generation 4 evolution, AP, and EX was cloned into the optimized targeting scaffold (**Figures 5-6A**). Desired sAPEX interaction at the mito-ER contact is depicted in the cartoon **Figures 5-6B**. After optimizing expression levels of both fragments and heme supplementation, we infected mito-AP into stable, monoclonal EX-ERM expressing COS7 cells, which are flat and thin for easier visualization of mitochondria and ER structures (**Figures 5-6C and 5-7A**). BP/streptavidin staining is observed in cells treated with rapamycin for 30 minutes. Consistent with the proximity-dependence of sAPEX reconstitution, no BP labeling is observed when rapamycin is omitted. We also stained the COS7 cells for mitochondrial and ER markers, and observed minimal disruption of organellar morphology, both before and after rapamycin addition (**Figure 5-7B; -5-7C**).



B



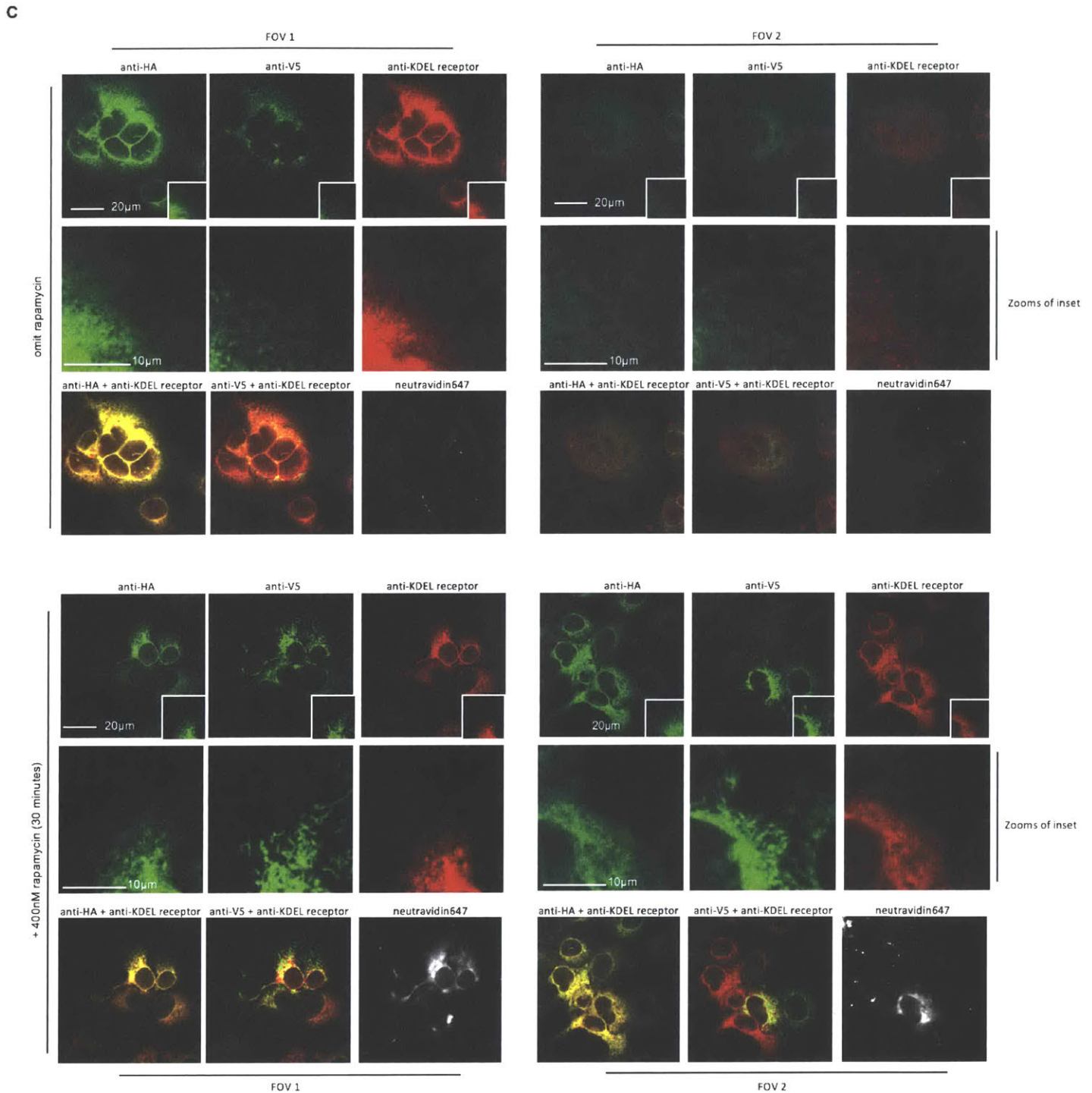


Figure 5-7. Examining sAPEX mito-ER targeting and morphology (A) Additional fields of view for mito-ER experiment shown in **Figure 5-6C**. Controls are also shown with each sAPEX fragment alone (right columns). Scale bar, 20 μm . Note that because BP labeling was performed live, biotinylated proteins can diffuse away from the site of labeling, leading to a diffuse

neutravidin staining pattern^{11,15}. **(B)** With matched conditions to **Figure 5-6C**, the morphology of the mitochondrial was visualized during confocal microscopy using MitoTracker Red. Scale bar, 20 μm unless otherwise indicated. The experiment has been biologically replicated twice. **(C)** Same as **(B)**, except the morphology of the ER was visualized during confocal microscopy using an antibody for the KDEL-receptor. Scale bar, 20 μm unless otherwise indicated. The experiment has been biologically replicated once.

After demonstrating the proof-of concept biotinylation by the detection of biotinylated proteins post sAPEX labeling, the final goal is to apply the tool for for proteomics – enabling the determination of the protein constituents at these mito-ER junctions.

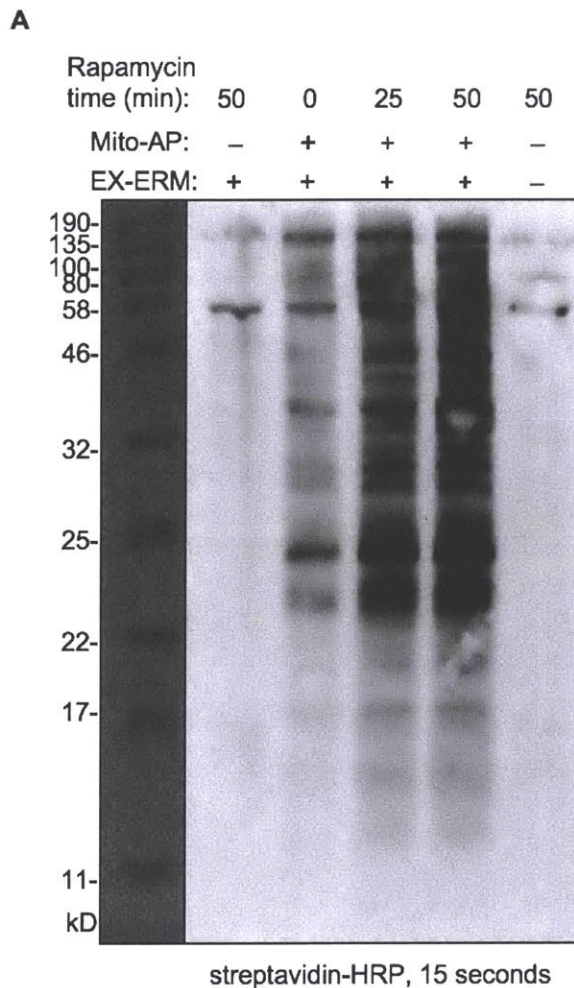


Figure 5-8. Western blot analysis of live biotin-phenol labeling of sAPEX targeted to the mito-ER **(A)** Monoclonal COS7 cells stably expressing EX-ERM were plated in 6 well plates and subsequently infected with lentivirus to express mito-AP, see **Figure 5-6A** for depiction of constructs. 46 hours later, cells were incubated with heme for 90 minutes prior to the 30 minute biotin-phenol incubation in heme-free media. Rapamycin omitted or added for an effective

concentration of 400 nM per well for 0, 25, and 50 minutes. Biotin-phenol labeling was performed for 1 minute before quenching, pelleting, and lysis. Lysates were run on a 12% SDS-PAGE gel, transferred to nitrocellulose membrane, incubated with streptavidin-HRP, and developed using Clarity reagent (Bio-Rad) for 5 min and imaged using an Alpha Innotech gel imaging system.

The monoclonal stable EX-ERM COS7 cells were tested on a larger scale, more similar to a mito-ER proteomics application. By using a streptavidin blot to stain for a biotinylation on generated cell lysates (**Figure 5-8**), we are able to get rapamycin dependent biotinylation under the right conditions, i.e., sufficiently low expression level regimes. In the process of preparing samples for both EM and for proteomics, stable COS7 cells expressing OMM-APEX2 and ERM-APEX2 were also prepared.

Using sAPEX for DAB labeling across the mito-ER

Utilizing optimized targeting sequences and expression levels of sAPEX resulted in PPI-dependent biotinylation that was detected at the mito-ER junction in COS7 cells. However, with these reduced expression level condition, the less sensitive DAB labeling assay required more optimization – viral titers that were sufficient for the biotin-phenol labeling assay were not for DAB labeling.

Preliminary testing of monoclonal stable EX-ERM COS7 cells with increased viral titer of the mito-AP increased the expression level; furthermore, allowing for at least 48 hour of protein synthesis and supplementing heme for 90 minutes at an effective concentration of 6 μ M prior to labeling was enough to generate DAB polymerization (**Figure 5-9**). While there is still detectable polymerization in the absence of rapamycin addition, the detected activity is indicative that there exists a concentration regime in which there is PPI-dependent reconstitution. The detectable polymerization was reproducible through attempts to optimize mito-AP expression levels. samples Preliminary samples which samples were treated with osmium, uranyl acetate, and embedded by durcupan resin were sent to our collaborator: Mark Ellisman, *University of California, San Diego*. Moreover, our collaborator had also tested DAB labeling in HEK 293T cells transfected to express sAPEX across the Mito-ER; while it did yield detectable polymerization, the high expression levels sAPEX unsurprisingly resulted in morphological perturbation of mito and ER organelles.

A

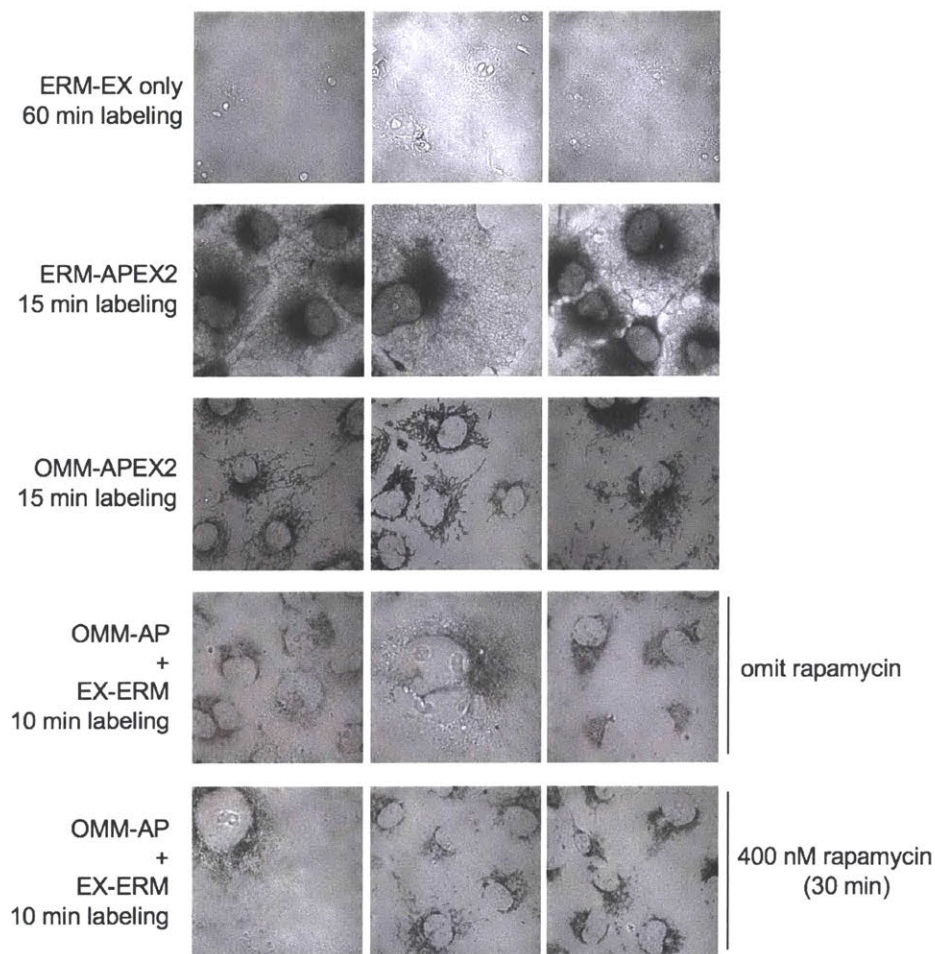


Figure 5-9. DAB labeling of sAPEX targeted to the mito-ER (A) Monoclonal COS7 cells stably expressing EX-ERM were plated in MatTek™ glass bottom imaging plates and subsequently infected to express mito-AP. In parallel, COS7 cells stably expressing full-length APEX2 targeted to the mitochondria and the ERM were also plated as a positive control. After 55 hours of protein expression, heme was supplemented for 90 minutes at an effective concentration of 6 μ M to sAPEX samples prior to fixation. Cells were then prepared for visualization by electron microscopy by previously published protocol⁹⁴. 3 representative fields of view are shown; images were taken by brightfield microscopy at 63x. Conditions of rapamycin addition, length of DAB labeling are as stated above.

Discussion

High-affinity fragments such as those of split GFP result in spontaneous and irreversible reconstitution⁹⁵; similarly, split HRP irreversibly reconstitutes in the ER without a PPI⁶¹. Split YFP utilizes fragments that reconstitute in a more PPI dependent manner^{96,97}, but it still suffers from background fluorescence from irreversible YFP self-assembly, especially at high expression levels^{98,99}. Lower affinity split proteins include split ubiquitin, which has a reported K_D of 7 μM ^{100,101}, and split firefly luciferase (Fluc) with a K_D of 6 μM ¹⁰². To our knowledge only NanoBiT, a Nanoluc complementation reporter, has been engineered specifically for both high conformational stability and low intrinsic affinity, with a reported K_D of 190 μM ¹⁰³. Unlike most other PCAs, sAPEX possesses the versatile capability to react with many different substrates. Through four rounds of evolution, we aimed to develop a sAPEX tool that had robust enzymatic activity while remaining low affinity to reduce background and artifacts.

We targeted sAPEX to the OMM and ERM of COS7 cells. To ensure proper targeting, a few targeting sequences were tested. The mito and ER targeting sequences obtained from the Takanari Inoue lab definitely helped to increase both protein expression level and proper organelle targeting. Despite multiple days of co-translation on membranes of organelles known to come into contact, sAPEX still demonstrated PPI-dependent reconstitution while avoiding perturbations to the organelles at levels of expression suitable for biotinylation. We note that some optimization was required to attain this PPI dependency; for instance, we changed the promoter of OMM-FKBP-EX from CMV to human ubiquitin promoter (hUBC) to reduce the overall protein expression level, which reduced self-assembly. As organelle contacts are not restricted to just mito-ER, sAPEX potentially provides the ability to investigate and map the protein residents at other organelle-organelle contact sites. Moreover, as sAPEX activity has been demonstrated to reconstitute in highly specific subcellular regions that are intractable for single gene constructs, it expands the toolkit of proximity labeling technologies.

Experimental Methods

See Table 5 for constructs used.

Cloning

See Table 1 for a list of genetic constructs used in this study, with annotated epitope tag, promoters, resistance, vector, linkers, etc. For cloning the constructs, PCR fragments were amplified using Q5 polymerase (New England BioLabs (NEB)) or PfuUltra II Fusion HS DNA polymerase (Agilent Technologies). The vectors were double-digested and ligated to gel-purified PCR products by T4 ligation or Gibson assembly. Ligated plasmid products were introduced by heat shock transformation into competent XL1-Blue bacteria. The APEX2 gene used for initial cut site screening was amplified from vimentin-APEX2 with codons optimized for mammalian expression. Mutants of AP were generated either using QuikChange mutagenesis (Stratagene) or isolated from individual yeast clones and transferred to simammalian expression vectors using standard cloning techniques.

Mammalian cell culture and transfection

HEK 293T and COS7 cells from ATCC (passages <20) were cultured as monolayers in complete growth media consisting of 50% DMEM (Dulbecco's Modified Eagle medium, Gibco) and 50% MEM (Minimum Essential medium, Gibco) supplemented with 10% w/v FBS (Fetal Bovine Serum, Sigma), 50 units/mL penicillin, and 50 µg/mL streptomycin at 37 °C under 5% CO₂. For confocal imaging experiments, cells were grown on 7 × 7 mm glass coverslips in 48-well plates (Corning). To improve the adherence of HEK 293T cells, the glass coverslips were pretreated with 50 µg/mL fibronectin (Millipore) in Dulbecco's phosphate-buffered saline (DPBS), pH 7.4, for 20 min at 37 °C, followed by two washes with growth media prior to cell plating. Cells were transfected at 70–85% confluence using Lipofectamine 2000 (Life Technologies), with 1.0 µL Lipofectamine 2000 and 200 ng per plasmid per 300,000 cells in serum-free media for 3-4 h, after which transfection media was replaced with fresh complete growth media. Cells were labeled and/or fixed 18–24 h after transfection. The cell line was not authenticated or tested for mycoplasma.

Lentivirus generation and mammalian cell infection

To generate viruses, HEK 293T cells plated in a T25 flask (Corning) were transfected at ~75-90% confluency with 5 mL of serum-free MEM containing the lentiviral plasmid encoding the gene of interest (2500 ng), the lentiviral packaging plasmids dR8.91 (2250 ng) and pVSV-G (250 ng), and 30 μ L of 1 mg/mL Polyethylenimine “Max” (Polysciences) pH 7.1 for 3-4 h, followed by media change with complete MEM. 48 h after transfection, the cell medium containing lentivirus was harvested, filtered through a 0.45- μ m filter, and added to other fresh cell cultures to induce expression of the gene of interest. Typically, for both HEK 293T and COS7 cultures plated in wells of a 48-well plate (0.95 cm² per well), cells were plated the night before in 1:1 complete growth media. HEK 293T and COS7 cells, at ~70% and ~85% confluency, respectively, were then infected by exchanging the overnight medium with 150-200 μ L of complete MEM and 50-100 μ L of viral medium. Cells were then incubated for 48 h at 37 °C with 5% CO₂ prior to heme incubation, labeling, and imaging.

Generation of stable cell lines

To generate stable cell lines, constructs of interest were cloned into the lentiviral vector pLX304 to make virus, as described above. Low passage HEK 293T cells plated in 6-well plates (Corning) were infected with 800 μ L-1000 μ L of lentivirus. After 2 days at 37°C, 5% CO₂, each well was trypsinized and replated into two T25 flasks in full growth media containing 8 mg/ml blasticidin S HCl (Calbiochem, CAS no. 3513_03_9). Media was changed to fresh 8 mg/ml blasticidin S HCl media every 24 h. Cells were trypsinized and replated before reaching full confluency within the first 7 days. After 7–10 days of selection, surviving cells can be used for experiments. To make cells stocks for long-term storage at -80 °C, cells from a ~90% confluent T75 flask are harvested, pelleted, resuspended 3x concentration in complete media with 5% DMSO, and frozen at a controlled rate to -80 °C overnight using a Mr. Frosty Freezing Container (Thermo Scientific™). COS7 cells stably expressing EX-FRB-ERM was prepared similarly, with a selection in media containing 10 mg/mL blasticidin S HCl.

Biotin-phenol labeling

Genes were introduced into HEK 293T or COS7 cells through either transient transfection with Lipofectamine 2000 or lentiviral infection. After 18–24 h (transfection) or 48 h (lentivirus), biotin-

phenol (bp) labeling was initiated by changing the medium to 200 μ L of fresh growth medium containing 500 μ M bp that was sonicated for at least 5 minutes to ensure bp was fully dissolved. Cells were incubated at 37 $^{\circ}$ C under 5% CO₂ for 30 min according to previously published protocols. To initiate labeling, 2 μ L of 100 mM H₂O₂ was spiked into each well, for a final concentration of 1 mM H₂O₂, and the plate was immediately gently agitated. To quench the reaction after 1 minute, the bp solution is aspirated, and cells were immediately fixed with 4% formaldehyde with 5 mM Trolox in PBS at room temperature for 5 minutes before continuing fixation on ice for an additional 25 minutes. Cells were then washed with chilled DPBS three times and permeabilized with pre-chilled methanol at -20° C for 10 min. Cells were washed again three times with DPBS and blocked with 3% BSA in DPBS for 1 h to overnight with rocking at 4 $^{\circ}$ C. To detect the expression of sAPEX fusions, cells were incubated with primary antibodies mouse anti-V5 (Life Technologies, Cat. No. R96025, 1:1500 dilution) and rabbit anti-HA (Cell Signaling, 1:1000 dilution) in 1% BSA in DPBS for 1 h to overnight at 4 $^{\circ}$ C followed by 4 \times 5 min washes with chilled DPBS. Cells were then incubated with a 1% BSA in DPBS solution containing secondary Alexa Fluor 488-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11001, 1:1000 dilution), Alexa Fluor 568-goat anti-mouse IgG (Life Technologies, Cat. No. A-11004, 1:1000 dilution), and homemade streptavidin–Alexa Fluor 647 (1:1000) for 25-45min at 4 $^{\circ}$ C with rocking. Cells were then washed 4 times for 5 min with chilled DPBS and imaged by confocal microscopy.

DAB labeling

HEK 293T and COS7 cells were plated on glass coverslips and then transfected or infected to express sAPEX fragments AP and EX targeted to the mito and ER, respectively. After sufficient protein expression, cells were fixed with 4% formaldehyde in PBS for 20 min at room temperature. Cells were subsequently washed twice with room temperature DPBS, incubated for 10 min in buffer containing 20 mM glycine to quench unreacted formaldehyde, and washed 3 more times with DPBS. A fresh solution of 0.5 mg/mL (1.4 mM) DAB (5.4 mg/mL of 3,3'-diaminobenzidine, Sigma, dissolved in 0.1 M HCl) was combined with 0.03% (v/v) (10 mM) H₂O₂ in DPBS for 15 min. To quench polymerization, the DAB solution was removed and cells were placed on ice and washed four times for 5 min with chilled DPBS, before imaging.

Labeling the mitochondrial and endoplasmic reticulum in COS7 cells for imaging by confocal microscopy

Preparing COS7 cells Mitochondria morphology was visualized using MitoTracker Red CMXRos (ThermoFisher, M7512); it was diluted to a final concentration of 200nM and added during the 30-minute biotin-phenol incubation. All subsequent steps are the same for imaging by microscopy, as the dye is retained after aldehyde-fixation. After following standard primary immunostaining protocol (see biotin-phenol labeling protocol), cells were then incubated with a 1% BSA in DPBS solution containing secondary Alexa Fluor 488-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11001, 1:1000 dilution), Alexa Fluor 405-goat anti-mouse IgG (Life Technologies, Cat. No. A-31553, 1:1000 dilution), and homemade streptavidin–Alexa Fluor 647 (1:1000) for 25-45min at 4 °C with rocking, to keep the MitoTracker Red channel free.

For visualizing the morphology of the endoplasmic reticulum, COS7 samples were prepared as standard biotin-phenol labeled samples. Having tested a variety of available antibodies against resident ER proteins, a polyclonal rabbit antibody against the KDEL-receptor had performed well. To use the antibody in tandem with our anti-HA rabbit antibody, two rounds of immunostaining was performed. After the first round of secondary immunostaining (same antibody conditions as the MitoTracker sample), cells were stringently washed 5x for 5 minutes each prior to a second fixation with 4% formaldehyde for 25 minutes on ice. Cells were then blocked with 2% BSA (w/v) in DPBS for one hour at 4 °C. Afterwards, a second round of primary immunostaining was performed using an antibody against the resident ER protein, KDEL receptor. (Santa Cruz Biotechnolgy, sc-33806) for one hour at 4 °C. After washing, and secondary immunostaining with Alexa Fluor 568-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11011, 1:2000 dilution in 2% BSA) for 30 minutes at 4 °C, cells were washed 5x for 5 minutes each before imaging.

Microscopy

Confocal imaging was performed on a Zeiss AxioObserver inverted confocal microscope with 10× air and 40× oil-immersion objectives, outfitted with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), 561 (DPSS) and 640-nm (diode) lasers (all 50 mW). All images were acquired and processed using Slidebook 5.0 or 6.0 software (Intelligent Imaging Innovations), through a 48× or 63× oil-immersion

objective CFP (405 laser excitation; 635/85 emission), Alexa Fluor 488 (491 laser excitation; 528/38 emission), Alexa Fluor 568 (561 laser excitation; 617/73 emission), Resorufin (491 laser excitation; 550/585), Alexa Fluor 647 (647 excitation; 680/30 emission), and differential interference contrast (DIC). Acquisition times ranged from 100 to 1000 ms. DAB labeled cells were imaged by bright field; acquisition time was 200 ms. Imaging conditions and intensity scales were matched for each data set presented together.

Streptavidin blotting

For streptavidin blotting, cells were grown on polystyrene 6-well plates (Corning) and labeled under the same conditions described above (see “Biotin-phenol labeling”). After 1 min of labeling, the cells were washed three times with room temperature quencher solution (10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in DPBS) and then scraped and pelleted by centrifugation at 3000g for 10 min. The cell pellet was then lysed by gentle resuspension into peroxidase-quencher containing RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 1× protease cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF (phenylmethylsulfonyl fluoride), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox). After 2 minutes on ice, the lysates were clarified by centrifugation at 15000 × g for 10 min at 4 °C before separation on 10% SDS-PAGE gel. For blotting analysis, gels were transferred to nitrocellulose membrane, stained by Ponceau S (10 min in 0.1% (w/v) Ponceau S in 5% acetic acid/water), and blocked with “blocking buffer” (3% (w/v) BSA and 0.1% Tween-20 in Tris-buffered saline) for 1 h at room temp or 4 °C overnight. The blots were immersed in streptavidin-HRP in blocking buffer (Thermo Scientific, cat. no. 21126, 1:3000 dilution) at room temperature for 1 h and then rinsed with blocking buffer 4 times for 5 min. For assessing comparative fragment expression level, identical gels and blots were prepared in parallel and immersed in blocking buffer containing either mouse anti-V5 (Life Technologies, R96025, 1:3000 dilution) or mouse anti-HA (Santa Cruz Biotechnology, G1817, 1:3000 dilution) for 1 h to overnight at 4°C. The blots were rinsed with blocking buffer 4 times for 5 minutes before being immersed in anti-Mouse-HRP (Bio-Rad, 1:3000) in blocking buffer at room temp for 1 h, rinsed with blocking buffer 4 times for 5 min again. All blots were developed using Clarity reagent (Bio-Rad) for 5 min and imaged using an Alpha Innotech gel imaging system.

Heme supplementation

Where indicated, heme was supplemented by dilution from a thawed 300 μ M heme stock solution, which was prepared by dissolving hemin-Cl (Sigma) in 10 mM NaOH with sonication for 3 min, aliquoted, flash frozen in liquid N₂, and stored at -80 °C. Heme stocks were used within 1 h after thawing and never reused. Heme stock is diluted to 5 μ M and 6 μ M in complete growth media and added to HEK 293T and COS7 cells, respectively, for 90 min prior to bp incubation (without heme) or DAB labeling. For yeast heme supplementation in Generations 2-4, washed yeast cells were pelleted at 7000 \times g for 1 min and resuspended to a concentration of 1.5 \times 10⁷ yeast cells/mL in .1% BSA in DPBS-B at the designated heme concentrations (Figure S3B, S4B, S5A) just prior to EX incubation.

Table 5. Plasmids used in Chapter 5

Plasmid name	plasmid vector	promoter	Expression in	Features
p117	PLX304	CMV	lentivirus	EX-7aa linker-HA-5aa linker-FRB-32aa linker-ERM
p118	PLX304	hUBC	lentivirus	OMM-12aa linker-V5-FKBP-10aa linker-V5-12aa linker-AP
p119	PCDNA3	CMV	mammalian	OMM-V5-FKBP-sYFPn
p120	PCDNA3	CMV	mammalian	sYFPc-flag-FRB-ERM
p121	PCDNA3	CMV	mammalian	flag-sGFPn(1-10)-10aa linker-OMM
p122	PCDNA3	CMV	mammalian	ERM-10aa linker-sGFP(11)c-myc
p123	PLX304	CMV	lentivirus	mCherry-KDEL
p124	PLX304	CMV	lentivirus	Mito-BFP
p125	PLX304	CMV	lentivirus	OMM-V5-FKBP-sYFPn-p2a-sYFPc-flag-FRB-ERM
p126	PLX304	CMV	lentivirus	OMM-APEX2
p127	PLX304	CMV	lentivirus	ERM-APEX2

References

1. Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science* (80-.). **339**, 1328–1331 (2013).

2. Hung, V. *et al.* Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol. Cell* **55**, 332–341 (2014).
3. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).
4. Hung, V. *et al.* Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* **6**, (2017).
5. Szabadkai, G. *et al.* Chaperone-mediated coupling of endoplasmic reticulum and mitochondrial Ca²⁺ channels. *J. Cell Biol.* **175**, 901–911 (2006).
6. Rizzuto, R. *et al.* Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca²⁺ responses. *Science (80-.)*. **280**, 1763–1766 (1998).
7. Kornmann, B., Osman, C. & Walter, P. The conserved GTPase Gem1 regulates endoplasmic reticulum-mitochondria connections. *Proc. Natl. Acad. Sci.* **108**, 14151–14156 (2011).
8. Kornmann, B. *et al.* An ER-mitochondria tethering complex revealed by a synthetic biology screen. *Science (80-.)*. **325**, 477–481 (2009).
9. Lewin, T. M., Van Horn, C. G., Krisans, S. K. & Coleman, R. A. Rat liver acyl-CoA synthetase 4 is a peripheral-membrane protein located in two distinct subcellular organelles, peroxisomes, and mitochondrial-associated membrane. *Arch. Biochem. Biophys.* **404**, 263–270 (2002).
10. Rusiñol, A. E., Cui, Z., Chen, M. H. & Vance, J. E. A unique mitochondria-associated membrane fraction from rat liver has a high capacity for lipid synthesis and contains pre-Golgi secretory proteins including nascent lipoproteins. *J. Biol. Chem.* **269**, 27494–27502 (1994).
11. Friedman, J. R. *et al.* ER Tubules Mark Sites of Mitochondrial Division. *Science (80-.)*. **334**, 358–362 (2011).
12. Murley, A. *et al.* ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. *Elife* **2013**, (2013).
13. De Brito, O. M. & Scorrano, L. Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* **456**, 605–610 (2008).
14. Filadi, R. *et al.* Mitofusin 2 ablation increases endoplasmic reticulum–mitochondria coupling. *Proc. Natl. Acad. Sci.* **112**, E2174–E2181 (2015).
15. Cosson, P., Marchetti, A., Ravazzola, M. & Orci, L. Mitofusin-2 Independent Juxtaposition of Endoplasmic Reticulum and Mitochondria: An Ultrastructural Study. *PLoS One* **7**, (2012).
16. Hirabayashi, Y. *et al.* ER-mitochondria tethering by PDZD8 regulates Ca²⁺ dynamics in mammalian neurons. *Science (80-.)*. **358**, 623–630 (2017).
17. Smirnova, E., Griparic, L., Shurland, D.-L. & Blik, A. M. van der. Dynamin-related Protein Drp1 Is Required for Mitochondrial Division in Mammalian Cells. *Mol. Biol. Cell* **12**, 2245–2256 (2001).
18. Henning, M. S. *et al.* PDZD8 is a novel moesin-interacting cytoskeletal regulatory protein that suppresses infection by herpes simplex virus type 1. *Virology* **415**, 114–121 (2011).
19. Rowland, A. A. & Voeltz, G. K. Endoplasmic reticulum-mitochondria contacts: Function of the junction. *Nature Reviews Molecular Cell Biology* (2012). doi:10.1038/nrm3440
20. De Brito, O. M. & Scorrano, L. An intimate liaison: Spatial organization of the endoplasmic reticulum-mitochondria relationship. *EMBO Journal* (2010). doi:10.1038/emboj.2010.177

21. Wideman, J. G., Gawryluk, R. M. R., Gray, M. W. & Dacks, J. B. The ancient and widespread nature of the ER-mitochondria encounter structure. *Mol. Biol. Evol.* (2013). doi:10.1093/molbev/mst120
22. Wieckowski, M. R. M. R., Giorgi, C., Lebedzinska, M., Duszynski, J. & Pinton, P. Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells. *Nat. Protoc.* (2009). doi:10.1038/nprot.2009.151
23. Yamagata, M. & Sanes, J. R. Transgenic strategy for identifying synaptic connections in mice by fluorescence complementation (GRASP). *Front. Mol. Neurosci.* (2012). doi:10.3389/fnmol.2012.00018
24. Ma, Y. & Taylor, S. S. A molecular switch for targeting between endoplasmic reticulum (ER) and mitochondria: Conversion of a mitochondria-targeting element into an ER-targeting signal in DAKAP1. *J. Biol. Chem.* (2008). doi:10.1074/jbc.M710494200
25. Komatsu, T. *et al.* Organelle-specific, rapid induction of molecular activities and membrane tethering. *Nat. Methods* **7**, 206–208 (2010).
26. Snapp, E. L. *et al.* Formation of stacked ER cisternae by low affinity protein interactions. *J. Cell Biol.* **163**, 257–269 (2003).
27. Martell, J. D., Deerinck, T. J., Lam, S. S., Ellisman, M. H. & Ting, A. Y. Electron microscopy using the genetically encoded APEX2 tag in cultured mammalian cells. *Nat. Protoc.* (2017). doi:10.1038/nprot.2017.065
28. 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).
29. Martell, J. D. *et al.* A split horseradish peroxidase for the detection of intercellular protein-protein interactions and sensitive visualization of synapses. *Nat. Biotechnol.* **34**, 774–780 (2016).
30. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90 (2002).
31. Hu, C. D., Chinenov, Y. & Kerppola, T. K. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* **9**, 789–798 (2002).
32. Chumakov, S. P., Kravchenko, Y. E. & Chumakov, P. M. Protein complementation as tool for studying protein-protein interactions in living cells. *Mol. Biol.* **46**, 627–638 (2012).
33. Horstman, A., Tonaco, I. A. N., Boutilier, K. & Immink, R. G. H. A Cautionary note on the use of split-YFP/BiFC in plant protein-protein interaction studies. *International Journal of Molecular Sciences* **15**, 9628–9643 (2014).
34. Jourdan, M. & Searle, M. S. Cooperative assembly of a natively like ubiquitin structure through peptide fragment complexation: Energetics of peptide association and folding. *Biochemistry* **39**, 12355–12364 (2000).
35. Xing, S., Wallmeroth, N., Berendzen, K. W. & Grefen, C. Techniques for the analysis of protein-protein interactions in vivo. *Plant Physiol.* pp.00470.2016 (2016). doi:10.1104/pp.16.00470
36. Paulmurugan, R. & Gambhir, S. S. Combinatorial library screening for developing an improved split-firefly luciferase fragment-assisted complementation system for studying protein-protein interactions. *Anal. Chem.* **79**, 2346–2353 (2007).
37. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).

Chapter 6. Reconstitution of sAPEX on a target RNA motif

Introduction

Having established that sAPEX can be reconstituted with sub-organelle specificity, we next wanted to test the functionality of this system at substantially smaller targets: individual macromolecular complexes. Another application for which the full length APEX2 genetic fusion strategy may be unsuitable is for profiling the interactome of specific cellular RNAs. While there are several robust methods to identify RNA interaction partners for specific proteins of interest⁵¹⁻⁵³, the converse problem—identifying proteins that interact with a particular RNA species—is much more challenging. One could envision fusing the APEX2 protein to a high-affinity RNA-binding protein (RBP; for example, the bacteriophage MS2 coat protein⁵⁴) allowing the peroxidase to be ectopically targeted to transcripts that are tagged with that RBP's cognate RNA motif. However, a major technical challenge would be the large pool of excess, catalytically active APEX2-RBP fusion protein that is not docked to the tagged RNA, resulting in background labeling that masks the specific signal (**Figure 6-1A;B**). When, in control reactions, we expressed full-length APEX2 as a fusion with MCP, we observed extensive nuclear labeling in live-cell biotinylation assays¹¹. However, the extent and localization of this signal was qualitatively unchanged in the presence or absence of MS2-tagged TERC RNA, implying that substantial biotinylation occurred outside of TERC complexes (**Figure 6-1C**).

A general solution to these problems and related ones could be a split form of APEX2, in which two inactive fragments of APEX2 reconstitute to give an active peroxidase (**Figure 6-2A;B**). One could use an intersectional approach to restrict APEX2 activity specifically to sites of interest, such as mito-ER contacts only, or specific RNA binding sites only – thus eliminating the background labeling from protein overexpression and off-targeting. As a first proof-of-principle experiment, we attempted to reconstitute sAPEX on ribonucleoproteins (RNPs) motivated both by the need for technologies that identify proteins interacting with specific RNAs¹⁰⁴⁻¹⁰⁶, and by the relative ease with which structured RNA domains—to which we could target our sAPEX2 components—can be designed *de novo*¹⁰⁷. Although such domains have also been used to recruit full-length APEX to a target RNA, we note the sizeable pool of catalytically active APEX2 that remained unbound under these conditions, thereby producing a high, nonspecific background

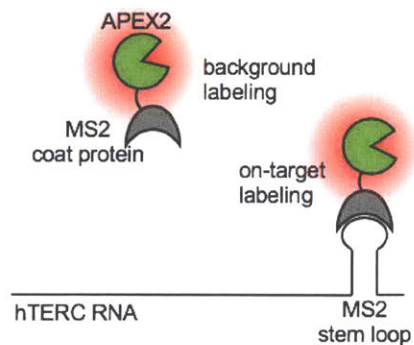
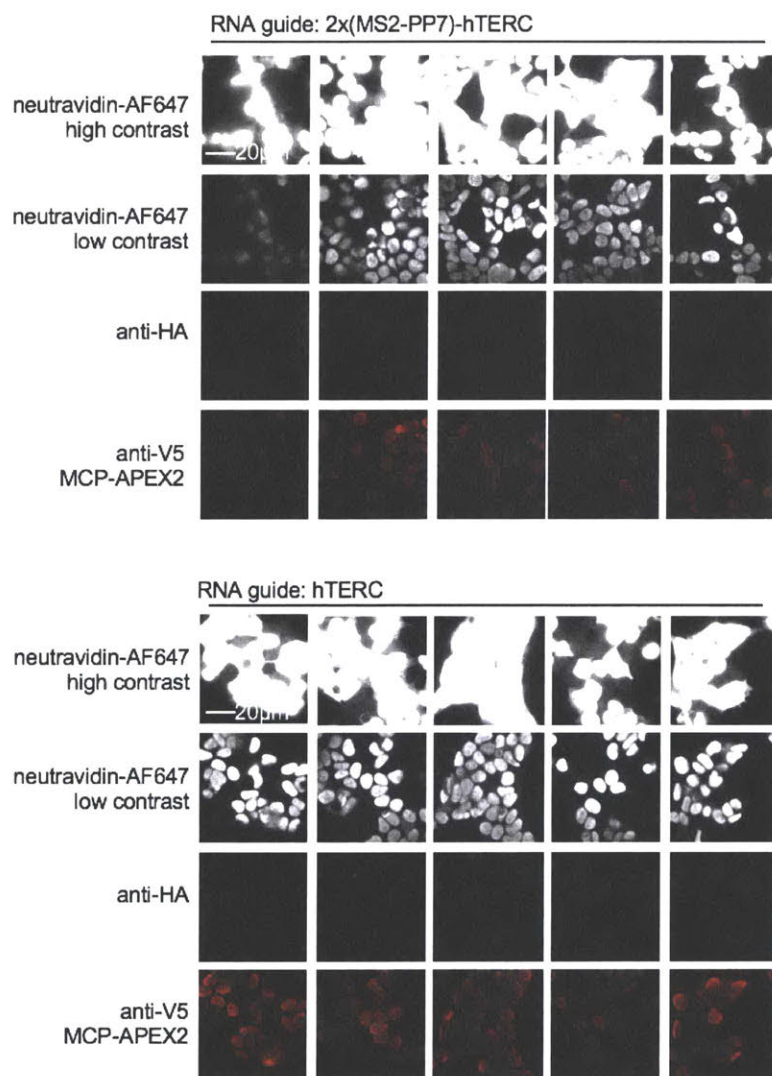
A**B****C**

Figure 6-1. Testing MS2 stem-loop coating protein fused to full length APEX2 to target it to an RNA binding site motif. (A) Constructs used to test full length APEX2 targeting to RNA with and without MS2 stem loops. MCP is the MS2 coat protein that binds to MS2 stem-loop. Protein fusions are targeted to the nucleus by a N-terminal nuclear localization signal (NLS). (B) For potential applications involving interaction mapping of specific cellular RNAs, full-length APEX2 could give high background due to excess, untargeted APEX2-MCP. (C) Biotin-phenol labeling to test full length APEX2 targeting to RNAs. HEK 293T cells stably expressing the MCP-APEX2 construct were transfected with the indicated RNA constructs. 22 hours later, cells underwent biotin-phenol incubation and labeling prior to being fixed and stained with neutravidin-AlexaFluor647 to visualize reconstituted peroxidase activity, anti-HA, and anti-V5 antibody to visualize APEX2 expression. Five fields of view are shown per condition. Scale bar, 20 μm . This experiment has three biological replicates.

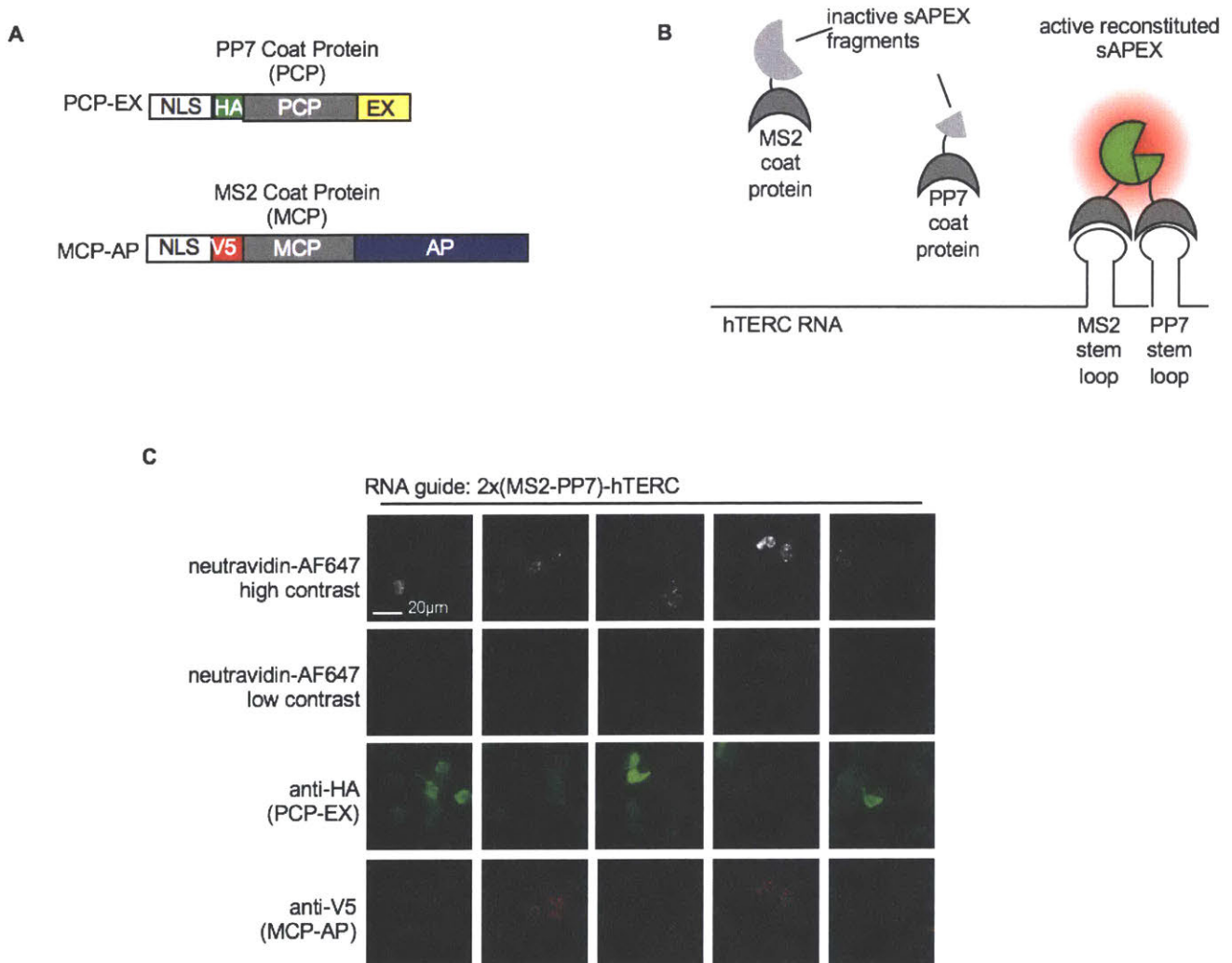
signal (**Figure 6-1**). To circumvent this problem, we reasoned that target transcripts could be instead appended with a structured RNA cassette that recruits two distinct RNA-binding proteins, each of which would be fused to a sAPEX fragment. Hence, catalytically viable APEX would be exclusively reconstituted at the surface of targeted transcripts (**Figure 6-2B**).

Applying sAPEX and optimizing expression levels of AP and EX fragment

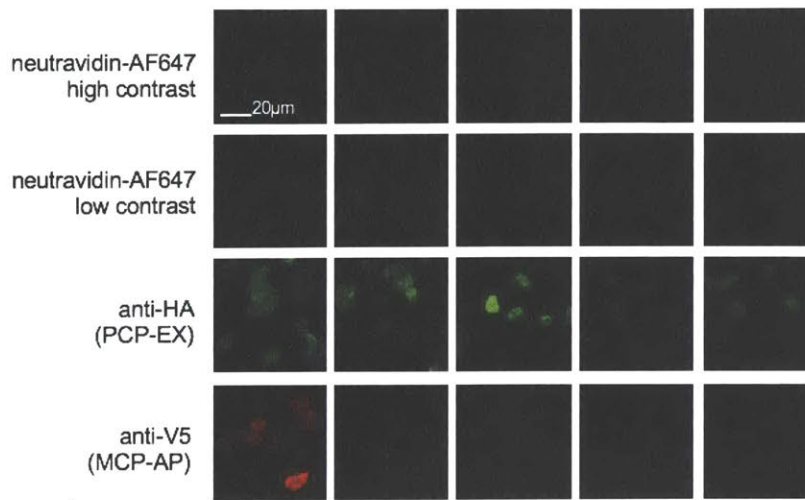
Towards this goal, we attempted to reconstitute sAPEX on TERC, the noncoding RNA component of the RNP telomerase, which synthesizes the ends of chromosomes in many clades of eukaryotes¹⁰⁸. In addition to providing the template for telomere synthesis, the TERC ncRNA is also thought to serve as a structural “scaffold” onto which the other holoenzyme components assemble¹⁰⁹. Critically, proper biogenesis of functional telomerase RNPs can be recapitulated when TERC RNA is transiently expressed from a plasmid, even if the RNA is appended with exogenous sequences at its 5' end¹¹⁰. We therefore designed a series of variants in which the TERC 5' terminus is appended by cassettes of motifs recognized by RNA-binding proteins (**Figure 6-3**). Work with designing and cloning the RNA motif guides were performed by our collaborator, Professor David Shechner, at University of Washington.

We chose to express AP and EX as fusions to the MS2 and PP7 bacteriophage nucleocapsid proteins—respectively termed MCP and PCP—which recognize disparate RNA motifs with exquisitely high affinity and specificity (cognate RNA $K_D \sim 1\text{nM}$, non-cognate RNA $K_D \sim 1\text{mM}$)

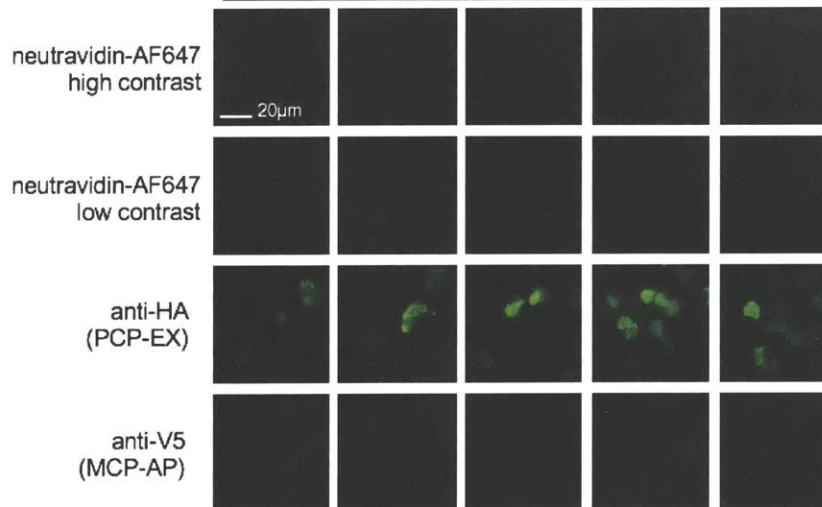
54,111,112,113 (Figure 6-2A). When testing single gene fusion of full length APEX2 in tandem with testing split APEX, the biotinylation pattern of full length APEX2 is indiscernible between TERC RNA guide containing the MS2 stem loop recognition motif and those that do not (Figure 6-1C).



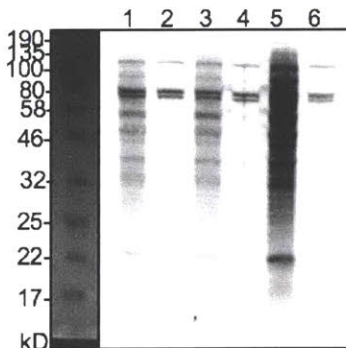
RNA guide: 2xMS2-hTERC + 2xPP7-hTERC



RNA guide: hTERC



D



SA-HRP blot
1 min exposure

1. 1000 ng MCP-AP, 1500 ng PCP-EX, 5000 ng 2x(MS2-PP7)-hTERC
2. 1000 ng MCP-AP, 1500 ng PCP-EX, 5000 ng hTERC
3. 1500 ng MCP-AP, 2000 ng PCP-EX, 5000 ng 2x(MS2-PP7)-hTERC
4. 1500 ng MCP-AP, 2000 ng PCP-EX, 5000 ng hTERC
5. 500 ng of MCP-FL, 5000 ng 2x(MS2-PP7)-hTERC
6. Untransfected

Figure 6-2. Testing split APEX targeted to an RNA binding site motif. (A) Constructs used to test sAPEX targeting to RNA. MCP is the MS2 coat protein that binds to MS2 stem-loop. PCP is the PP7 coat protein that binds to PP7 RNA stem-loop. Protein fusions are targeted to the nucleus by a N-terminal nuclear localization signal (NLS). (B) Rationale for using sAPEX instead of full-length APEX2. For potential applications involving interaction mapping of specific cellular RNAs, full-length APEX2 could give high background due to excess, untargeted APEX2-MCP (Figure 6-1B). On the other hand, cells expressing sAPEX (with N and C-terminal fragments fused to MCP and PCP, respectively) would localize peroxidase activity and labeling only to hTERT (human telomerase RNA component) RNA sites with adjoining MS2 and PP7 stem-loops, as shown. (C) Immunofluorescence analysis of biotin-phenol labeling. HEK 293T cells stably expressing the MCP-AP were transfected with PCP-EX expression plasmid and the indicated RNA constructs. The experiment was done in tandem with Figure 6-1C. 22 hours later, cells underwent biotin-phenol incubation and labeling prior to being fixed and stained with neutravidin-AlexaFluor647 to visualize reconstituted peroxidase activity, anti-HA antibody to visualize EX expression, and anti-V5 antibody to visualize AP expression. Same experiment with HEK 293T cells stably expressing MCP-APEX2 is shown at right for comparison. Five fields of view are shown per condition. Scale bar, 20 μm . (D) Western blot analysis of split APEX targeted to RNA binding motif. HEK 293T cells were plated in a 6-well plate. After one night of growth, cells were transfected with the listed amounts of DNA as shown to the right. 24 hours later, biotin-phenol labeling was performed and cells were quenched, pelleted, lysed for a western blot. Streptavidin-HRP was utilized to visualize the extent of biotinylation for each condition.

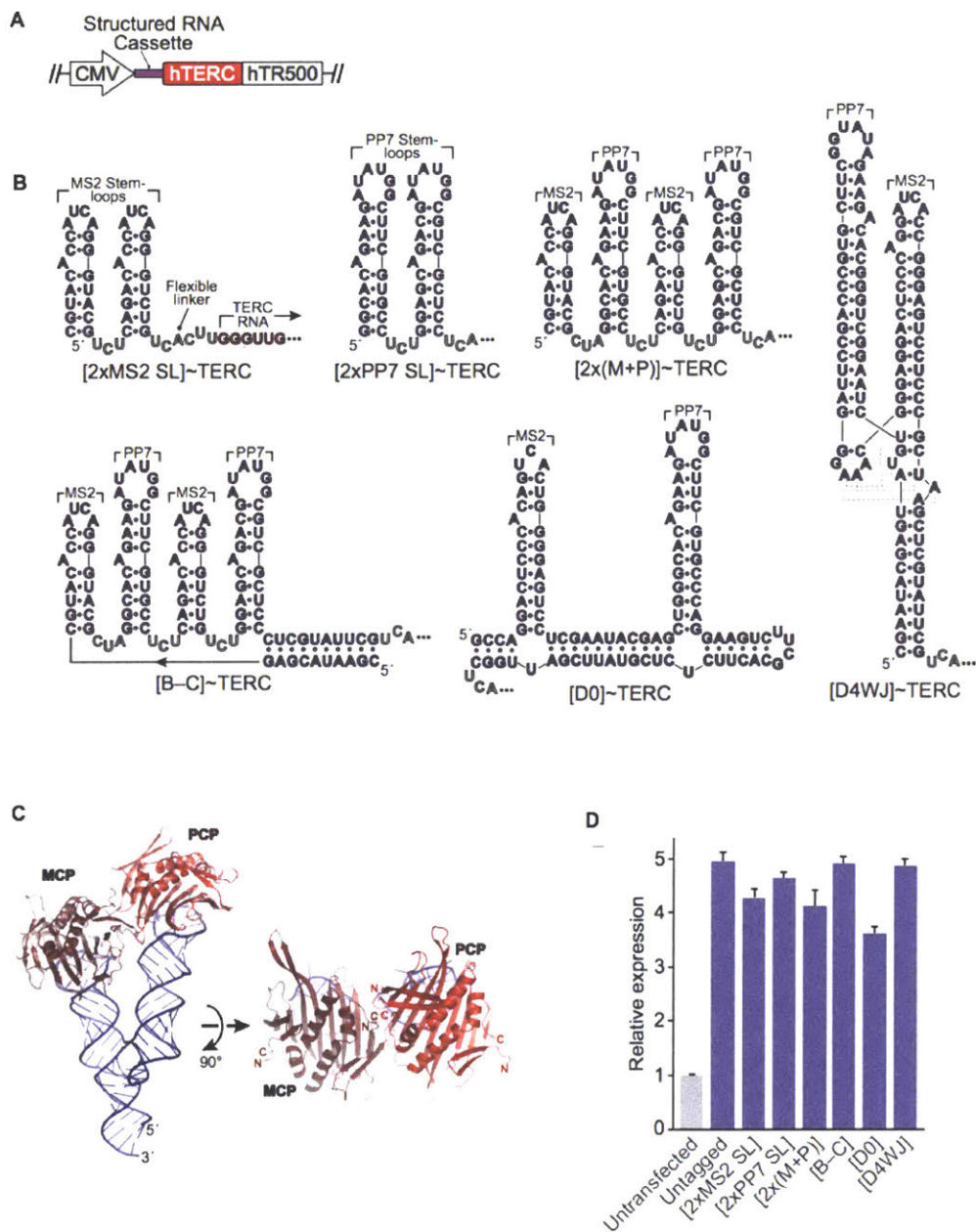


Figure 6-3. Noncoding RNA construct design. (A) General schematic of the tagged TERC RNA expression system. A cytomagalovirus promoter (CMV) drives expression of human telomerase RNA (hTERC, red) appended at its 5' terminus with a structured cassette (purple). TERC 3'-end processing is mediated by the hTR500 block (gray), corresponding to the 500 bp of native genomic sequence downstream of the TERC 3'-terminus (PMID: 12769858). (B) Predicted secondary structures of the RNA cassettes tested. Linker residues denoted in gray; the 3'-terminal UCAUUU linker is present on all constructs. Red nucleotides indicate the 5'-end of the TERC ncRNA. Cassettes contain the following elements: [2xMS2 SL], two MS2 stem-loops; [2x PP7 SL], two PP7 stem-loops; [2x(M+P)], two pairs of alternating MS2 and PP7 stem loops; [B-C], a "bracketed

cassette,” in which the [2x(M+P)] construct is enclosed within an additional helix; [D0], the “D0” RNA scaffold, previously demonstrated to co-localize MS2– and PP7–fusion proteins *in vivo* (PMID: 21700839); [D4WJ], a “docked 4-way junction” cassette, derived from “tecto-RNA” constructs designed to promote parallel co-proximity of two RNA helices (PMID: 20876687). Gray dotted lines denote tertiary contacts that mediate the interhelical docking interaction (PMID: 8781224). **(D)** Three-dimensional model of the D4WJ construct, bound to MCP and PCP (modeled on PDB IDs 2BU1, 2QUX, and 1GID). Orthogonal views are shown. Note the close proximity between the coat promoters’ N– and C–termini (“N” and “C”) near the protein–protein interface. Modeling was performed in COOT (PMID: 20383002); figure was generated using Pymol (Schrödinger, LLC). **(D)** Quantitative RT–PCR expression analysis of TERC constructs transiently expressed in HEK 293T cells. A common pair of primers that target the TERC RNA core was used to survey all constructs. Relative expression (scaled to untransfected control cells, which naturally express TERC) was quantified via the $\Delta\Delta C_T$ method, normalized to GAPDH. Data correspond to the mean of four technical replicates. Error bars: standard deviation.

However, under the same assay conditions, expression of MCP–AP and PCP–EX resulted in biotinylation that was dependent on the presence of a TERC RNA bearing both of the cognate MS2 and PP7 motifs (**Figure 6-2C**). Critically, we observed no such biotinylation when we replaced this TERC variant with a pair of constructs that each bore one of the two cognate sites. Moreover, while the extent of biotinylation by sAPEX is significantly lower (as compared to MCP–APEX2) when using the doubly-tagged TERC construct, the labeling pattern appears more punctate, as would be predicted from localizing APEX2 to the discrete subnuclear foci characteristic of interphase telomerase¹¹⁴. To scale up the experiment and test the extent of biotinylation of split APEX compared to full length APEX2 by western blot analysis, HEK 293T cells were transfected to express either full length APEX2 protein (**Figure 6-1A**) or split APEX fusion proteins (**Figure 6-2A**). hTERC RNA with or without MS2 and PP7 stem loop motifs were introduced by transfection. In the presence of both stem loop motifs, biotinylation by reconstituted split APEX is detected by the streptavidin-HRP blot (**Figure 6-2D**). Reproducibly in this scaled up biotin-phenol labeling experiment, biotinylation was not observed when hTERC was transfected without both the MS2 and PP7 recognition motif.

When comparing the extent of biotinylation in HEK 293T cells transfected to express split APEX and full-length APEX2, it is clear that the overall signal is much weaker in the case of split APEX. However, collectively, the data suggests that sAPEX activity can be specifically reconstituted on a target RNA by nucleating protein fragment assembly on a structured RNA cassette.

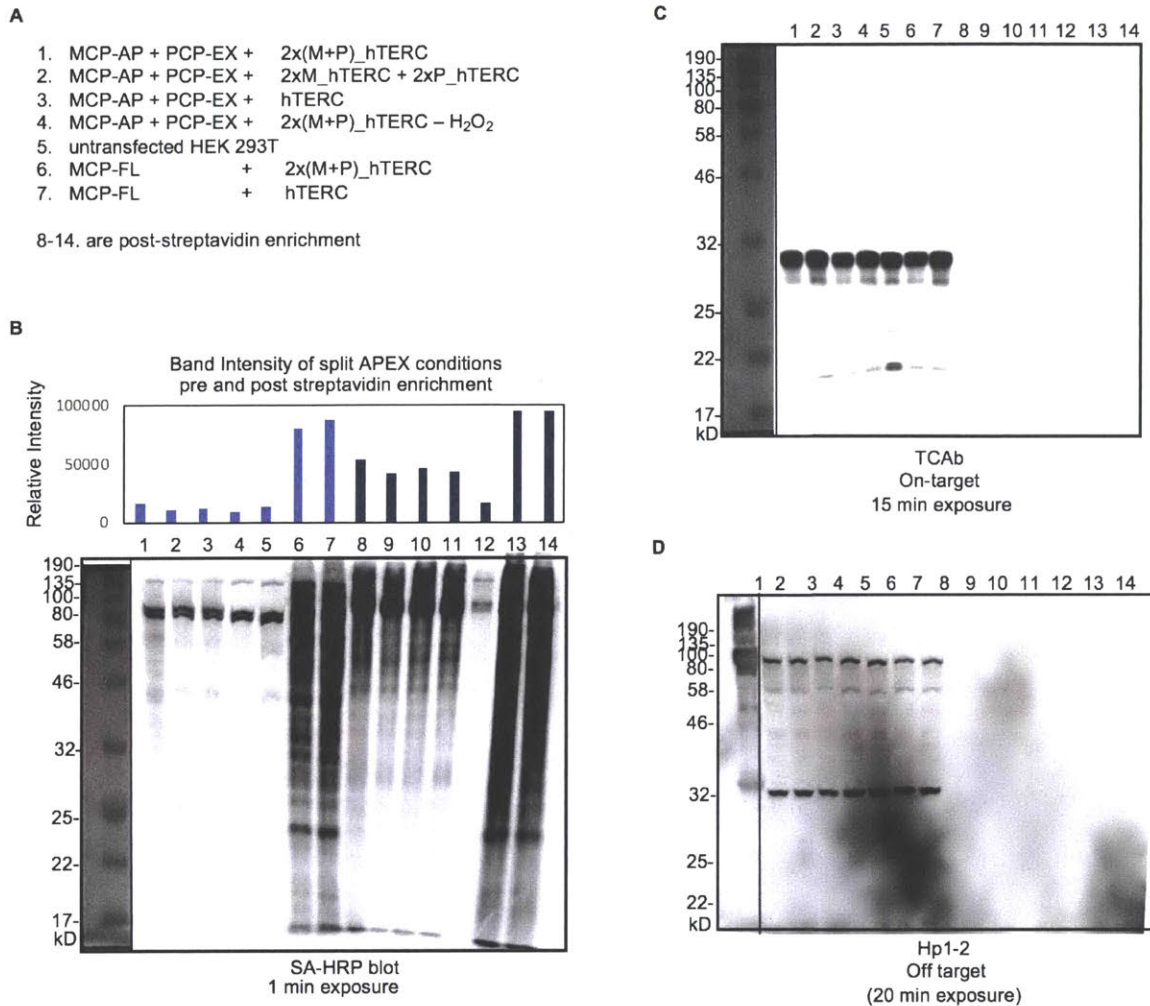


Figure 6-4. Streptavidin-enrichment of biotin-phenol labeled HEK 293T cells expressing split APEX. (A) Description of the plasmids transfected for each lane. HEK 293T cells were plated in T-25 flasks. At around 75% confluency, cells were transfected to express split APEX or full-length APEX2 and some combination of plasmids to express the hTERC RNA with and without MS2 and PP7 binding motifs. Lanes 1-7 are the original whole cell lysate; lanes 8-14 are the elutes after streptavidin-enrichment (B) Whole cell lysates and enriched eluates analyzed by 12% SDS-PAGE gel and transferred for western blot analysis. SA-HRP was utilized to detect the extent of biotinylation. Bar graph shown above the western blot utilized ImageJ to quantify the band intensities for each condition. (C) Telomerase Cajal body protein 1 (TCAb-1) mediates the accumulation of telomerase (in which hTERC is the the RNA subunit that provides the template for the catalytic hTERT) in the nucleus¹¹⁵. It is a known RBP to hTERC and was used as positive standard for sufficient biotinylation. TCAb-1 has a molecular weight of 59 kDA. (D) Heterochromatin protein 1 (Hp1-2) is a nuclear protein with no annotation of hTERC interaction was utilized as an off-target protein to test for specificity of the biotinylation. Molecular weight of Hp1-2 is a 23 kDA.

Following up on the results of detectable biotinylation by transfection of both protein and RNA components into HEK 293T cells, we scaled up our experiment to try streptavidin-bead enrichment (**Figure 6-4A;B**). Enrichment of labeled material would enable us to analyze the extent and specificity of the biotinylation in each condition. While the streptavidin-enrichment did increase the overall amount of biotinylated material run in each western blot lane (as quantified by the developed streptavidin-HRP blot intensity), the difference in the extent of labeling between the different split APEX conditions was barely discernible. Additionally, full length APEX2 has much larger extent of biotinylation; however, post-enrichment blots have shown that there is just insufficient labeled material to detect either the on or the off-target RNA binding proteins for both full length and split APEX in all conditions. The large extent of labeling by full length and the lack of detectable on and off-target nuclear proteins is very surprising. The visualized bands are not at the expected weights of either the on and off-target proteins (**Figure 6-4C;D**). This is likely a result of just non-specific binding and over development of the blots. Thus, there is much need for optimization with regards to expression and generating robust signal reconstitution, as well as validation of the labeling assay itself.

One of the biggest impediments with this current application is the expression of both fragments with little degradation. As noted previously from expression in the cytosol, the small fragment, EX, is prone to degradation, and while we previously thought that this might be an advantage – in that, excess EX fragments that were not stabilized via reconstitution with its complementary AP fragment would be degraded to further reduce any background signal. However, it may take time for each fragment to come together on the RNA scaffold, and protein degradation could occur prior to that more stable form.

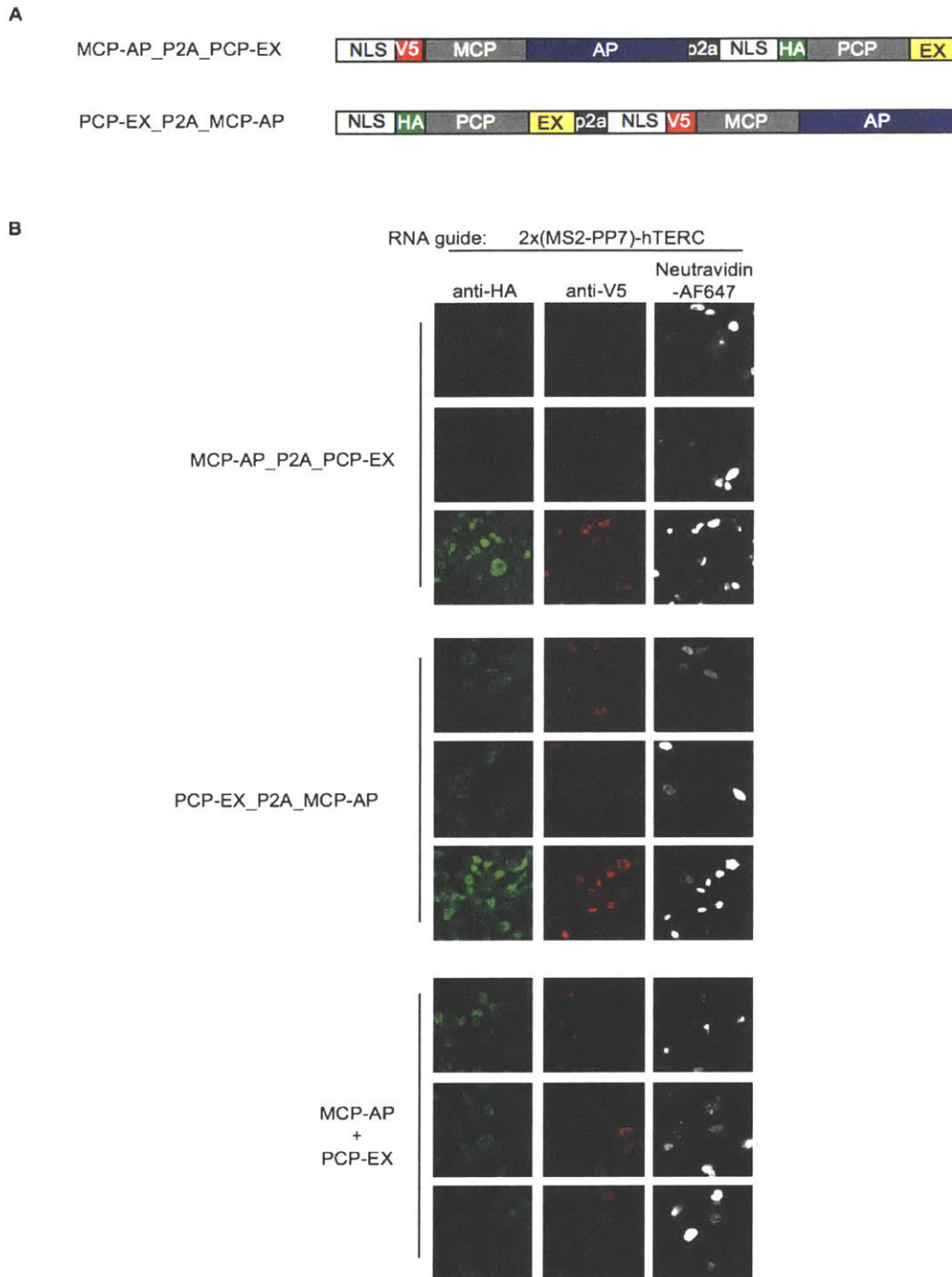


Figure 6-5. Testing Split APEX2 P2A constructs. A) Constructs cloned - expression of split protein fusions MCP-AP and PCP-EX were combined into a single plasmid using a P2A self-cleaving peptide sequence. The plasmid, PLX304, can be utilized to generate lentivirus as well as provide blasticidin drug resistance. Two orientations were tested. (B) Stable HEK 293T cells were selected using infection by lentivirus to express the P2A constructs from (A) using low expression promoter hUBC. HEK 293T cells stably expressing PCP-EX under the high expression promoter

PKG was also generated. All stable cell lines were plated and then transfected with the plasmid to express RNA, 2x(MS2-PP7)-hTERC. Stable PCP-EX cells were also transfected to express MCP-AP under hUBC promoter. After 24 hours of protein expression, biotin-phenol labeling was performed on live cells. Afterwards, cells were fixed, immunostained, and imaged by confocal microscopy.

One of the first things we tried optimizing with regards to expression, was using P2A¹¹⁶ to link MCP-AP and PCP-EX expression as separate proteins into one plasmid (**Figure 6-5A**). We thought this would ensure that transfected cells would express both fragments at equal stoichiometries. Furthermore, stable cell lines selection would lead to cells with the expression both fragments. In parallel, due to the known stability issues of the smaller EX fragment, fusion protein PCP-EX was cloned into a plasmid that had a much stronger promoter, to increase the overall expression level (PKG). The PLX208 vector was utilized as it gave a drug resistance to Hygromycin, allowing for the future possibility of selecting for double stables with the other drug marker, blasticidin.

However, immunostaining and imaging by confocal microscopy revealed that despite the cell lines having undergone stable cell selection, there was still significant expression and extent of reconstitution variation. In all three tested stable cell lines, imaging revealed many fields of view in which there were cells that had very robust labeling even with nearly undetectable levels of EX fragments. Overall, activity generally correlated to AP expression. Surprisingly, there were even instances though that the BP labeling pattern was not reflective of the overlap between the EX and AP fragment immunostaining (**Figure 6-5B**). Additional assays were performed to test the effects of heme supplementation. While it appeared to yield benefits to overall signal but the non-specific biotinylation in cells without robust expression of both split APEX fragments was still an observed phenomenon. Overall, the P2A constructs demonstrated too much variations in expression, and the single fragment PCP-EX stable cells was deemed more promising.

With the initial test of stable cells expressing PCP-EX transfected with MCP-AP with the 3 different scaffold RNA conditions looking promising (**Figure 6-6A**), selections to make a double stable HEK 293T cells were underway utilizing the double drug selection. The hygromycin resistance of the PCP-EX cells was combined by infection of those cells to express MCP-AP and provide blasticidin resistance (**Figure 6-6B**).

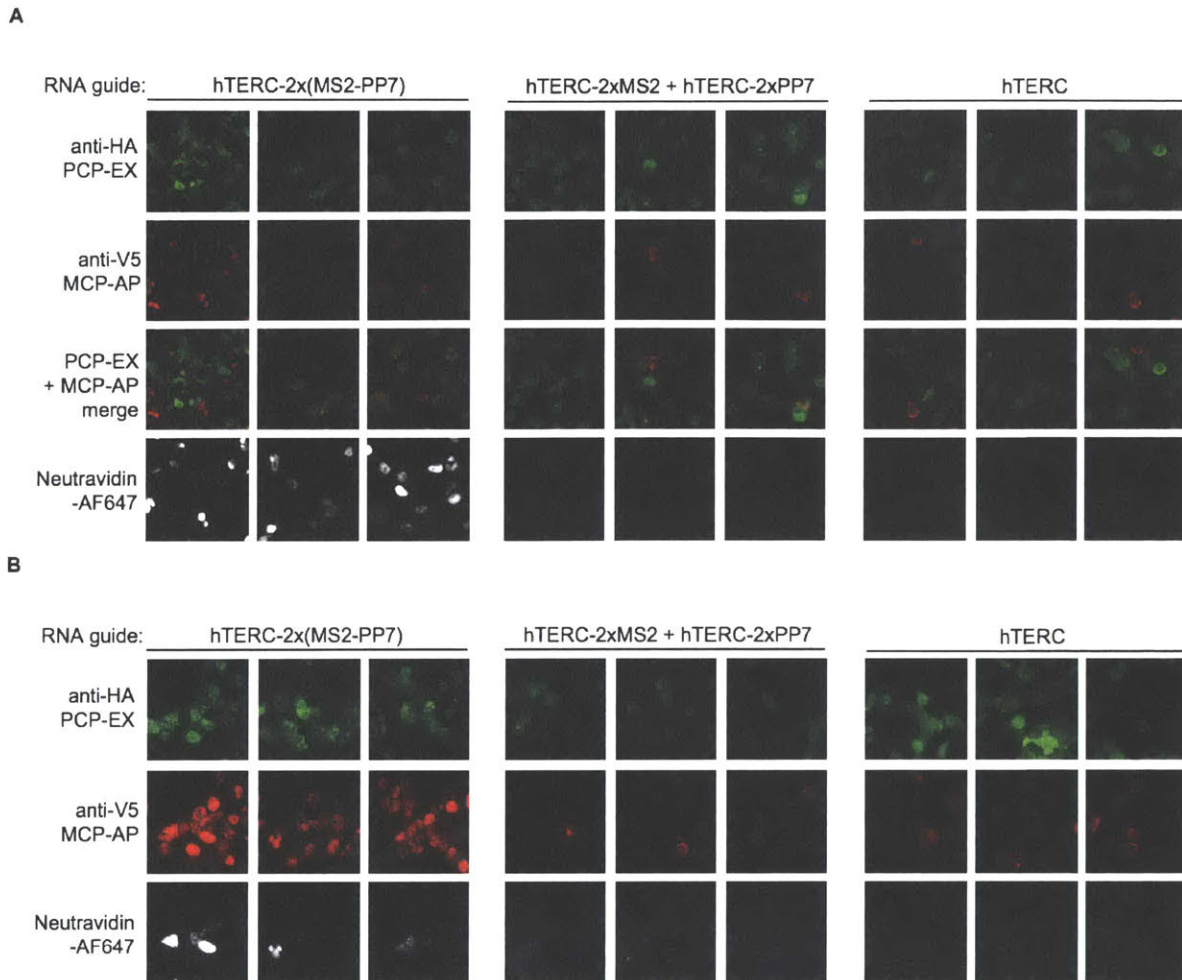


Figure 6-6. Testing HEK 293T cells stably expressing PCP-EX against double stable cells expressing PCP-EX and MCP-AP. (A) HEK 293T cells stably expressing PKG promoted PCP-EX with hygromycin resistance were plated in 49-well plates. After overnight growth, cells were transfected to express hUBC promoted MCP-AP and the indicated RNA guide. After a day of protein expression, cells were biotin-phenol labeled live, fixed, immunostained, and imaged by confocal microscopy. Three fields of view are shown for each condition, magnification 40x. (B) In parallel with the experiment in (A), HEK 293T cells that were selected to stably express PKG promoted PCP-EX (hygromycin resistance) and hUBC promoted MCP-AP (blasticidin resistance) were plated and subsequently transfected to express the indicated RNA guides. Three fields of view, magnification at 63x.

Unfortunately, the double stable cells did not improve overall signal - the total reconstitution was very low across the cells. We observed that in the double stable cells, despite

individual selection of each fusion split APEX fragment to guarantee expression of both fragments in each cell, there is still a large discrepancy in the expression levels from cell – this is highly dependent on the presence of the template RNA scaffold that brings the two fragments together. Without it, the expression level becomes increasingly poor for both fragments (**Figure 6-6B** middle and right columns). In both tested expression platforms, biotinylation was still only detectable in the presence of the RNA scaffold that drives reconstitution.

Discussion

Given the purportedly broad array of novel RNAs and RNA-mediated processes that have eluded mechanistic dissection¹⁰⁴ and the growing number of diseases now thought to be mediated by aberrant RNA-protein interactions^{105,106}, there is a great need for identifying the interaction partners of specific RNAs of interest. Conventional approaches address this problem by capturing and enriching a target RNA, either using affinity-tagged antisense oligonucleotides that isolate endogenous transcripts^{30,117–119}, or by affinity tagging the transcript of interest itself^{120,121}. However, these approaches are often limited by the abundance of the target transcript, or are confounded by the unpredictable performance of RNA-based affinity tags¹²², and by nonspecific RNA-protein interactions formed during lysis and enrichments¹²³.

Using RNA scaffolding motifs to target the reconstitution of sAPEX eliminates high off-target background signal from protein overexpression. sAPEX demonstrates promise for BP labeling around specific RNAs, which indicates potential future applications to elucidate interactomes and interaction partners of an RNA of interest. One caveat of sAPEX is that expression of EX fragment alone is prone to degradation, especially when not co-expressed with AP; this characteristic became advantageous in further reducing off-target reconstitution, effectively further reducing excess pools of protein.

We applied sAPEX reconstitution using a RNA scaffold through the use of fusions of each fragment to an RNA-motif binding protein. While enzymatic activity via biotin-phenol labeling assay yielded detectable biotinylation only in the presence of the RNA scaffold that drives reconstitution, overall reconstitution was low and could not provide enough material for streptavidin-enrichment to analyze for specificity of labeling.

Due to the RNA motif dependent reconstitution, many efforts were made to improve the system for more robust sAPEX enzymatic activity. One major focus of these optimizations, was

to improve the overall labeling intensity. P2A and double stable cell lines were generated to attempt the guaranteed expression of both AP and EX fragments inside the nucleus of HEK 293T cells. The double stable cells may have provided a slight boost in overall expression level of both fragments - however, the overall labeling intensity was significantly decreased.

While this observation was confounding, we had also continually noticed that there was protein stability issues for the small EX fragment which may contribute to the poor reconstituted activity. Efforts have recently been dedicated to improving the stability of the small fragment by changing orientations, linkers, effective copy numbers. From preliminary results, it appears that certain orientations in which the EX fragment is not C-terminal does improve fragment expression. Stable cells expressing EX-PCP-EX-NLS subsequently infected with MCP-AP provided decent robust expression of the EX fragment and BP labeling that resembled the overlap of the cells that express both fragments. However, the general overall expression level still requires further optimization. In conclusion, while there are many areas in which the sAPEX tool could be improved and optimized for each application, it has demonstrated co-proximity dependent reconstitution.

Methods

See Table 6 for plasmids used in this chapter

Cloning

See Supplementary Table 1 for a list of genetic constructs used in this study, with annotated epitope tag, promoters, resistance, vector, linkers, etc. For cloning the constructs, PCR fragments were amplified using Q5 polymerase (New England BioLabs (NEB)) or PfuUltra II Fusion HS DNA polymerase (Agilent Technologies). The vectors were double-digested and ligated to gel-purified PCR products by T4 ligation or Gibson assembly. Ligated plasmid products were introduced by heat shock transformation into competent XL1-Blue bacteria. The APEX2 gene used for initial cut site screening was amplified from vimentin-APEX2 with codons optimized for mammalian expression. Mutants of AP were generated either using QuikChange mutagenesis (Stratagene) or isolated from individual yeast clones and transferred to simammalian expression vectors using standard cloning techniques.

To generate tagged RNA expression constructs (**Figure 4D–G; Figure S8**), the TERC ncRNA and the hTR500 3′-processing block (PMID: 12769858) were amplified as a single product from HEK 293T genomic DNA, via nested PCR. DNA fragments encoding the structured RNA cassettes (**Figure S8**) were synthesized by primer extension from overlapping oligonucleotides, or as gBlocks (Integrated DNA Technologies, Inc.). TERC and cassette inserts were ligated into the ncRNA transient expression vector pCMV–SV40pA (PMID: 26030444), from which the existing polyadenylation site had been removed, via standard two- or three-piece restriction-digest cloning.

MCP- and PCP-tagged constructs (**Figure 4D–G; Figure S8**) were generated as follows. Genes encoding single protomers of the phage coat proteins were amplified from Aplasmids pHA_MS2-VP64 and pHA_PP7-VP64 (PMID: 26030444). Upstream fragments encoding the Kozak sequence, Nuclear Localization Sequence (NLS), and V5 (MCP) or Flag (PCP) tags were synthesized by primer extension from overlapping oligos (IDT, Inc.). Full length APEX2 was amplified from mito-V5-APEX2 (PMID:72480) Three-fragment pools (upstream fragments, phage coat protein, APEX2) were restriction digested and ligated with T4 DNA ligase (New England Biolabs). This ligation mixture was then used as template for a subsequent round of PCR, to isolate the full chimeric insert. These full-length PCR products were then inserted into pCMV–SV40pA by standard restriction-digest cloning. For cloning MCP-AP and PCP-EX, similar methods were employed.

Biotin-phenol labeling

Genes were introduced into HEK 293T cells through either transient transfection with Lipofectamine 2000 or lentiviral infection. After 18–24 h (transfection) or 48 h (lentivirus), biotin-phenol (bp) labeling was initiated by changing the medium to 200 μ L of fresh growth medium containing 500 μ M bp that was sonicated for at least 5 minutes to ensure bp was fully dissolved. Cells were incubated at 37 °C under 5% CO₂ for 30 min according to previously published protocols. To initiate labeling, 2 μ L of 100 mM H₂O₂ was spiked into each well, for a final concentration of 1 mM H₂O₂, and the plate was immediately gently agitated. To quench the reaction after 1 minute, the bp solution is aspirated, and cells were immediately fixed with 4% formaldehyde with 5 mM Trolox in PBS at room temperature for 5 minutes before continuing

fixation on ice for an additional 25 minutes. Cells were then washed with chilled DPBS three times and permeabilized with pre-chilled methanol at -20°C for 10 min. Cells were washed again three times with DPBS and blocked with 3% BSA in DPBS for 1 h to overnight with rocking at 4°C . To detect the expression of sAPEX fusions, cells were incubated with primary antibodies mouse anti-V5 (Life Technologies, Cat. No. R96025, 1:1500 dilution) and rabbit anti-HA (Cell Signaling, 1:1000 dilution) in 1% BSA in DPBS for 1 h to overnight at 4°C followed by 4×5 min washes with chilled DPBS. Cells were then incubated with a 1% BSA in DPBS solution containing secondary Alexa Fluor 488-goat anti-rabbit IgG (Life Technologies, Cat. No. A-11001, 1:1000 dilution), Alexa Fluor 568-goat anti-mouse IgG (Life Technologies, Cat. No. A-11004, 1:1000 dilution), and homemade streptavidin–Alexa Fluor 647 (1:1000) for 25-45min at 4°C with rocking. Cells were then washed 4 times for 5 min with chilled DPBS and imaged by confocal microscopy.

Streptavidin blotting

For streptavidin blotting, cells were grown on polystyrene 6-well plates (Corning) and labeled under the same conditions described above (see “Biotin-phenol labeling”). After 1 min of labeling, the cells were washed three times with room temperature quencher solution (10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox in DPBS) and then scraped and pelleted by centrifugation at 3000g for 10 min. The cell pellet was then lysed by gentle resuspension into peroxidase-quencher containing RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, $1\times$ protease cocktail (Sigma Aldrich, catalog no. P8849), 1 mM PMSF (phenylmethylsulfonyl fluoride), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox). After 2 minutes on ice, the lysates were clarified by centrifugation at $15000 \times g$ for 10 min at 4°C before separation on 10% SDS-PAGE gel. For blotting analysis, gels were transferred to nitrocellulose membrane, stained by Ponceau S (10 min in 0.1% (w/v) Ponceau S in 5% acetic acid/water), and blocked with “blocking buffer” (3% (w/v) BSA and 0.1% Tween-20 in Tris-buffered saline) for 1 h at room temp or 4°C overnight. The blots were immersed in streptavidin-HRP in blocking buffer (Thermo Scientific, cat. no. 21126, 1:3000 dilution) at room temperature for 1 h and then rinsed with blocking buffer 4 times for 5 min. For assessing comparative fragment expression level, identical gels and blots were prepared in parallel and immersed in blocking buffer containing either mouse anti–V5 (Life Technologies, R96025, 1:3000

dilution) or mouse anti-HA (Santa Cruz Biotechnology, G1817, 1:3000 dilution) for 1 h to overnight at 4°C. The blots were rinsed with blocking buffer 4 times for 5 minutes before being immersed in anti-Mouse-HRP (Bio-Rad, 1:3000) in blocking buffer at room temp for 1 h, rinsed with blocking buffer 4 times for 5 min again. All blots were developed using Clarity reagent (Bio-Rad) for 5 min and imaged using an Alpha Innotech gel imaging system.

Microscopy

Confocal imaging was performed on a Zeiss AxioObserver inverted confocal microscope with 10× air and 40× oil-immersion objectives, outfitted with a Yokogawa spinning disk confocal head, a Quad-band notch dichroic mirror (405/488/568/647), and 405 (diode), 491 (DPSS), 561 (DPSS) and 640-nm (diode) lasers (all 50 mW). All images were acquired and processed using Slidebook 5.0 or 6.0 software (Intelligent Imaging Innovations), through a 48× or 63× oil-immersion objective CFP(405 laser excitation; 635/85 emission), Alexa Fluor 488 (491 laser excitation; 528/38 emission), Alexa Fluor 568 (561 laser excitation; 617/73 emission), Resorufin (491 laser excitation; 550/585), Alexa Fluor 647 (647 excitation; 680/30 emission), and differential interference contrast (DIC). Acquisition times ranged from 100 to 1000 ms. DAB labeled cells were imaged by bright field; acquisition time was 200 ms. Imaging conditions and intensity scales were matched for each data set presented together.

RNA structural modeling

The Docked Fourway Junction construct (“D4WJ”, **Figure S8B**) was designed based on analogous teco-RNA devices (PMID: 20876687), supplemented by three-dimensional modeling as performed in COOT (PMID: 20383002). Briefly, parallel helices of the tetraloop–tetraloop-receptor interaction in the P4–P6 domain of the *T. thermophila* Group I intron (PMID: 8781224, PDB: 1GID) were used as structural templates on which regular A–form RNA duplexes (generated automatically in COOT) were aligned. These duplexes then served as templates on which the structures of RNA-bound MCP and PCP proteins (PDB IDs: 2BU1, 2QUX, respectively, PMIDs 11720290, 18066080) were aligned. The register of alignment (*i.e.*, the number of base pairs between the MS2 and PP7 loops and the core four-way junction) were systematically altered so as to co-proximate the N- and C-termini of the coat protomers, while minimizing the apparent overall

steric clash between the proteins. Images of the final model (**Figure S8C**) were rendered in PyMol (Schrödinger, LLC).

RNA expression analysis

Quantitative RT–PCR analysis of TERC constructs (**Figure S8D**) was performed in a manner similar to that described previously (PMID: 26030444). Briefly, HEK 293T cells were grown in 6-well dishes, as above, and at 80% confluency were transiently transfected with each of the TERC expression constructs, using lipofectamine 2000 (Thermo Fisher Scientific, LLC). Two days thereafter, cell media was aspirated, cells were washed once briefly with warmed PBS, and total cellular RNA was harvested by extraction with Trizol (Thermo Fisher) followed by precipitation with isopropanol, using Glyco Blue (Thermo Fisher) as a carrier. Pellets were resuspended in water and further purified using RNEasy mini-prep columns (QIAGEN), following the manufacturer’s “RNA cleanup” protocol with on-column DNase treatment. 500 ng of purified RNA was used as template for reverse transcription, using Super Script III Reverse Transcriptase (Thermo Fisher), primed with random hexamers (Thermo Fisher). Transcripts were quantified in quadruplicate on a 7900HT Fast Real-Time PCR System (Applied Biosystems) using Rox-normalized Universal SYBR Green Master Mix (Roche) and the following primers:

GAPDH_F: TTCGACAGTCAGCCGCATCTTCTT

GAPDH_R: GCCCAATACGACCAAATCCGTTGA

TERC_F: CGCCTTCCACCGTTCATTC

TERC_R: GGCCAGCAGCTGACATTTT

Data were analyzed using Realtime qPCR Miner (PMID: 16241897), as described previously (PMID: 26030444).

Table 6. Plasmids used in Chapter 6				
Plasmid name	plasmid vector	promoter	Expression in	Features
P128	PLX304	hUBC	lentivirus	NLS-V5-MS2 coating protein-12 aa linker-AP
P129	PLX304	hUBC	lentivirus	NLS-HA-PP7 coating protein-12 aa linker-EX
P130	PLX304	hUBC	lentivirus	NLS-V5-MS2 coating protein-12 aa linker-APEX2
P131	pNEB193	pCAG	mammalian	hTERC MS2 PP7 MS2 PP7
P132	pNEB193	pCAG	mammalian	hTERC MS2 MS2
P133	pNEB193	pCAG	mammalian	hTERC PP7 PP7
P134	pNEB193	pCAG	mammalian	hTERC
P135	PLX208	PKG	lentivirus	NLS-HA-PP7 coating protein-12 aa linker-EX
P136	PLX304	hUBC	lentivirus	MCP-AP-P2A-PCP-EX
P137	PLX304	hUBC	lentivirus	PCP-EX-P2a-MCP-AP

References

- Hendrickson, D., Kelley, D. R., Tenen, D., Bernstein, B. & Rinn, J. L. Widespread RNA binding by chromatin-associated proteins. *Genome Biol.* **17**, (2016).
- Sibley, C. R. Individual nucleotide resolution UV cross-linking and immunoprecipitation (iCLIP) to determine protein–RNA interactions. in *Methods in Molecular Biology* **1649**, 427–454 (2018).
- Garzia, A., Morozov, P., Sajek, M., Meyer, C. & Tuschl, T. PAR-CLIP for discovering target sites of RNA-binding proteins. in *Methods in Molecular Biology* **1720**, 55–75 (2018).
- Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).
- Rhee, H.-W. *et al.* Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science (80-.)*. **339**, 1328–1331 (2013).
- Kopp, F. & Mendell, J. T. Leading Edge Review Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* **172**, 393–407 (2018).
- Pereira, B., Billaud, M. & Almeida, R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer* **3**, 506–528 (2017).
- Osborne, R. J. R. J. & Thornton, C. A. RNA-dominant diseases. *Human Molecular Genetics* **15**, (2006).
- Grabow, W. & Jaeger, L. RNA modularity for synthetic biology. *F1000Prime Rep.* **5**, (2013).
- Blackburn, E. H., Greider, C. W. & Szostak, J. W. Telomeres and telomerase: The path from maize, Tetrahymena and yeast to human cancer and aging. *Nature Medicine* **12**, 1133–1138 (2006).

11. Zappulla, D. C. & Cech, T. R. RNA as a flexible scaffold for proteins: Yeast telomerase and beyond. in *Cold Spring Harbor Symposia on Quantitative Biology* **71**, 217–224 (2006).
12. Fu, D. & Collins, K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Mol. Cell* **11**, 1361–1372 (2003).
13. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA. *Biochemistry* **34**, 10600–10606 (1995).
14. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
15. Lim, F. & Peabody, D. S. RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* **30**, 4138–44 (2002).
16. Theimer, C. A. *et al.* Structural and Functional Characterization of Human Telomerase RNA Processing and Cajal Body Localization Signals. *Mol. Cell* **27**, 869–881 (2007).
17. Yuan, P. *et al.* Telomerase Cajal body protein 1 depletion inhibits telomerase trafficking to telomeres and induces G1cell cycle arrest in A549 cells. *Oncol. Lett.* (2014). doi:10.3892/ol.2014.2306
18. Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. *Sci. Rep.* (2015). doi:10.1038/srep16273
19. Lee, N., Moss, W. N., Yario, T. A. & Steitz, J. A. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. *Cell* **160**, 607–618 (2015).
20. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
21. McHugh, C. A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
22. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science (80-.).* **349**, 1DUIMMY (2015).
23. Walker, S. C., Scott, F. H., Srisawat, C. & Engelke, D. R. RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes. *Methods Mol. Biol.* **488**, 23–40 (2008).
24. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W. J. Identifying proteins that bind to specific RNAs - Focus on simple repeat expansion diseases. *Nucleic Acids Research* **44**, 9050–9070 (2016).
25. Walker, S. C., Good, P. D., Gipson, T. A. & Engelke, D. R. The dual use of RNA aptamer sequences for affinity purification and localization studies of RNAs and RNA-protein complexes. *Methods Mol. Biol.* **714**, 423–444 (2011).
26. MILI, S. Evidence for reassociation of RNA-binding proteins after cell lysis: Implications for the interpretation of immunoprecipitation analyses. *RNA* **10**, 1692–1694 (2004).