Title
Enhancing Bioproduction with Enzyme, Compartment, and Pathway Engineering

Author
Russ, Zachary Nicholas

Publication Date
2016

Enhancing Bioproduction with Enzyme, Compartment, and Pathway Engineering

by

Zachary Nicholas Russ IV

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John E. Dueber, Chair
Professor Adam P. Arkin
Professor Tanja Kortemme
Professor David F. Savage

Summer 2016

Enhancing Bioproduction with Enzyme, Compartment, and Pathway Engineering

Abstract

Enhancing Bioproduction with Enzyme, Compartment, and Pathway Engineering

by

Zachary Nicholas Russ IV
Joint Doctor of Philosophy
with University of California, San Francisco
in Bioengineering
University of California, Berkeley
Professor John E. Dueber, Chair

Microbial fermentations are often the most efficient and environmentally friendly option for producing valuable chemicals. However, producing the correct molecule without losing material to side products and spontaneous degradation can prove challenging, especially working within the constraints of a living organism. To address this problem, we present three separate strategies for directing flux and controlling reactions, each operating on a different pathway. We apply enzyme-level engineering to enhance the benzylisoquinoline alkaloid (BIA) pathway, where a cytochrome P450 is responsible for both product formation and degradation. Next, we demonstrate compartment engineering by sequestering prodeoxyviolacein (PDV) pathway enzymes in the peroxisome to control the flow of intermediates and reduce side products. Finally, we rearrange pathways themselves by using biological protecting groups to protect reactive and unstable molecules, producing stable indican from unstable indoxyl in the indigo pathway. These three strategies – enzyme, compartment, and protecting group engineering – can be applied individually or together to enhance microbial yields of valuable products while minimizing byproducts.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank **my parents** first and foremost. I'm not sure how you properly thank the people who have given you everything, but there it is.

I appreciate the wonderful professors who took a risk on me and let me work in their labs. You mentored me in how to conduct experiments, manage and report results, and generally learn scientific thinking: **Joonil Seog, William Bentley, Laura Segatori, J. Christopher Anderson, Minkui Luo, Arash Komeili**, and especially **John Dueber**. I also appreciate all I've learned from **Mrs. Hill, Mrs. Norris-Krevor, Mrs. Weston, Mr. Flowers, Steve Rokita, and Philip DeShong**. They taught me very different things, but they have one thing in common – they are exemplary teachers.

Thanks also are due to the senior grad students who mentored me: **Josh Kittleson, Weston Whitaker,** and **Will DeLoache**. Your remarkable skill, versatility, and saintly patience make you true role models and excellent mentors. I'm grateful for everything you've taught me.

I would be remiss to not mention **Luke Latimer**. Your relentless drive encouraged me to achieve more, and your friendship lit the darkest depths and the latest nights, which, in grad school, are many. **iGEM 2013**, **SPSB 2014**, **Anderson Lab, Arkin Lab**, and rest of the **Dueber Lab**: it has been my great pleasure to work with, mentor, and learn from you.

Finally, I'd like to thank my coauthors and collaborators – couldn't have done it without you!
An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose: **William C. DeLoache, Lauren Narcross, Andrew M. Gonzales, Vincent J.J. Martin**
Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways: **William C. DeLoache**
Employing a glucosyl biochemical protecting group for sustainable indigo dyeing (patent): **Tammy M. Hsu, Luke N. Latimer, Bernardo Cervantes, Ramya L. Prathuri, Shyam P. Bhakta, Arthur M. Fong**; also **Raul Leal, Richmond Sarpong**

Chapter 1. Introduction

1.1. The Big Picture

Organic chemistry has provided us with a wealth of products – plastics, fuels, fertilizers, medicines and more – but there is a growing need to produce these chemicals (or alternatives) in large quantities using clean, safe, and sustainable methods[1]. These clean alternatives, termed "green chemistry," have been frequently underpinned by biocatalysis and bioproduction of the chemicals in question[2]. As bioengineers and synthetic biologists, we hope to offer our own contributions by making the process of engineering organisms to produce chemicals simpler and more efficient.

When moving from organic synthesis to biosynthesis, one major challenge crops up: a loss of control. In organic synthesis, one can tune solvent, temperature, pH, and control the presence or absence of catalysts and reactants as needed to maximize yield and minimize byproducts. These tweaks are largely unavailable in live cells, where the reaction conditions are dictated by the cell's physiology. So the question, first and foremost, is: What *can* we control in bioproduction?

In this work, we examine three different strategies that nature already uses to push the boundaries of what can be achieved within a cell. In the process, we develop new dye-based assays that can be used to probe multiple pieces of the cell's metabolism, allowing us to study cell biology even as we try to control the reactions happening inside. Finally, after demonstrating this control, we explore how these individual strategies can be extended and combined to yield new options for research and development.

As an aside, I would like to mention an interesting pattern: Dyes have cropped up at the beginning of modern organic chemistry as the foundation for chemical giants (BASF originally made aniline dyes[3]) and Nobel Prizes (Adolf von Baeyer, for the synthesis of indigo[4,5]; Gerhard Domagk, for the first antibacterial drug, Prontosil, a red aniline dye[6]). Indigo also appeared at the beginning of modern biotechnology as one of Amgen's first products[7,8]. Here, we're presenting indigo, betaxanthin, and prodeoxyviolacein dyes as our research tools and possible products – so perhaps there are exciting times ahead!

1.2. Organization

The next three chapters outline complementary strategies for controlling biochemistry. In each chapter, we gain new insights into how biochemical reactions can be controlled and metabolic flux redirected by interventions including enzyme engineering, compartmentalization, and refactoring a pathway through introduction of new chemical species. These approaches are summarized in **Figure 1-1**. Each individual chapter is a manuscript, with a section appended in which extensions to the work are discussed.

Chapter 2 discusses control and modifications at the enzyme engineering level. In this published work[9], we discover a new tyrosine hydroxylase activity in a previously characterized enzyme, the cytochrome P450 CYP76AD1 from *Beta vulgaris*. This function allows CYP76AD1 to convert L-tyrosine into L-DOPA, a crucial step for linking central metabolism to the benzylisoquinoline alkaloid (BIA) pathway. This, in turn, allows yeast to produce valuable BIA medicines including morphine, and we demonstrate the production of a branch-point intermediate, (*S*)-reticuline. We probed CYP76AD1's tyrosine hydroxylase (desired) and DOPA oxidase (undesired) activities using a novel assay in which L-DOPA is converted to a fluorescent product, betaxanthin, with DOPA oxidase products converted to a colorful product, betanidin. Using the betaxanthin screen, we obtained a 7.4-fold improvement in dopamine productivity by mutating CYP76AD1. **In the followup**, we return to CYP76AD1 to look at how its reductase partner influences DOPA oxidase activity, create orthogonal partners, and compare the natural hydroxylase CYP76AD6 to engineered CYP76AD1.

Chapter 3 covers compartmentalization of enzymes to achieve control over flux. This published work[10] focuses on repurposing the peroxisome as a synthetic organelle – a synthetic organelle being a compartment that can be adapted to the needs of your chemical pathway without significant impact on the function of the cell. Such a compartment would allow you to sequester the enzymes and intermediates of your pathway away from the host cell, preventing cross-reactivity and potential host toxicity. To lay the groundwork for the synthetic organelle, we developed a prodeoxyviolacein (PDV)-based colorimetric assay for peroxisomal protein import, and used this screen to generate an optimized, modular peroxisomal targeting tag (ePTS1). We also demonstrated the first ever *in vivo* polymer exclusion assay, which we used to measure the peroxisomal membrane's permeability and confirm that the peroxisome is indeed as permeable as previously suspected. Finally, we took a simple two-enzyme pathway and demonstrated that, thanks to an impermeant intermediate, flux could be directed between an on-target product and a spontaneous byproduct depending on enzyme compartmentalization. Furthermore, we uncovered evidence of substrate channeling under specific, heavily enzyme-limited conditions. **In the followup**, we examine what the source of the peroxisomal permeability could be, test alternative peroxisomal targeting tags, and discuss use of the peroxisome in more ambitious projects.

Chapter 5 examines reworking pathways themselves to control reactivity. This as-yet-unpublished work features the use of biological protecting groups, xenobiotic metabolism-derived chemical modifications that can be easily added and removed, as a way to stabilize highly reactive products. The product in this case is indoxyl, a soluble molecule that reacts with air to form indigo, passing through a reduced intermediate, leucoindigo. The ability to release indoxyl on demand inside the fibers of cotton yarn and generate leucoindigo would enable dyeing of denim without using any of the hazardous reducing agents now used to generate leucoindigo in an industrial context. To make this stabilized indoxyl, we copied the system that indigo-producing plants naturally use – protecting the free indoxyl with a glucosyltransferase to make stable indican. To do so, it was necessary to sequence the transcriptome of *Polygonum tinctorium* and match the data with protein mass spectrometry to identify the appropriate glucosyltransferase. Having discovered the sequence of two glucosyltransferases (PtUGT1 and PtUGT2), we were able to scale up and successfully dye several swatches of cotton without using any reducing agent. **In the followup,** we look at the broader picture of biological protecting groups to see the full potential of this strategy, discuss further improvements to reductant-free indigo dyeing, and demonstrate how biological protecting groups can be used in combination with organic chemistry to unlock new syntheses.

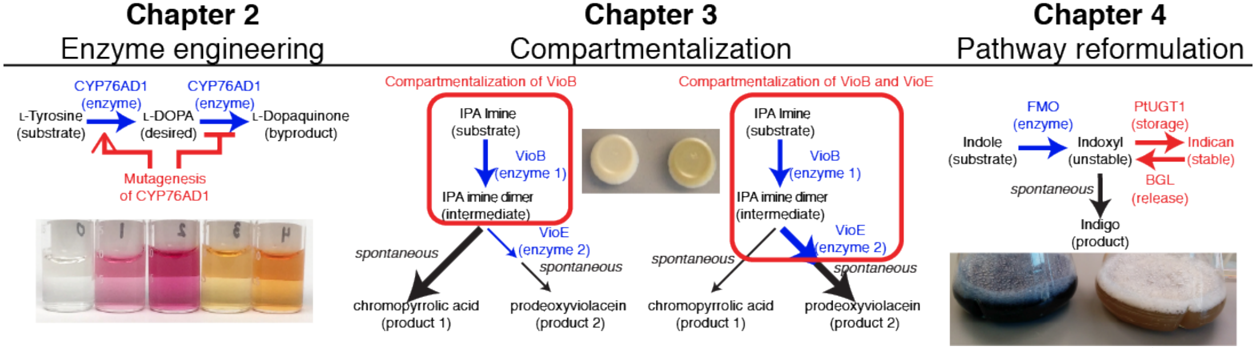


Figure 1-1. Graphical summary of the thesis. Colorimetric screens used in each work are inset.

Chapter 2. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose

2.1. Introduction

Plant specialized metabolites are a rich source of chemical diversity that has given rise to a host of prominent pharmaceuticals[11]. Despite their widespread use[12], many of these compounds are still manufactured from plant extracts because they are too structurally complex to be cost-effectively produced by total organic synthesis. Target metabolites often accumulate at low levels in plants, and efforts to improve yields are hindered by limitations in plant metabolic engineering—namely complex pathway regulation, a lack of genetic tools, and long development cycles[13]. Such problems have also made it challenging to fully explore the potential structural diversity of natural product derivatives[14].

Microbially-based production systems overcome many of these barriers and are poised to transform the manufacturing and drug discovery processes for many natural products[15]. Next-generation DNA sequencing has accelerated the elucidation of complex biochemical pathways in plants such that they can now be transplanted to easily engineered and scalable industrial production hosts[16]. Meanwhile, the development of powerful synthetic biology tools for genetically manipulating cells has made the engineering of microorganisms with new properties considerably faster, cheaper, and easier[17,18]. Commercial production of the anti-malarial drug precursor artemisinic acid in the yeast *Saccharomyces cerevisiae* is a pioneering example of these technologies revolutionizing pharmaceutical manufacturing[19].

Benzylisoquinoline alkaloids (BIAs) are a large family of L-tyrosine-derived plant specialized metabolites with a variety of therapeutic uses[20]. This class of compounds includes the opioid analgesics morphine and codeine, the antibiotics sanguinarine and berberine, the muscle relaxants (+)-tubocurarine and papaverine, and the cough suppressant noscapine. Several analgesics (e.g. oxycodone) are synthetic derivatives of the BIA intermediate thebaine[20]. Because of their structural complexity, most BIAs cannot be chemically synthesized at commercial scale and are instead extracted from plants—most notably the opium poppy (*Papaver somniferum*)[21]. In light of recent advances in our understanding of BIA biosynthesis, these molecules have become high-value targets for production via microbial fermentation.

A major achievement in microbial BIA production was marked by the synthesis of the key intermediate (*S*)-reticuline from glucose in the bacterium *Escherichia coli*[22]. (*S*)-reticuline is the final intermediate to be shared between the major branches of the BIA pathway. Despite high titers[23], no steps downstream of (*S*)-reticuline have since been demonstrated in *E. coli*. However, an impressive number of these downstream steps have been successfully reconstituted in *S. cerevisiae* leading to the recent production of many valuable BIAs, albeit from cost-prohibitive intermediates such as norlaudanosoline and thebaine[24–26]. These products include morphine, codeine, hydrocodone, hydromorphone, oxycodone, (*S*)-tetrahydroberberine, and dihydrosanguinarine. *S. cerevisiae* is amenable to the functional expression of endomembrane-localized cytochrome P450s and therefore a more suitable host for the downstream steps of BIA biosynthesis[18]. P450s play an important role in BIA synthesis and diversification, especially in the generation of new backbones, such as morphinans (CYP719B1), aporphines (CYP80G2), bisbenzylisoquinolines (CYP80B2), and benzophenanthridines (CYP82N4)[20]. While yeast have been co-cultured with reticuline-producing *E. coli* to catalyze downstream steps to magnoflorine and scoulerine[27], large-scale production of BIAs will necessitate that all steps be consolidated into a single microorganism to minimize carbon loss and simplify fermentation processes. Although there have been recent successes in establishing microbial co-cultures and sequential fermentations in the laboratory[23,28,29], industrial fermentations frequently favor monocultures over more complicated, multi-strain strategies.

Production of (*S*)-reticuline from central metabolites in *S. cerevisiae* has proven surprisingly difficult given the early successes in *E. coli*. While reticuline has been synthesized from fed norlaudanosoline in yeast[24], the upstream steps required for the synthesis of the norlaudanosoline or the natural BIA backbone (*S*)-norcoclaurine from L-tyrosine have remained elusive. This is primarily attributable to: i) difficulties at the first biosynthetic step from L-tyrosine to L-3,4-dihydroxyphenylalanine (L-DOPA) and ii) poor activity of the norcoclaurine synthase (NCS)[30]. Prior to this work, there were two known families of enzymes broadly referred to as tyrosine hydroxylases. Plants and animals utilize tyrosine 3-monooxygenases (E.C. 1.14.16.2), which require a cofactor (tetrahydrobiopterin) not found in yeast[31]. Copper-containing tyrosinases (E.C. 1.14.18.1), used by many organisms for melanin production, exhibit both tyrosine hydroxylase and DOPA oxidase activities to produce L-dopaquinone from L-tyrosine[32,33]. While bacterial tyrosinases have been used for L-DOPA production in *E. coli*[22], functional expression of these enzymes in *S. cerevisiae* remains a challenge.

We report here the development of an enzyme-coupled biosensor for the BIA pathway intermediate L-DOPA. This biosensor takes advantage of a plant biosynthetic enzyme that converts L-DOPA into the yellow, fluorescent pigment betaxanthin[34]. Using this sensor, we identified a tyrosine hydroxylase that was highly active in *S. cerevisiae* and were able to further improve its activity via PCR mutagenesis to increase L-DOPA titer 2.8-fold and dopamine titer 7.4-fold. This enzyme is a cytochrome P450 from the sugar beet *Beta vulgaris* and represents the first known example of a P450 capable of L-tyrosine hydroxylation. While the wildtype version of this enzyme catalyzes an additional unwanted oxidation of L-DOPA into L-dopaquinone, one of the beneficial mutations that we identified suppresses this off-pathway activity. Our mutant tyrosine hydroxylase enabled the production of dopamine in yeast and, when coupled to a newly identified norcoclaurine synthase from opium poppy, allowed for the synthesis of the BIA intermediates (*S*)-norcoclaurine and (*S*)-reticuline from glucose. This breakthrough connects, for the first time, the central metabolism of yeast to the downstream steps of BIA biosynthesis that have previously been demonstrated in *S. cerevisiae* and will contribute to the development of a microorganism capable of producing high-value BIAs at commercial scale.

2.2. Results

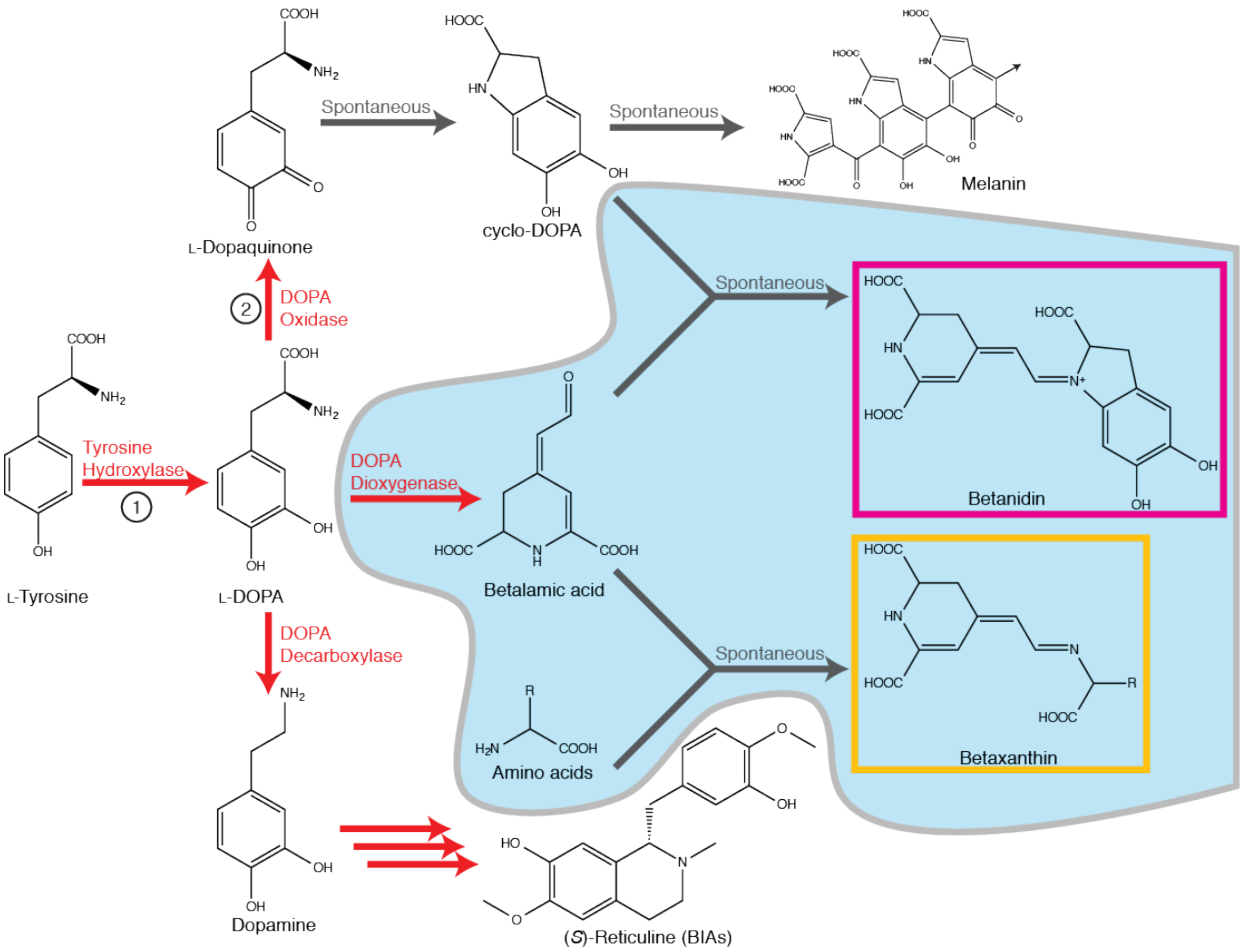


Figure 2-1. Utilizing the betaxanthin pathway as a biosensor for L-DOPA in yeast. To improve L-DOPA production in yeast cells, we expressed DOPA dioxygenase (DOD), which enabled rapid monitoring of upstream enzymatic activity through the production of plant pigments (shaded in blue). Enzymes that produce L-DOPA via tyrosine hydroxylation (activity 1) often catalyze a second step to L-dopaquinone (DOPA oxidase, activity 2). This second activity is undesirable for the production of BIAs from tyrosine because it introduces a competing pathway to melanin. When co-expressed with a tyrosine hydroxylase, DOD will generate betalamic acid, a reactive aldehyde that undergoes spontaneous condensation with amino acids and other amines to form betaxanthins, which are yellow and fluorescent. Increases in tyrosine hydroxylase activity can be easily detected by increased betaxanthin production. Tyrosine hydroxylases with a high level of DOPA oxidase activity will produce L-dopaquinone. After a spontaneous conversion to cyclo-DOPA, L-dopaquinone can undergo condensation with betalamic acid to form betanidin, a violet pigment. While betanidin is prone to oxidative polymerization in normal yeast media, the addition of ascorbic acid as a reducing agent stabilizes betanidin, allowing DOPA oxidase activity to be easily monitored. (*S*)-Reticuline is a key intermediate in the BIA pathway that represents a branch point from which the majority of BIAs can be produced. Red arrows denote enzymatic reactions; black arrows denote spontaneous reactions.

2.2.1. Development of an enzyme-coupled L-DOPA biosensor

A long-standing inability to achieve L-tyrosine hydroxylation to L-DOPA in yeast prompted us to develop an enzyme-coupled biosensor for L-DOPA that could be used to quickly screen candidate tyrosine hydroxylases for activity. DOPA dioxygenase (DOD) is a plant enzyme found in members of the order Caryophyllales that converts L-DOPA into a yellow, highly fluorescent family of pigments called betaxanthins (**Figure 2-1** and **Figure 2-2a**)[35,36]. These betaxanthins, all of which have similar optical properties, are the result of a spontaneous reaction between free amines and betalamic acid produced by DOD[37]. For simplicity, we will refer to this entire family of molecules collectively as betaxanthin.

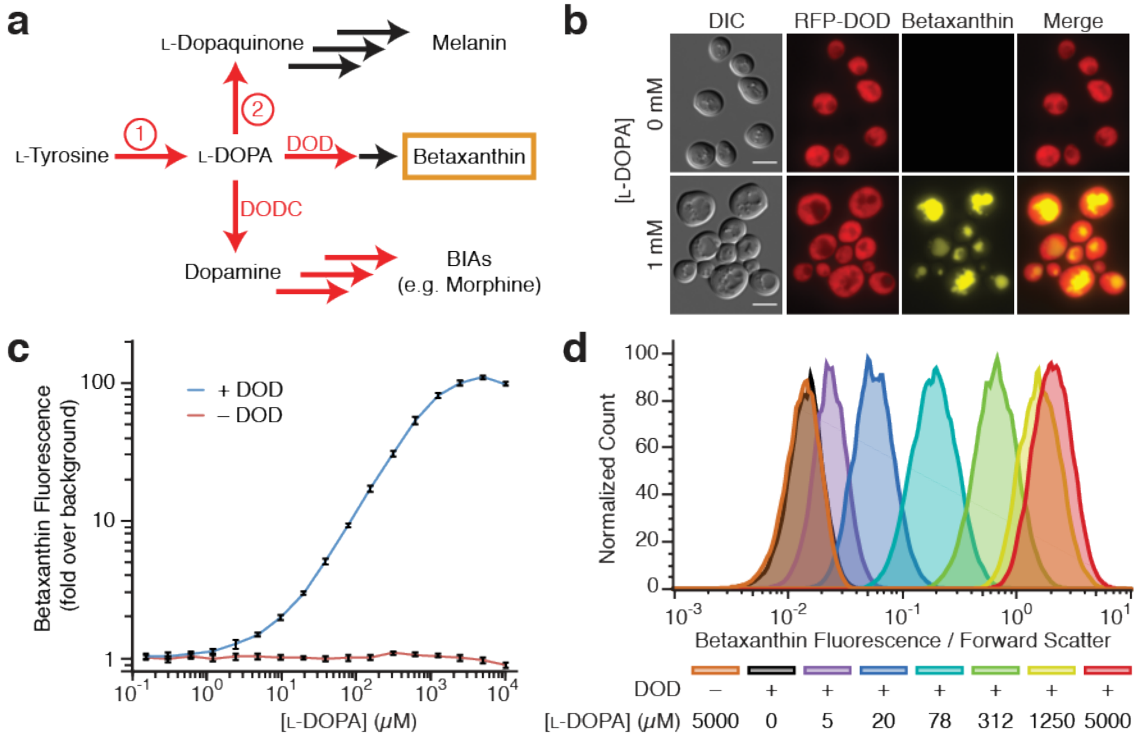


Figure 2-2. Development and characterization of an enzyme-coupled L-DOPA biosensor. (**a**) L-DOPA can be acted on by three distinct enzymes. For the production of benzylisoquinoline alkaloids (BIAs), DOPA decarboxylase (DODC) is the desired activity. However, many enzymes that produce L-DOPA (tyrosine hydroxylases, activity 1) also have a second activity (DOPA oxidase, activity 2) that introduces a competing pathway towards melanin. DOPA dioxygenase (DOD) converts L-DOPA into the fluorescent, yellow pigment betaxanthin in plants and was used in yeast as an L-DOPA biosensor to optimize tyrosine hydroxylase activity. Red arrows denote enzymatic reactions; black arrows denote spontaneous reactions. See **Figure 2-1** for a more detailed pathway diagram. (**b**) Fluorescence microscopy of yeast expressing DOD with the red fluorescent protein mKate2 fused to the N-terminus, grown in media with and without L-DOPA. Scale bars, 10 μm. (**c**) Dose-response curve for a DOD-expressing strain (blue) and a wildtype control strain (red) grown in media supplemented with a range of L-DOPA concentrations ranging from 10,000 – 0.15 μM in two-fold increments.

Cellular betaxanthin fluorescence was measured on a microplate fluorometer. Error bars are mean ± s.d. of twelve biological replicates. (**d**) Flow cytometry histograms for DOD-expressing yeast cells (and a wildtype control, orange) grown in media supplemented with a range of L-DOPA concentrations. Cells were gated by forward and side scatter as indicated in **Figure 2-4**. Betaxanthin fluorescence was normalized to forward scatter to account for variation in cell size. Cell counts were normalized as a percent of the mode of each population.

We expressed in yeast a DOD variant from the flowering plant *Mirabilis jalapa*[35] and found that cells grown in media supplemented with L-DOPA produced enough betaxanthin to be easily detected by eye in both the supernatant and the cell pellet. Fluorescence microscopy indicated that betaxanthin, which is highly water soluble, accumulates in the yeast vacuole in addition to getting pumped out of the cell through unknown mechanisms (**Figure 2-2b**). Because a substantial fraction of betaxanthin remains intracellular, L-DOPA levels can be quantified via cellular fluorescence. Using a microplate fluorometer, we generated a dose-response curve for a DOD-expressing strain grown in 0.1-10,000 μM L-DOPA (**Figure 2-2c**). Changes in betaxanthin fluorescence were detectable across a thousand-fold range, 2.5-2,500 μM L-DOPA, with the most sensitive response observed between 25-250 μM L-DOPA (**Figure 2-3**).

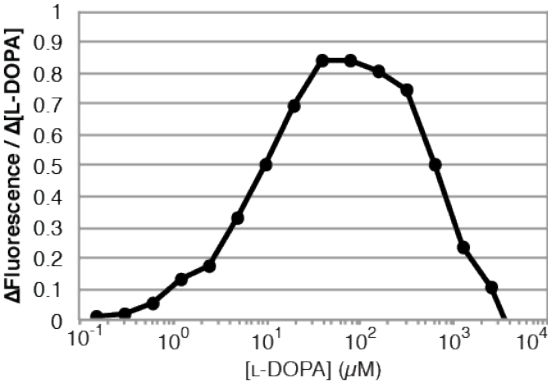


Figure 2-3. Marginal change in betaxanthin fluorescence in response to extracellular L-DOPA concentration. Betaxanthin fluorescence was measured via microplate fluorometer in a DOD-expressing strain (Strain 3) grown a range of extracellular L-DOPA concentrations (see **Figure 2-2c**). In this figure, the finite difference derivative was taken between each data point to calculate the marginal change in betaxanthin fluorescence as the L-DOPA concentration increased.

The sensor demonstrated a 110-fold dynamic range and low sample-to-sample variability, yielding a Z'-factor of 0.91, where a value between 0.5-1.0 typically denotes an excellent high throughput assay[38]. Changes in betaxanthin fluorescence were also readily detectable at the single cell level via flow cytometry (**Figure 2-2d**). Using this detection technique we achieved a 145-fold dynamic range for the sensor, however, we observed considerable cell-to-cell variability, as is common for single cell measurements, despite both gating cells by forward and side scatter and normalizing fluorescence by forward scatter to account for differences in cell size (**Figure 2-4** and **Figure 2-5**).

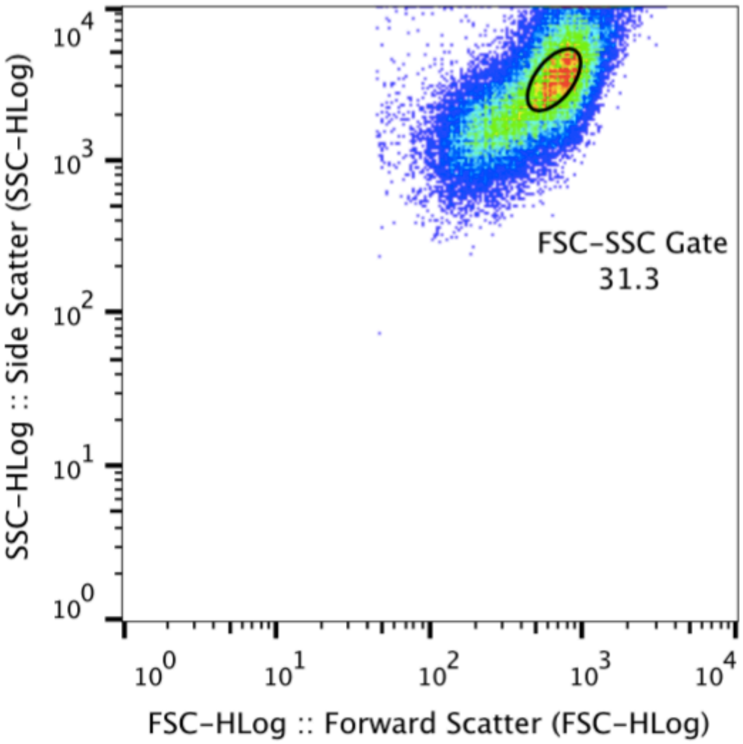


Figure 2-4. Forward- and side-scatter flow cytometry gating conditions. To remove cells with outlier morphologies including dead and conjoined cells, a subset accounting for the center third of cells (31.3%) was selected for measurement. The gated population, shown above within the black ellipse, was used for further quantification as shown in **Figure 2-2d**.

Analysis of our flow cytometry data indicated that a tyrosine hydroxylase variant with two-fold improved activity would be expected to be enriched between 25 and 50-fold over the rest of the population in a given round of fluorescence activated cell sorting (FACS) if L-DOPA production fell in the linear range of the biosensor. While the ability to cell sort is promising for enabling future screens and applications, we elected to use colony-based colorimetric and fluorometric screening of our tyrosine hydroxylase libraries, as this technique proved sufficient for our purposes (**Figure 2-6**).

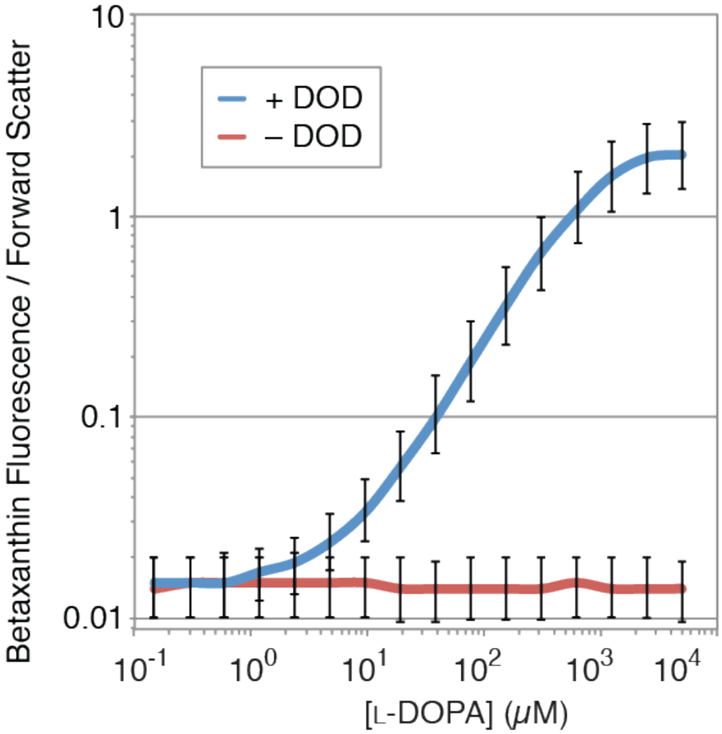


Figure 2-5. DOD-based betaxanthin biosensor responds to a thousand-fold extracellular L-DOPA concentration range as measured by flow cytometry. A DOD-expressing yeast strain (Strain 3) and a nonexpressing strain (Strain 1) were grown under conditions of varying extracellular L-DOPA as in **Figure 2-2c**. Cellular betaxanthin fluorescence was measured using a flow cytometer (**Figure 2-2d**) with gating parameters as indicated in **Figure 2-4**. The biosensor response of betaxanthin fluorescence was measured by flow cytometry and then divided by forward scatter (FSC) to account for variability in cell size. The above plot shows the median fluorescence at each L-DOPA concentration. Error bars show 5th and 95th percentiles.

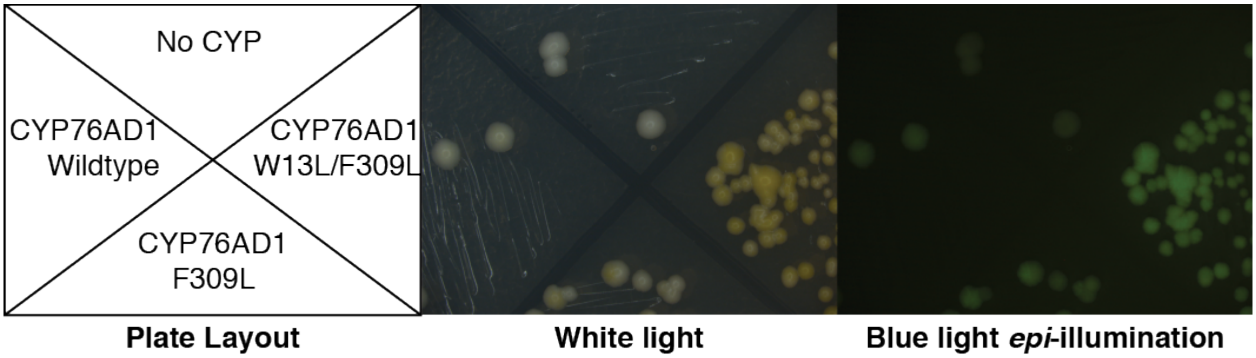


Figure 2-6. Fluorescence-based plate screening of betaxanthin producers and non-producers. Yeast expressing DOD (Strain 3) was transformed with a plasmid expressing either CYP76AD1 (Plasmid 3 from **Table 2-2**), F309L (Plasmid 4), or W13L/F309L (Plasmid 5) or an empty plasmid control (Plasmid 2), streaked onto SD-Leucine plates, and grown for 48 hours at 30˚C. Both plates were imaged under white (left) or blue light *epi*-illumination using an amber filter (right) on a Leica dissecting microscope at minimal tube magnification (0.71X base magnification) with DFC300FX camera.

Importantly, our enzyme-coupled biosensor is designed to be specific for L-DOPA as it should be capable of differentiating between the tyrosine hydroxylase and DOPA oxidase activities exhibited by tyrosinases. For the production of BIAs, the secondary DOPA oxidase activity is undesirable as it diverts L-DOPA to melanin[39]. While a tyrosinase from the button mushroom *Agaricus bisporus* (polyphenol oxidase 2, AbPPO2) has been successfully expressed in yeast, this enzyme had over 10-fold higher activity on L-DOPA than it did on L-tyrosine[40]. Given the existence of bacterial tyrosinases with relatively low DOPA oxidase activity[41], we hypothesized that our betaxanthin production assay could be used alongside enzyme mutagenesis to alter the relative substrate specificity for L-DOPA and L-tyrosine in enzymes that preferentially act on L-DOPA and are active in yeast. In this manner, undesired DOPA oxidase activity (activity 2 in **Figure 2-1** and **Figure 2-2a**) could be minimized and tyrosine hydroxylase activity maximized (activity 1).

2.2.2. Identification of a tyrosine hydroxylase active in *S. cerevisiae*

We first sought a tyrosine hydroxylase variant that yielded measurable levels of betaxanthin when expressed with DOD to use as a starting point for enzyme mutagenesis. While we expected this search to involve screening large variant libraries from cDNA and/or bioinformatic curation, we fortuitously found an acceptable candidate without resorting to these techniques. We note that future work to screen such libraries with our enzyme-coupled biosensor might reveal additional insights into this enzyme class and even identify variants with higher activity than those described in this work. Two candidate enzymes were selected for initial screening—AbPPO2 from *A. bisporus* (discussed above) and CYP76AD1 from the sugar beet *Beta vulgaris*. The latter is a recently identified cytochrome P450 DOPA oxidase[42]. This enzyme is responsible for the production of the violet pigment betanidin (a betaxanthin analog) in beets (**Figure 2-1**). Functional expression of CYP76AD1 in yeast was used to elucidate this enzyme's role in determining beet color, making it an attractive candidate for our purposes[42]. Notably, tyrosine hydroxylase activity was not previously reported for this enzyme.

Both AbPPO2 and CYP76AD1 were codon optimized for yeast and expressed along with DOD. As expected based on previous work, AbPPO2 showed extremely low, but detectable, betaxanthin production. Surprisingly, CYP76AD1 exhibited considerable tyrosine hydroxylase activity, leading to levels of betaxanthin that were clearly visible in colonies growing on agar plates (**Figure 2-7a**). This result was particularly unexpected given previous work in which CYP76AD1 and a DOD were co-expressed in yeast and yielded no observable betaxanthin[42]. One experimental difference that could account for these contradictory results is the identity of the DOD enzyme used. While our strains utilized a DOD from *M. jalapa*, previous experiments with CYP76AD1 were conducted using a *B. vulgaris* DOD, which has a K_m of 7 mM[43]—well above the levels that are likely to accumulate in yeast given the competing DOPA oxidase activity of CYP76AD1.

After achieving clearly detectable betaxanthin synthesis with CYP76AD1, we tested for L-DOPA production in the absence of DOD. CYP76AD1 yielded 1.3 mg/l L-DOPA, almost 20-times the levels obtained with AbPPO2. This result suggested our betaxanthin biosensor would have utility in further optimizing L-DOPA production, as the betaxanthin levels from AbPPO2 and CYP76AD1 were predictive of L-DOPA titers (**Figure 2-7b**). We hypothesized that titers could be improved by abolishing CYP76AD1's DOPA oxidase activity and set out to perform mutagenesis and screening with our L-DOPA biosensor.

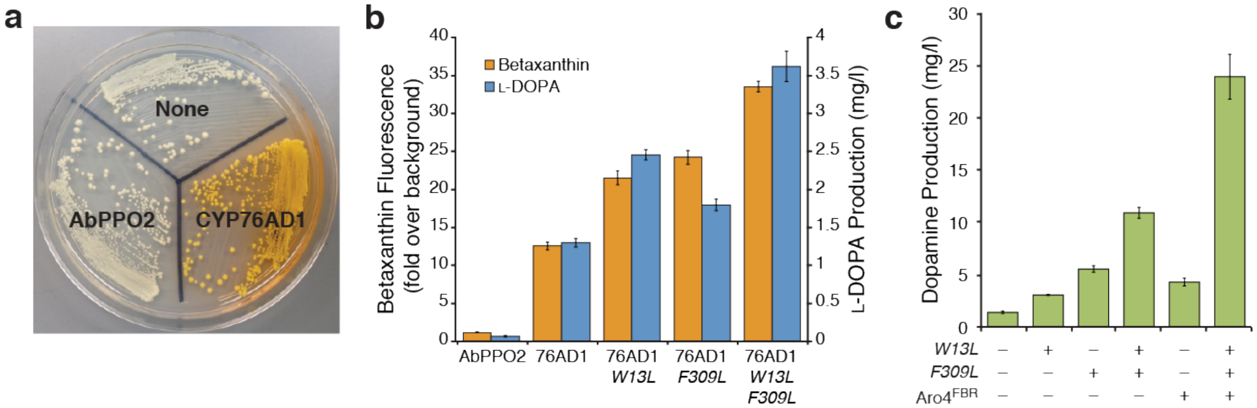
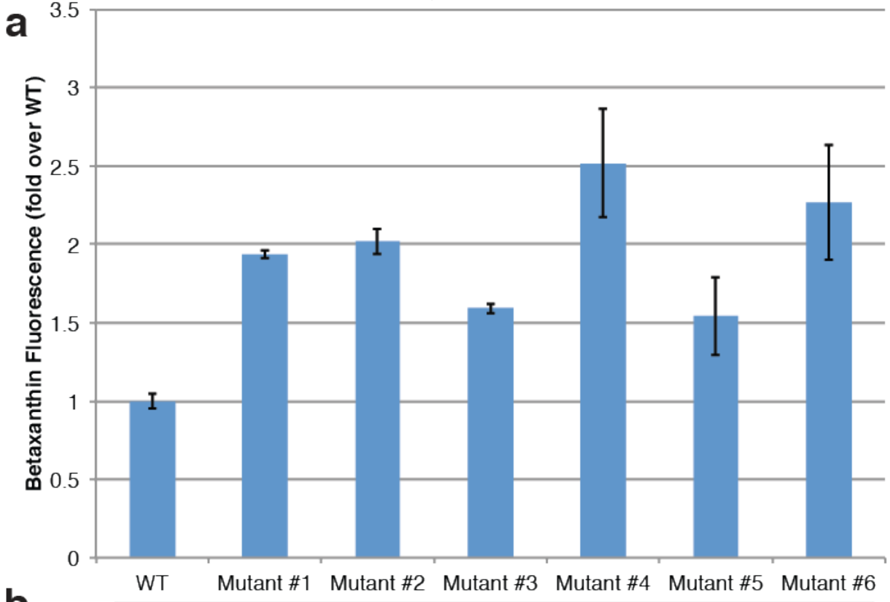


Figure 2-7. Isolation and improvement of a tyrosine hydroxylase functional in yeast. (**a**) Yeast cells expressing DOD alone or with a candidate tyrosine hydroxylase (either AbPPO2 from *A. bisporus* or CYP76AD1 from *B. vulgaris*), streaked on an agar plate with synthetic complete media. (**b**) Quantification of tyrosine hydroxylase activity for enzyme variants isolated with the betaxanthin biosensor screen. Yellow bars represent betaxanthin fluorescence in cells co-expressing a tyrosine hydroxylase variant with DOD as measured by cellular fluorescence on a microplate fluorometer. Fluorescence values were normalized to the autofluorescence of a control strain that expressed no tyrosine hydroxylase. Blue bars show L-DOPA titer after 48 hours of growth for strains expressing a candidate tyrosine hydroxylase as measured by LC/MS of culture supernatants and comparison to a standard curve. Amino acid substitutions for enzyme mutants are denoted in italics (76AD1 = CYP76AD1). Error bars indicate mean ± s.d. of four biological replicates. (**c**) Dopamine titers in the culture supernatant of cells co-

expressing DOPA decarboxylase (DODC) and a CYP76AD1 mutant as measured by LC/MS and comparison to a standard curve. The presence of the W13L and/or F309L mutation in a given variant is indicated with a "+". A strain overexpressing a feedback resistant mutant of Aro4p, known to increase intracellular tyrosine levels (Aro4^{FBR}), was also tested with select CYP76AD1 variants. Error bars indicate mean values ± s.d. of four biological replicates.

2.2.3. Improving tyrosine hydroxylase activity for dopamine production

We performed error-prone PCR to generate a library of CYP76AD1 mutants that was transformed into a yeast strain expressing DOD. This mutant library contained approximately 200,000 members with an average mutation rate of approximately 4 mutations/gene. The difference in betaxanthin production between library clones was sufficiently large that obvious differences between colonies could be detected by eye with the aid of blue light *epi*-illumination and a amber-green emission filter to amplify the betaxanthin fluorescence signal (**Figure 2-6**). We visually selected 17 of the highest betaxanthin producers for sequence analysis and confirmation of activity using microplate fluorometry. Six of the 17 clones sequenced contained the same F309L missense mutation; among those with this mutation, there were 2 distinct genotypes (**Figure 2-8**).



b

	Mutations				
Mutant #1	*87A>C*	L141I			
Mutant #2	*123T>A*	F309L	E465D		
Mutant #3	D2E	*150G>A*	*327C>T*	Y380H	
Mutant #4	*9T>C*	W13L	*1236T>C*	*1281T>C*	
Mutant #5	*180T>A*	F127L	*576G>A*	S232T	*714G>A*
Mutant #6	*684C>T*	F309L			

Figure 2-8. CYP76AD1 mutants yield increased betaxanthin production. Six mutants from the first round of betaxanthin screening were selected to undergo DNA shuffling.

Mutant #2 was independently isolated two times, and Mutant #6 was independently isolated four times. All other mutants were only isolated once. (**a**) CYP76AD1 mutants 1-6 were genomically integrated into a strain expressing DOD. Betaxanthin fluorescence was measured on a microplate fluorometer and normalized to cells expressing wildtype CYP76AD1. Bars represent mean values ± 1 s.d. of six biological replicates. (**b**) Sequence of CYP76AD1 mutants 1-6. Silent mutations (italicized) are indicated with their nucleotide substitution and indexed by their distance from the start codon.

We narrowed our set of hits to six by removing duplicate genotypes and any mutant proteins that showed less than a 1.5-fold improvement in betaxanthin production upon isolation and retransformation into the L-DOPA sensor base strain (**Figure 2-8**). These six mutants were used to construct a DNA shuffling library in order to randomly combine potentially beneficial mutations. The resulting library was again transformed into yeast expressing DOD and screened for betaxanthin production through a combination of visual screening and plate-based fluorescence measurements. We visually selected 67 high-fluorescence colonies to grow in liquid media, and secondary fluorescence screening was done in a microplate format (**Figure 2-9**). Out of the 28 highest producers that we sequenced from the DNA shuffling library, 20 had incorporated the same two mutations (W13L and F309L), suggesting that the effects of these mutations were additive or multiplicative. An additional silent mutation that was linked to W13L (9T>C) was also enriched for and included in future experiments involving the W13L mutation. For brevity, this linked combination will be referred to as W13L.

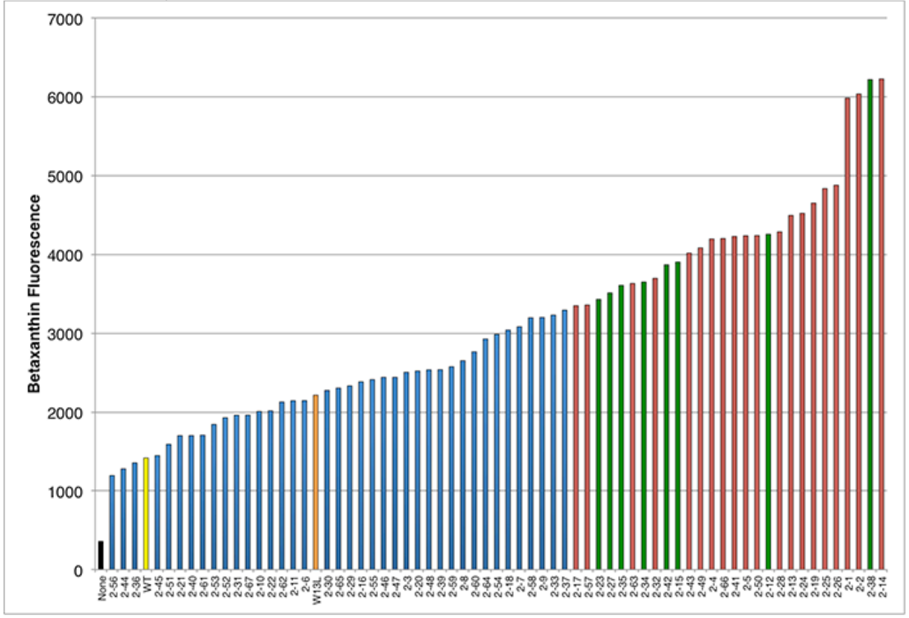


Figure 2-9. Analysis of CYP76AD1 mutants from the second round of mutagenesis. Sixty-seven colonies from the Golden Gate shuffling library (transformed into Strain 3)

were picked by visual inspection for high betaxanthin production. Colonies were grown in SD-uracil and measured for betaxanthin fluorescence on a microplate fluorometer. The 28 highest producing clones were sequenced. Red bars represent sequenced library members with both the W13L and F309L mutations incorporated into CYP76AD1. Green bars represent library members that were sequenced and did not have W13L and F309L mutations. Blue bars represent library members that were not sequenced. No CYP76AD1 control (Plasmid 2) = black; WT CYP76AD1 control (Plasmid 3) = yellow; CYP76AD1 W13L control (Plasmid 5) = orange.

Because W13L and F309L appeared to be our two best mutations, we analyzed their effects on betaxanthin and L-DOPA production individually and in combination. Compared to wildtype CYP76AD1, the W13L and F309L single mutants yielded 1.7-fold and 1.9-fold more betaxanthin respectively (Z-scores = 16.2 and 21.0) when measured on a microplate fluorometer (**Figure 2-7b**). When combined, the double mutant showed a net increase in betaxanthin fluorescence of 2.7-fold over wildtype CYP76AD1 with a Z-score of 10.0 relative to the F309L single mutant. When betaxanthin fluorescence was compared via flow cytometry, the mutant populations were easily distinguishable, showing a 4.3-fold improvement in mean fluorescence for the double mutant compared to wildtype CYP76AD1 (**Figure 2-10**).

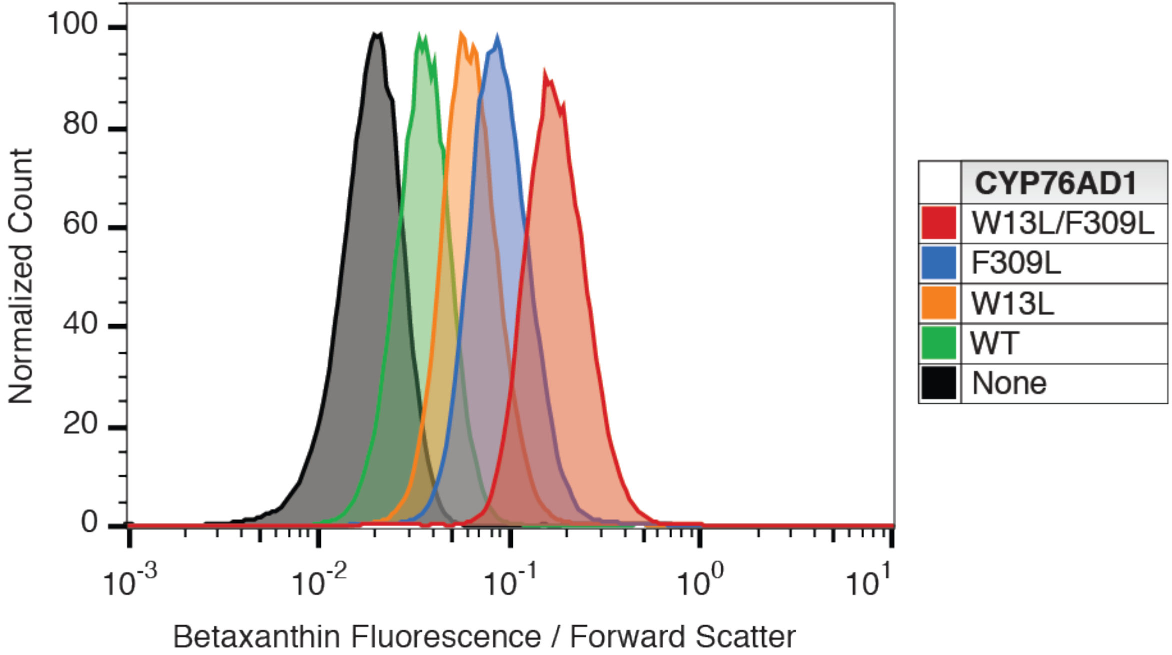


Figure 2-10. Flow cytometry fluorescence measurement can distinguish CYP76AD1 mutants in yeast. Yeast strains expressing DOD and a mutant of CYP76AD1 (Strains 4, 5, 7, 8, and 9) were grown in SD-uracil. Cellular betaxanthin fluorescence was measured using a flow cytometer with gating parameters as designated in **Figure 2-4**. Fluorescence was divided by forward scatter (FSC) to account for variability in cell size, and a histogram of each strain of cells was generated and superimposed. Cell counts

were normalized as a percent of the mode of each population. At least 15,000 cells per condition are represented on the histograms.

In the absence of DOD, CYP76AD1 with the W13L, F309L, and W13L/F309L mutations improved upon the L-DOPA titer of wildtype CYP76AD1 by 1.9-fold, 1.4-fold, and 2.8-fold, respectively (**Figure 2-7b**). Overall, there was strong linear correlation between betaxanthin fluorescence and intracellularly produced L-DOPA ($R^2 = 0.92$), although there was a slight discrepancy between the two measurements for the F309L mutant.

We next tested the ability of CYP76AD1 to catalyze the synthesis of the downstream BIA intermediate dopamine when expressed with a DOPA decarboxylase (DODC) from *Pseudomonas putida* that is highly specific for L-DOPA[44]. The specificity of DODC is important, as decarboxylation of L-tyrosine would introduce a side pathway to tyramine and reduce flux to dopamine[22]. When combined into a single strain, both activities resulted in dopamine production, something that has not been demonstrated previously in *S. cerevisiae*. Using wildtype CYP76AD1, we achieved a dopamine titer of 1.5 mg/l (**Figure 2-7c**). Both the W13L and F309L single mutants improved dopamine production; in fact, the effect of the F309L mutant on dopamine titer was considerably larger than had been observed for L-DOPA titer (3.8-fold vs. 1.4-fold respectively). The W13L/F309L double mutant of CYP76AD1 increased the level of dopamine production to a surprisingly high 10.8 mg/l, 7.4-fold greater than the wildtype enzyme.

Given the high tyrosine hydroxylase activity of the CYP76AD1 W13L/F309L double mutant, we hypothesized that L-tyrosine availability might be limiting dopamine production in our strain. Therefore, we overexpressed a feedback-insensitive mutant of the L-tyrosine pathway enzyme Aro4p that is known to markedly increase the intracellular L-tyrosine concentration in *S. cerevisiae*[45]. Doing so resulted in an additional 2.2-fold increase in dopamine levels. In this strain background, we achieved a final dopamine titer of 23.8 mg/l when cells were grown for 48 hours in 2X synthetic complete media with 4% glucose (**Figure 2-7c**).

2.2.4. Characterization of reduced DOPA oxidase activity

We were interested in characterizing the mechanism by which our CYP76AD1 mutants led to increased tyrosine hydroxylase activity and started by looking at changes in enzyme expression. Western blot analysis showed that the W13L mutation yielded 2.8-fold more protein, while the F309L mutation led to a 1.6-fold increase in expression (**Figure 2-11a**). These changes were consistent with measurements of protein expression made via fluorescent protein fusion (**Figure 2-12**). Fluorescence microscopy of these fusion proteins showed that neither mutation affected localization of CYP76AD1 to the endoplasmic reticulum (**Figure 2-13**). While increased expression is sufficient to explain the improvements in L-DOPA and dopamine production caused by the W13L mutation, the 3.8-fold improvement in dopamine titer caused by the F309L mutation cannot be fully explained by a 1.6-fold increase in protein abundance. Therefore, we investigated potential changes in catalytic activity resulting from this mutation.

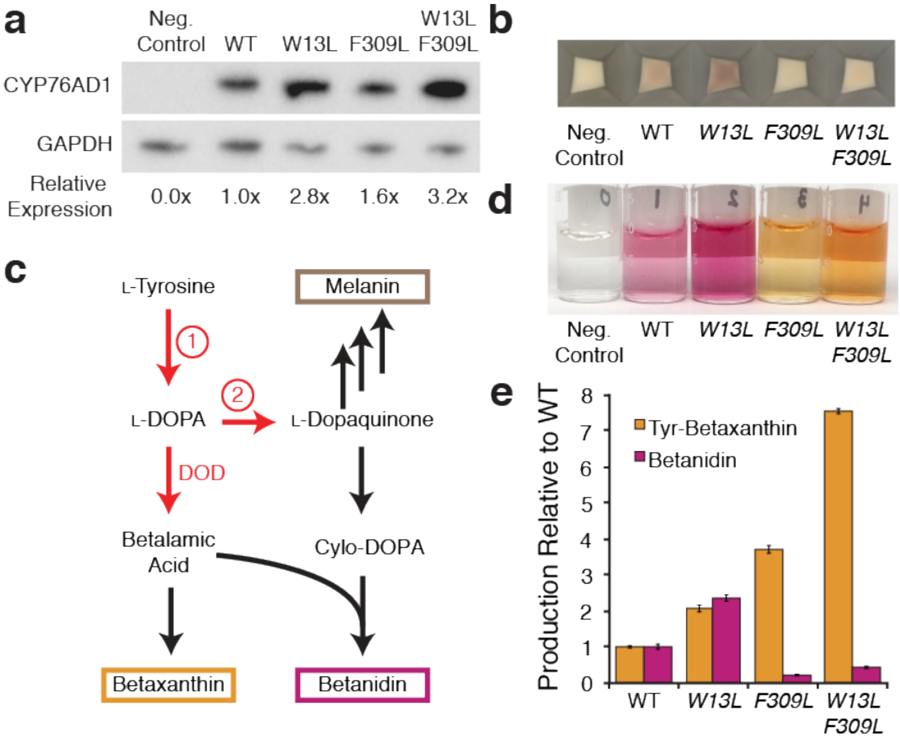


Figure 2-11. Characterization of reduced DOPA oxidase activity in CYP76AD1 mutants. (**a**) Western blot comparing expression level of CYP76AD1 mutants in yeast. Enzymes contain a C-terminal FLAG-tag. An anti-GAPDH antibody was used as a loading control. Relative expression represents the ratio of anti-FLAG intensity to anti-GAPDH intensity, normalized to wildtype CYP76AD1. (**b**) Pelleted yeast cells expressing CYP76AD1 mutants (and Aro4^{FBR}). Neg. Control denotes a strain not expressing CYP76AD1. (**c**) DOD generates the intermediate betalamic acid that can either react with free amines to produce betaxanthins (a yellow family of pigments) or with the L-dopaquinone derivative cyclo-DOPA to produce betanidin (a violet pigment). Betanidin is labile but can be stabilized with the addition of the reducing agent ascorbic acid. Relative levels of betanidin production were used to compare DOPA oxidase activity (activity 2) in CYP76AD1 variants. See **Figure 2-1** for a more

detailed pathway diagram. (**d**) Culture supernatant from strains co-expressing DOD and a CYP76AD1 mutant grown in minimal media with ascorbic acid. Neg. Control indicates no CYP76AD1 expression. (**e**) LC/MS analysis of tyrosine-betaxanthin (yellow) and betanidin (violet) levels in the supernatants from (**d**). Tyrosine-betaxanthin was selected as a representative member of the betaxanthins since many different species exist in the culture supernatant. Peak areas were fit to relative standard curves generated by serial dilution of the samples to obtain levels of betanidin and tyrosine-betaxanthin relative to the wildtype CYP76AD1 sample. Error bars represent mean values ± s.d. of four biological replicates.

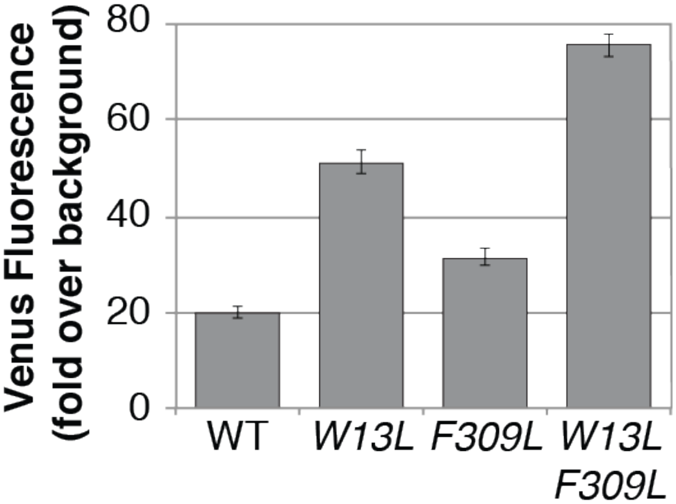


Figure 2-12. Measurement of expression levels of Venus-tagged CYP76AD1 mutants. As a complement to **Figure 2-11a**, expression levels of different CYP76AD1 mutants were quantified using strains expressing a C-terminal Venus fusion protein (Strains 41, 42, 43, and 44). Fluorescence measurements were taken using a microplate fluorometer and normalized to the fluorescence value of cells not expressing any fluorescent protein (Strain 1). Error bars represent mean values ± s.d. of ten biological replicates.

While pelleting L-DOPA-producing cultures with feedback-insensitive Aro4p overexpressed, we noticed that cells expressing wildtype CYP76AD1 or the W13L single mutant were brown in color compared to cells expressing a version of CYP76AD1 with the F309L mutation incorporated (**Figure 2-11b**). We suspected that this brown color was likely due to the accumulation of melanin generated by CYP76AD1's DOPA oxidase activity (**Figure 2-11c**). Because of challenges associated with accurate quantification of melanin and its precursor L-dopaquinone (**Figure 2-14**), we again turned to the betaxanthin pathway to analyze this apparent change in CYP76AD1's catalytic activity.

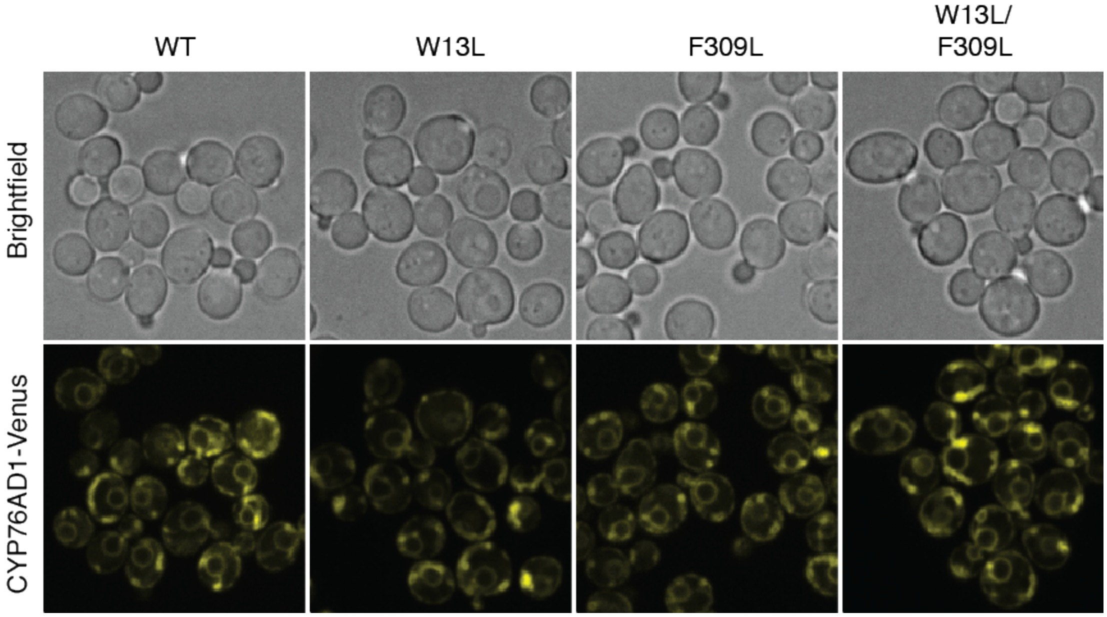


Figure 2-13. Venus-tagged CYP76AD1 mutants localize to the endoplasmic reticulum. Strains expressing a C-terminal Venus fusion of CYP76AD1 mutants (Strains 41, 42, 43, and 44) were examined using confocal fluorescence microscopy. The localization of Venus-tagged CYP76AD1 appears to be primarily at the endoplasmic reticulum and does not show obvious differences in localization between any of the mutants.

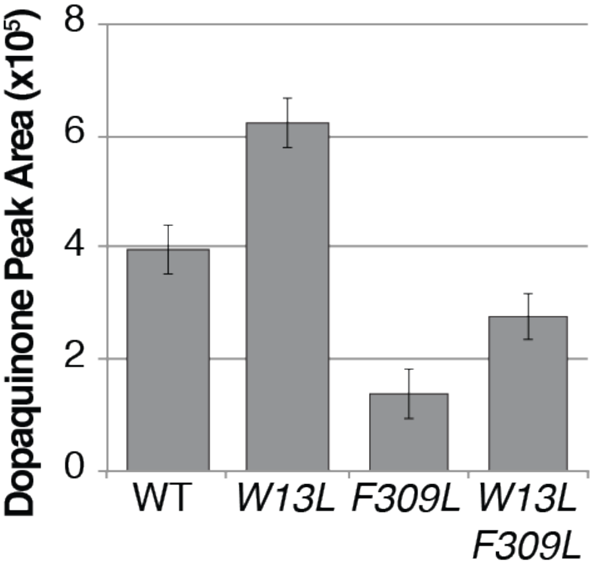


Figure 2-14. The F309L mutation decreases the amount of dopaquinone produced. Yeast strains expressing mutants of CYP76AD1 (Strains 10, 11, 12, 13) were grown in 2X SD-uracil supplemented with 10 mM ascorbic acid. Dopaquinone was measured using the same LC/MS protocol as was used for dopamine quantification, except the ion extracted from the data was *m/z* 196.06, representing dopaquinone $[M+H]^+$. Note that a chemical standard for L-dopaquinone could not be obtained. Bars represent mean values ± s.d. of four biological replicates.

As stated previously, the pathway from L-DOPA to betaxanthin goes through the intermediate betalamic acid, which undergoes spontaneous condensation with amines (**Figure 2-1**). If L-dopaquinone is generated via a DOPA oxidase, however, betalamic acid can react with the L-dopaquinone derivative cyclo-DOPA to form a violet pigment called betanidin (**Figure 2-1** and **Figure 2-11c**).

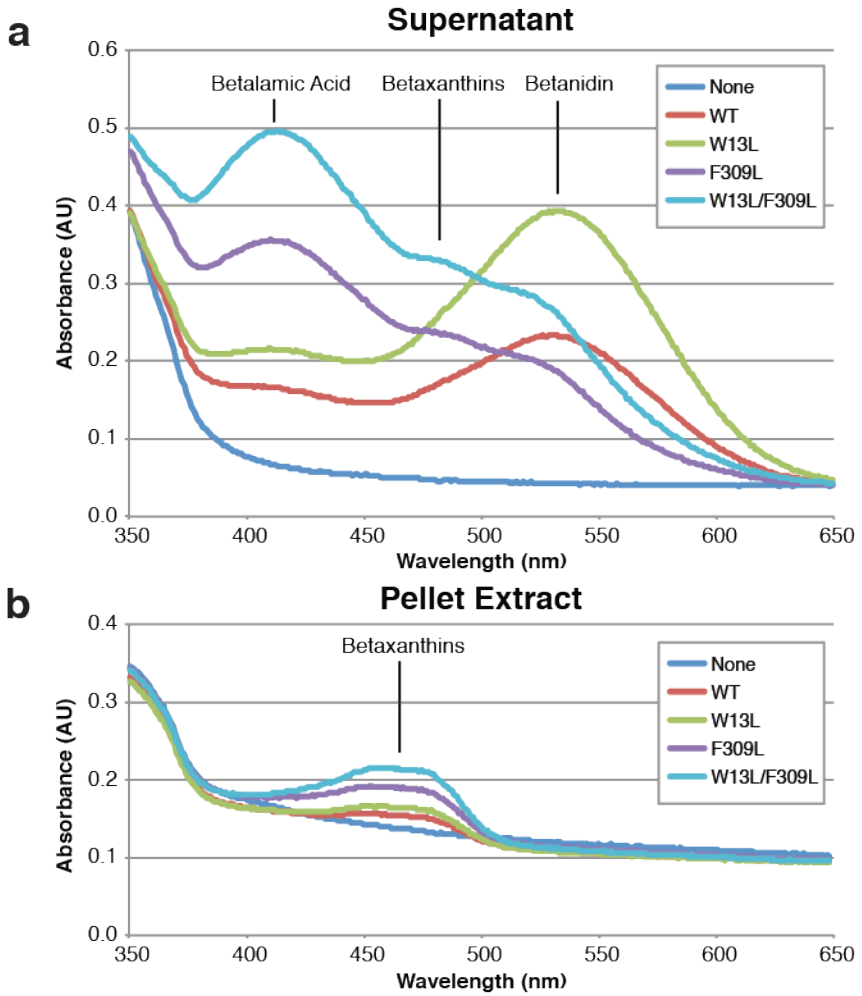


Figure 2-15. Absorbance spectra of supernatants and corresponding pellet extracts. Supernatants are from **Figure 2-11d**. Measured absorbance maxima: betalamic acid = 413nm, betaxanthins = 460-480nm, betanidin = 533nm. Based on comparisons of spectra between supernatant and cellular fluorescence, it appears that a large fraction of the betaxanthin remains intracellular relative to betanidin and betalamic acid. (**a**) Absorbance spectra of yeast supernatants measured using a microplate spectrophotometer with a 1 nm wavelength step. (**b**) Absorbance spectra of yeast pellet acetonitrile extracts, also measured using microplate spectrophotometer with a 1 nm wavelength step.

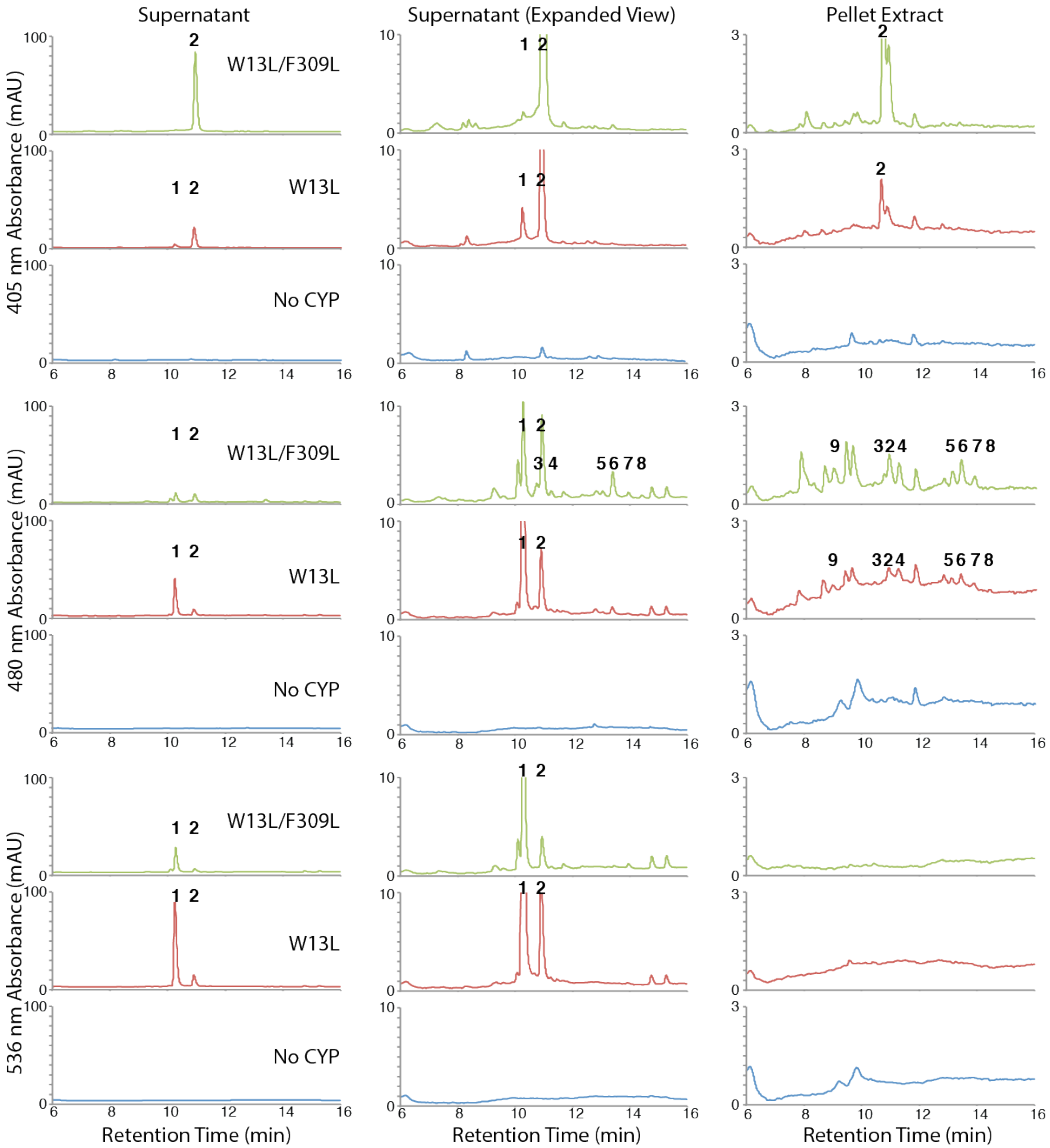


Figure 2-16. HPLC traces for CYP76AD1/DOD strains. Yeast expressing DOD and no CYP76AD1 (Strain 4, blue), W13L (Strain 8, red) or W13L/F309L (Strain 9, green) were grown in minimal media with 10 mM ascorbic acid. Supernatants and extracts of pellets were separated on a C18 column. 405 nm (betalamic acid), 480 nm (betaxanthin), and 536 nm (betanidin) absorbance traces are shown. Where identified, peaks were labeled with numbers directly above the corresponding peak. 1, Betanidin; 2, Betalamic acid; 3, Tyrosine-betaxanthin; 4, Valine-betaxanthin; 5-6, (Iso)Leucine-betaxanthin; 7, Phenylalanine-betaxanthin; 8, Tryptophan-betaxanthin; 9, Proline-betaxanthin.

We used betanidin production as an indicator of DOPA oxidase activity in yeast. Doing so required supplementing the growth media with ascorbic acid as a reducing agent to prevent spontaneous betanidin oxidation that causes the pigment to polymerize and lose its violet color[46]. When co-expressed with DOD, candidate tyrosine hydroxylases with high DOPA oxidase activity should produce a violet supernatant containing high levels of betanidin, while enzymes with low DOPA oxidase activity will generate a yellow supernatant that is composed mostly of betaxanthin and betalamic acid. As expected, when wildtype CYP76AD1 was tested in this betaxanthin/betanidin assay, the supernatant was violet, indicative of high DOPA oxidase activity (**Figure 2-11d**). While the W13L mutation increased the concentration of pigments in the culture supernatant, the overall violet color did not change (**Figure 2-15**). In contrast, incorporation of the F309L point mutation into CYP76AD1 produced a shift in color from violet to yellow, which was also observed for the W13L/F309L double mutant. This change demonstrates that the F309L mutation reduces DOPA oxidase activity. To better quantify the culture supernatants, we measured betanidin and betaxanthin using liquid chromatography–mass spectrometry (LC/MS). Tyrosine-betaxanthin was selected as a representative member of the betaxanthin species since many different betaxanthins can form from the spontaneous conjugation of an amine with betalamic acid (**Figure 2-16**). Our results showed that betanidin levels decreased by 80% in the F309L mutant, compared to wildtype CYP76AD1 (**Figure 2-11e**). This decrease in betanidin was accompanied by a 3.7-fold increase in tyrosine-betaxanthin production. When quantified via LC/MS, the changes in tyrosine-betaxanthin caused by the mutations showed strong linear correlation (**Figure 2-17**, $R^2 = 0.999$) to the dopamine levels measured in **Figure 2-7c**. In fact, betaxanthin titers were more predictive of dopamine titers than was direct measurement of L-DOPA in the absence of a downstream enzyme.

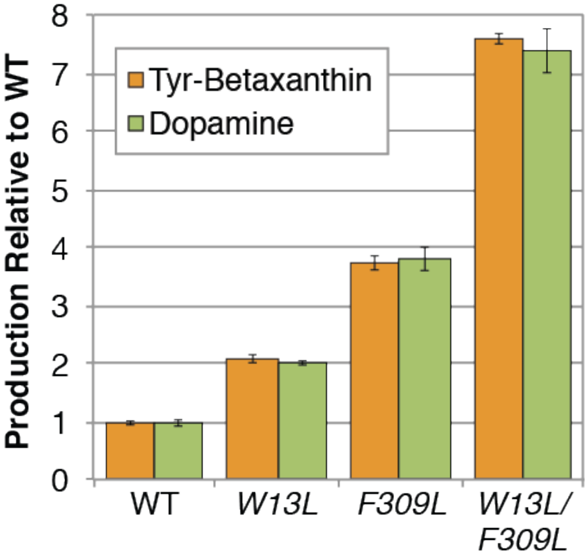


Figure 2-17. The effects of CYP76AD1 mutations on tyrosine-betaxanthin and dopamine-betaxanthin are highly correlated when measured via LC/MS. Dopamine production from **Figure 2-7c** and tyrosine-betaxanthin production from **Figure 2-11e** are plotted relative to wildtype levels of production. Strains 4, 5, 7, 8, 9 (betaxanthin) and 15, 16, 17, 18, 19 (dopamine) were used. Error bars represent mean values ± s.d. of four biological replicates.

To further investigate the importance of the F309 residue on DOPA oxidase activity, we performed a protein alignment of CYP76AD1 to its three known orthologs (CYP76AD2, CYP76AD3, and CYP76AD4), which showed that this residue is conserved in each (**Figure 2-18**). To determine whether the F309L mutation would have similar effects in these orthologs, we codon optimized the genes for expression in yeast and tested wildtype and F309L mutants in the betaxanthin/betanidin assay. In all cases, introduction of the mutation produced a shift from violet to yellow supernatants (**Figure 2-19**). We further aligned the CYP76AD1 protein sequence to its seven closest paralogs from the recently sequenced beet genome and found that all seven contained the same F309L substitution that we had identified through PCR mutagenesis (**Figure 2-20**)[47]. In beets, inactivation of CYP76AD1 abolishes betanidin production while betaxanthin production persists, suggesting the presence of a dedicated tyrosine hydroxylase for betanidin synthesis[42]. We tested the two closest paralogs for tyrosine hydroxylase activity in *S. cerevisiae* but observed no betaxanthin production (data not shown). Efforts to determine a crystal structure for CYP76AD1 were unsuccessful, however, a homology model was generated by threading CYP76AD1 onto human CYP1A2 (**Figure 2-21**)[48]. The model provided little insight into the mechanism of the F309L mutation's effect on CYP76AD1's behavior as the residue was 14 Å from the enzyme's predicted active site.

```
CYP76AD2        MDNATLAMILTIWFISINFIKMFFYHQNTKLSLPPGPKPLPIIGNILEVGKKPHRSFANL 60
CYP76AD4        MDNATLAMLLAIWFISFHFIKMLFTNQSTKL-LPPGPKPLPIIGNILEVGKKPHRSFANL 59
CYP76AD1        MDHATLAMILAIWFISFHFIKLLFSQQTTKL-LPPGPKPLPIIGNILEVGKKPHRSFANL 59
CYP76AD3        MDFLTLVMILSIIFFFYNLLKMKFTTHSDAQ-LPPGPKPMPIFGNIFELGEKPHRSFANL 59
                **  **.*:*:* *:  :::*: *  :.    *******:**:***:*:*:*********

CYP76AD2        AKIHGPLISLRLGSVTTIVVSSAEVAKEMFLKKDQPLSNRNVPNSVTAGDHHKLTMSWLP 120
CYP76AD4        AKIHGPLISLKLGSVTTIVVSSAEVAKEMFLKKDQPLSNRTVPNSVTAGDHHKLTMSWLP 119
CYP76AD1        AKIHGPLISLRLGSVTTIVVSSADVAKEMFLKKDHPLSNRTIPNSVTAGDHHKLTMSWLP 119
CYP76AD3        AKTHGPLMSLRLGSVTTIVVSSAEVAKEMFLKNDQSLADRSVPNSVTAGDHHKLTMSWLP 119
                ** ****:**:************:********:*:.*::*.:******************

CYP76AD2        VSPKWRNFRKITAVHLLSPLRLDACQSLRHAKVQQLYQYVQECALKGQSVDIGKAAFTTS 180
CYP76AD4        VSPKWRNFRKITAVHLLSPLRLDACQSLRHAKVQQLFQYVQECAQKGQAVDIGKAAFTTS 179
CYP76AD1        VSPKWRNFRKITAVHLLSPQRLDACQTFRHAKVQQLYEYVQECAQKGQAVDIGKAAFTTS 179
CYP76AD3        VSPKWKNFRKITAVHLLSPQRLDACHALRHAKVKQLYEYVQECALKGEAVDIGKAAFTTS 179
                *****:************* *****:::*****:**::****** **::***********

CYP76AD2        LNLLSKLFFSKELACHKSHESQELKQLIWNIMEDIGKPNYADYFPILGCIDPLGIRRRLA 240
CYP76AD4        LNLLSKLFFSKELASHKSRESQEFKQLIWNIMEDIGKPNYADYFPILGCVDPSGIRRRLA 239
CYP76AD1        LNLLSKLFFSVELAHHKSHTSQEFKELIWNIMEDIGKPNYADYFPILGCVDPSGIRRRLA 239
CYP76AD3        LNLLSNLFFSVELANHTSNTSQEFKQLIWDIMEDIGKPNYADYFPLLKYVDPSGIRRRLA 239
                *****:**** *** *.*. ***:*:***:***************:*  :** *******

CYP76AD2        ANFDKLISVFQTIISERLENDIN--SNATTNDVLDVLLQLYKQKELSMGEINHLLVDIFD 298
CYP76AD4        SNFDKLIEVFQCIIRQRLERNP---STPPTNDVLDVLLELYKQNELSMGEINHLLVDIFD 296
CYP76AD1        CSFDKLIAVFQGIICERLAPDSSTTTTTTTDDVLDVLLQLFKQNELTMGEINHLLVDIFD 299
CYP76AD3        ANFDKLIDVFQSFISKRLSSAYS--SATSLDDVLDVLLKLLKEKELNMGEINHLLVDIFD 297
                ..***** *** :* :**       : .. :*******:* *::**.*************

CYP76AD2        AGTDTTSSTFEWVMAELIRNPKMMEKAQQEIHEVLGKD-RQIQESDIIKLPYLQALIKET 357
CYP76AD4        AGTDTTSSTFEWVMAELIRNPEMMAKAQDEIEQVLGKD-RQIQESDIIKLPYLQAIIKET 355
CYP76AD1        AGTDTTSSTFEWVMTELIRNPEMMEKAQEEIKQVLGKD-KQIQESDIINLPYLQAIIKET 358
CYP76AD3        AGTDTTSNTFEWAMAELMRNPIMMKRAQNEIALVLGKDNATIQESDIANMPYLQAIIKET 357
                *******.****.*:**:*** ** :**:**  *****   ****** ::*****:****

CYP76AD2        LRLHPPTVFLLPRKADMDVELYGYVVPKDAQILVNLWAIGRDSQVWEKPNVFLPERFLGS 417
CYP76AD4        LRLHPPTVFLLPRKADTDVELYGYIVPKDAQILVNLWAIGRDSQAWENPKVFSPDRFLGC 415
CYP76AD1        LRLHPPTVFLLPRKADTDVELYGYIVPKDAQILVNLWAIGRDPNAWQNADIFSPERFIGC 418
CYP76AD3        LRLHPPTVFLLPRKAITNVKLYGYIVPKNAQILVNLWAIGRDPKVWKNPNEFLPDRFLNS 417
                ***************  :*:****:***:*************.:.*::.. * *:**:..

CYP76AD2        DVDVKGRDFGLLPFGAGKRICPGMNLAIRMLTLMLATLLQFFNWKLEDGMNPQDLDMDEK 477
CYP76AD4        EIDVKGRDFGLLPFGAGKRICPGMNLAIRMLTLMLATLLQFFNWKLQDGMSLEDLDMEEK 475
CYP76AD1        EIDVKGRDFGLLPFGAGRRICPGMNLAIRMLTLMLATLLQFFNWKLEGDISPKDLDMDEK 478
CYP76AD3        DIDVKGRDFGLLPFGAGRRICPGMNLAYRMLTLMLATLLQSFDWKLPHRNSPLDLDMDEK 477
                ::***************:********* ************ *:***    .  ****:**

CYP76AD2        FGIALQKNKPLEIIPSLRH 496
CYP76AD4        FGIALQKTKPLRIIPVSRY 494
CYP76AD1        FGIALQKTKPLKLIPIPRY 497
CYP76AD3        FGIALQKTKPLEIIPLIKY 496
                *******.***.:**  ::
```

Figure 2-18. Multiple sequence alignment of CYP76AD1 and orthologs. W13 and F309 residues are highlighted in yellow. Amino acids are color-coded based on their properties (red = small; blue = acidic; magenta = basic; green = hydroxyl, sulfhydryl, amine; grey = other). Asterisks indicate fully conserved residues, colons and periods indicate strong and weak conservation respectively. CYP76AD1 (*Beta vulgaris*, UniProt: I3PFJ5); CYP76AD2 (*Amaranthus cruentus*, UniProt: I3PFJ7); CYP76AD3 (*Mirabilis jalapa*, UniProt: I3PFJ8); CYP76AD4 (*Celosia cristata*, Uniprot: M9RR47).

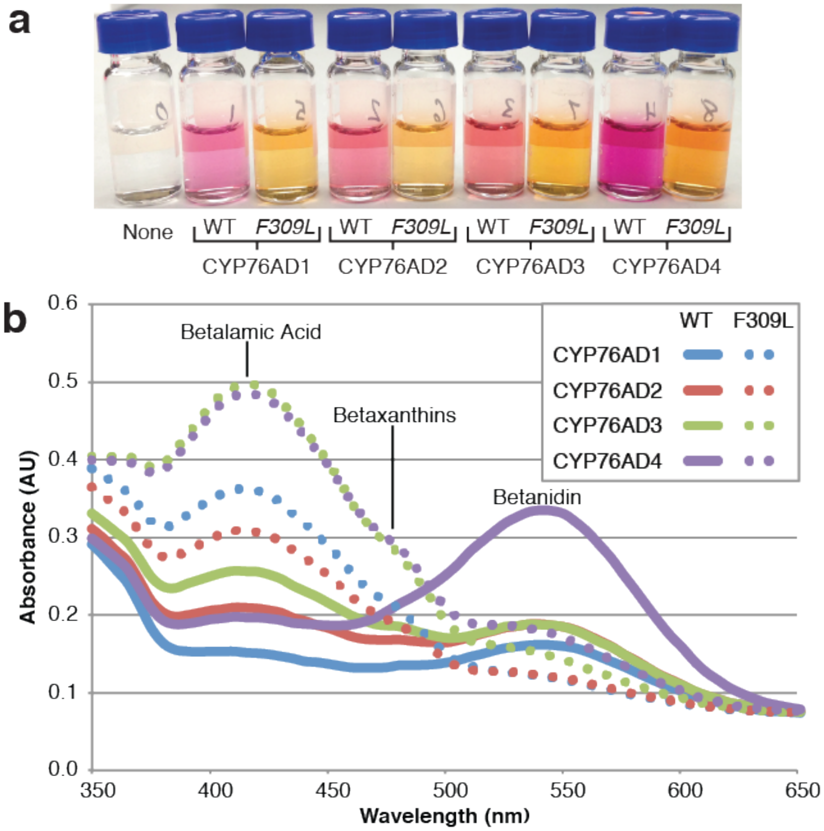


Figure 2-19. Effects of the F309L mutation on CYP76AD1 and its orthologs. (**a**) Culture supernatant from strains co-expressing DOD and one of four CYP76AD orthologs (Strains 5, 7, 45, 46, 47, 48, 49, 50) grown in minimal media with ascorbic acid. All wildtype (WT) enzymes produced a red supernatant, while mutants with an F309L substitution produced yellow supernatants, indicative of reduced DOPA oxidase activity. (None = DOD expression alone, Strain 4). Note: F309L indicates mutation of the residue corresponding to F309 in CYP76AD1 based on sequence alignment, not necessarily the exact position of the mutated amino acid in each ortholog. (**b**) Absorbance spectra of the supernatants shown in (**a**).


a.

```
Bv1u_022460_qtnn       MEYFTTLLLLLSIIILTILLSTKL--------FTKSNLPPGPKPWPIIGNILELGKLPHQAVDKLSKTYGPILSLKLGSITTIVISSPEIVKEMFLEHDLALSSRPSPDASRVGNHNKFS 112
Bv1u_022470_qtfp       MEYYT-LSLIFVPIIFTTLFFLQT--------LSKSKLPPGPKPWPIIGSLHKLGDRPHRAVAELSKIYGPIMSLKLGTITTIVISSPEIVKELFLKHDLAVSSRTVPNAARAVDHDKFS 111
Bv1u_022400_apow       -------------------------------------------------SIHKLGDKPHHVVAKLSKIYGPIMSLKLGSITTIVISSPEIAKEMFLEHDLALSSRPMQTKS----LKKFS 233
Bv1_001290_xkhk        MDYYTTLL-ILCSIFFAFFHIYKLISPSSKLTSNDSRLPPGPKPIPILGNLSHLGDSPHLSLANLAKTYGPLMSLKFGSITTIVVSSSIVAKEMFQKHDLTLSSRHASAAVRANGHDKCS 119
Bv1_001300_sidm        MDYYTTLFFILLPIFFALLYLYVFKRNPTFTTNNNARLPPGPKPIPILGNLPHLGDKPHHSLANLAKTYGPLMSLKFGSITTIVVSSSIVAKEMFQKHDLTLSSRHVSAAVRANGHDKFS 120
Bv9_228610_yqeq        MDNATLAV--ILSILFVFYHIFKS----FFTNSSSRRLPPGPKPVPIFGNIFDLGEKPHRSFANLSKIHGPLISLKLGSVTTIVVSSASVAEEMFLKNDQALANRTIPDSVRAGDHDKLS 114
Bv9_228860_ickx        --------------------------------------------MPIFGNIFDLGEKPHRSFANLAKIHGPLVSLQLGSVTTVVVSSADVAKEMFLKNDQALANRTIPDSVRAGDHDKLS 76
CYP76AD1               MDHATLAM--ILAIWFISFHFIKL----LFSQQTTKLLPPGPKPLPIIGNILEVGKKPHRSFANLAKIHGPLISLRLGSVTTIVVSSADVAKEMFLKKDHPLSNRTIPNSVTAGDHHKLT 114
                                                                        .: .:*. **  . :*:* :**::**:*::**:*:**. :.:*:* ::* .::.*            .* :

Bv1u_022460_qtnn       IVWLPVSPKWRDLRKIATIQLFTTQRLDSSQELRQIKVNELVDYVRQCCEKGLPVDVGKAGFTTTLNMLSNTFFSMDLASHASSNSQEFKDLVWSLLEEGAKPNVSDFFPIVRELDLQGV 232
Bv1u_022470_qtfp       MVWLPVCPKWRDLRKIATIQLFTTQRLDTSQFLRQKKVKELVEYARQCCEKGVALDIGKAGFTTTLNLLSNTFFSMDLASYDSLDSQEFKDLVWHLLEEGARPNVSDFFPLVKHFDLQGV 231
Bv1u_022400_apow       MVWLPVCPKWRHLRKIATLQLFTTQRLDISQVLRHTKVKELMEYAQQCCENNLPVDIGKAAFTTSLNLLSNTIFSMDLASHVSSNSQEFKDIVWNIMES--RPNVLDYIPLVRKLDLQGV 351
Bv1_001290_xkhk        IAWLPVCAKWRSLRKISAIHLFSSQKLDSSQALRQEKVSKLIDYVKECCNVGEEIDVGGVAFTTSLNLLSNTFFSFDLASYNSSDSGEFKELVWKIMEEIGKPNLVDCFPMLRFLSVFSV 239
Bv1_001300_sidm        MAWLPVGPKWRALRKIATIHLFSSQRLDSSQALRREKVSKLIDYVKECCNVGEAIDVGGVAFTTSLNLLSNTFFSFDLASYNSSDSGEFKELVWKIMEEIGKPNLADCFPMLRFLSVFSV 240
Bv9_228610_yqeq        MSWLPVSQKWRNMRKISAVQLLSNQKLDASQPLRQTKVKQLLSYVQDCSKKMQPVDIGRAAFTTSLNLLSNTFFSIELASHESSASQEFKQLMWNIMEEIGRPNYADFFPILGYIDPFGI 234
Bv9_228860_ickx        MSWLPVSAKWRNLRKISAVQLLSTQRLDASQAHRQSKVQQLLEYVHDCSKKGQPVDIGRAAFTTSLNLLSNTFFSVELASHESSASQEFKQLMWNIMEEIGRPNYADFFPILGYLDPFGI 196
CYP76AD1               MSWLPVSPKWRNFRKITAVHLLSPQRLDACQTFRHAKVQQLYEYVQECAQKGQAVDIGKAAFTTSLNLLSKLFFSVELAHHKSHTSQEFKELIWNIMEDIGKPNYADYFPILGCVDPSGI 234
                       : ****  *** :***::::*:: *:** .*  *: **.:* .*.::*.:    :*:* ..***:**:**: :**.:** : *  * ***:::* ::*.  :** *  :*::  ..  .:

Bv1u_022460_qtnn       SKNRRVHMKKLMGIFEEIIDGRLTKLKDVK------DDVLSTLLKLVKDEE--LNLDDVKHMLMDLFLAGTDTTSITLEWAMTELLRNPEKMEKVQIELDKVLGKD-SSLQESMISKLPY 343
Bv1u_022470_qtfp       LKTTTSYLKKLIGIFEEIIDKRLKDPTDVK------DDVLSTLLKLVEDDE--LSLDDVKHLLADLFIAGTDTTSNTLEWAMAELLRNPEKMEKAQAEINKVLGKD-SSMQENDISKLPY 342
Bv1u_022400_apow       LKRKRSYFKKIMGVFEEIIDVRLKDPTDVK------DDVLGTLLKLVKDEE--LSLHDVKHMLFDLFLAGTDTTSSTLEWAMTELLRNPKVMEKAQIEIDQVLGKD-GSMQELDIAKLPY 462
Bv1_001290_xkhk        KGKLLGYDNKLNEVFENIIQKRLQNYCGDSS---SGGDVLDTLLRLMKENELDLDLGDIKHLLMDFFTAGTDTTSSTLEWAMTELLRNPEKMAKAQVELEQVLGKN-KVVGEFDISKLPY 355
Bv1_001300_sidm        NYKVMVYGNRLNDVFEDIIQNRLISSSADK----IGGDVLDTLLRLMKENESELSLDDIKHLLMDFFTAGTDTTSSTLEWAMTELLRNPEKLAKAQAELEQVVGKN-KVVKEADISKLPY 355
Bv9_228610_yqeq        RRRLAGYFDKLIDVFQDIIRERQKLRSSNSSGAKQTNDILDTLLKLHEDNE--LSMPEINHLLVDIFDAGTDTTASTLEWAMAELVKNPEMMTKVQIEIEQALGKDCLDIQESDISKLPY 352
Bv9_228860_ickx        RRRLAGYFDQLIAVFQDIIGERQKIRSANLSGGKQTNDILDTLLNLYDEKE--LSMGEINHLLVDIFDAGTDTTASTLEWAMAELVKNPYMMVKVQDEIEKAIGKGCSMVQESDISKLPY 314
CYP76AD1               RRRLACSFDKLIAVFQGIICERLAPDSSTTT-TTTTDDVLDVLLQLFKQNE--LTMGEINHLLVDIFDAGTDTTSSTFEWVMTELIRNPEMMEKAQEEIKQVLGKD-KQIQESDIINLPY 350
                               .::  :*: **  *            .*:*..**.* .:.*  * : :::*:* *:* ******: *:**.*:**:**  : *.* *:.:.:**.   : *  * :***

Bv1u_022460_qtnn       IQAIVKETLRLHPPTPFLIPHKAEKDVLLCNYLVPKNSIIWVNLWSIARSPSVWPNPESFSPERFLEMEIDIKGRDFKLIPFGSGRRMCPGMPLAYRMTHMLLATLLHSFNWKYGEA--- 460
Bv1u_022470_qtfp       VQAIVKETFRLHPVTPFLVPHKAEKDILLGNYLVPKNSTIWVNVWSIGRNPSVWSKPELFSPERHLELEIDVKGQAFELMPFGSGRRTCPGMPLAYRMTHLMLATLLHSFHWKYGHGK-- 460
Bv1u_022400_apow       IQALVKEILRLHPPAPFLIPHMAIEDVQLCGYLVPKQSTIWVNVWSIGRDPSVWTKSKMFSPERFLEKEIDVKGRNFELIPFGSGRRICPGMPLAYRMVHLTLATLLHSFNWKYVNEASI 582
Bv1_001290_xkhk        LQAIVKETLRMHPPTVFLLPRKANNDVELYGYVVPKNAQVFVNVWAISRDPNHWENPNSFSPERFIEHEIDMKGQDFGLIPFGAGRRICPGDMLAFRMLNLMLGSLLHGFNWKVGDGI-- 473
Bv1_001300_sidm        LQAIIKETLRMHPPTVFLLPRKANNDVKLYGYIVPKNAQIFVNLLAISRDPTHWKNPDLFSPERFFELEIDLKGHDFGFIPFGAGRRTCPGDTLAFRMLNLMLGSLLHCFNWTFRD---- 471
Bv9_228610_yqeq        LQGIIKETLRLHPPTVFLLPRKADNDVELYGYVVPKNAQVLVNLWAIGRDPKVWKNPEVFSPERFLDCNIDYKGRDFELLPFGAGRRICPGLTLAYRMLNLMLATLLQNYNWKLEDGI-- 470
Bv9_228860_ickx        LQAIIKETLRLHPPTVFLLPRKADADVELYGYIVPKNAQVLVNLWAIGRDPKVWKNPEVFSPERFLESNIDYKGRDFELLPFGAGRRICPGLTLAYRMLNLMMANFLHSYDWKLEDGM-- 432
CYP76AD1               LQAIIKETLRLHPPTVFLLPRKADTDVELYGYIVPKDAQILVNLWAIGRDPNAWQNADIFSPERFIGCEIDVKGRDFGLLPFGAGRRICPGMNLAIRMLTLMLATLLQFFNWKLEGDI-- 468
                       :*.::** :*:** : **:*: *  *: * .*:***:: : **: :*.*.*. * :.. *****.:  :** **: * ::***:*** ***  ** **  : :..:*: :.*.

Bv1u_022460_qtnn       SPKDIDMKEKFGLTLQKAQPLQAIPIPR- 488
Bv1u_022470_qtfp       SPEDIDMEEKFGITLQKVEPLQAIPISR- 488
Bv1u_022400_apow       ATCTIDVEEKFGITLQKAEPLQAIPLPR- 610
Bv1_001290_xkhk        SPQDLDMTDKFGITIQKAIPLRALPIPK- 501
Bv1_001300_sidm        -DEDLDMNDKFGITIQKAKPLHVIPISKL 499
Bv9_228610_yqeq        NPKDLDMDEKFGITLQKVKPLQVIPVPRN 499
Bv9_228860_ickx        HPKDLDMDEKFGITLQKVKPLQVIPVPRK 461
CYP76AD1               SPKDLDMDEKFGIALQKTKPLKLIPIPRY 497
                           :*: :***:::**. **: :*:.:
```

b.

	1	2	3	4	5	6	7	8
1. Bv1u_022460_qtnn		74.2%	66.8%	56.6%	55.8%	54.2%	55.9%	53.4%
2. Bv1u_022470_qtfp	74.2%		67.6%	55.1%	55.6%	54.7%	56.1%	51.5%
3. Bv1u_022400_apow	66.8%	67.6%		50.9%	50.6%	52.0%	53.7%	51.2%
4. Bv1_001290_xkhk	56.6%	55.1%	50.9%		80.0%	58.3%	60.3%	56.3%
5. Bv1_001300_sidm	55.8%	55.6%	50.6%	80.0%		56.9%	59.8%	55.9%
6. Bv9_228610_yqeq	54.2%	54.7%	52.0%	58.3%	56.9%		87.9%	71.3%
7. Bv9_228860_ickx	55.9%	56.1%	53.7%	60.3%	59.8%	87.9%		71.6%
8. CYP76AD1	53.4%	51.5%	51.2%	56.3%	55.9%	71.3%	71.6%	

Figure 2-20. Multiple sequence aligment of CYP76AD1 with its 7 closest paralogs in *B. vulgaris*. (**a**) The F309 residue is highlighted in yellow. Note that all paralogs contain the F309L substitution identified by PCR mutagenesis and screening. Expression of Bv9_228610_yqeq or Bv9_228860_ickx in *S. cerevisiae* did not lead to betaxanthin production. Amino acids are color-coded based on their properties (red = small; blue = acidic; magenta = basic; green = hydroxyl, sulfhydryl, amine; grey = other). Asterisks indicate fully conserved residues, colons and periods indicate strong and weak conservation respectively. (**b**) Percent identities between paralogs.

Figure 2-21. Homology model of CYP76AD1 threaded onto human microsomal CYP1A2. (The CYP76AD1 protein sequence was uploaded to ModBase servers (http://modbase.compbio.ucsf.edu) for processing with the ModPipe pipeline. CYP76AD1 was aligned to human CYP1A2 (PDB ID: 2HI4) with a ModPipe Quality Score (MPQS) of 1.24059. F309 is highlighted in yellow, while CYP1A2's substrate, α-naphthoflavone, is highlighted in green and heme is highlighted in red. F309 appears to be fairly distal from where L-tyrosine or L-DOPA might bind.

2.2.5. Production of (*S*)-reticuline from glucose

Having achieved efficient dopamine production from L-tyrosine in *S. cerevisiae*, we sought to extend the pathway towards downstream BIA intermediates. The first committed step in BIA biosynthesis is the formation of the backbone molecule (*S*)-norcoclaurine via condensation of dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA), a reaction that is catalyzed by NCS (**Figure 2-22a**). 4-HPAA is produced endogenously in *S.*

cerevisiae as an intermediate of the Ehrlich pathway for amino acid catabolism[49]. In this pathway, L-tyrosine is converted to 4-HPAA through the sequential action of Aro8p/Aro9p and Aro10p prior to being broken down into tyrosol or 4-hydroxyphenylacetate (4-HPA) by a host of redundant enzymes.

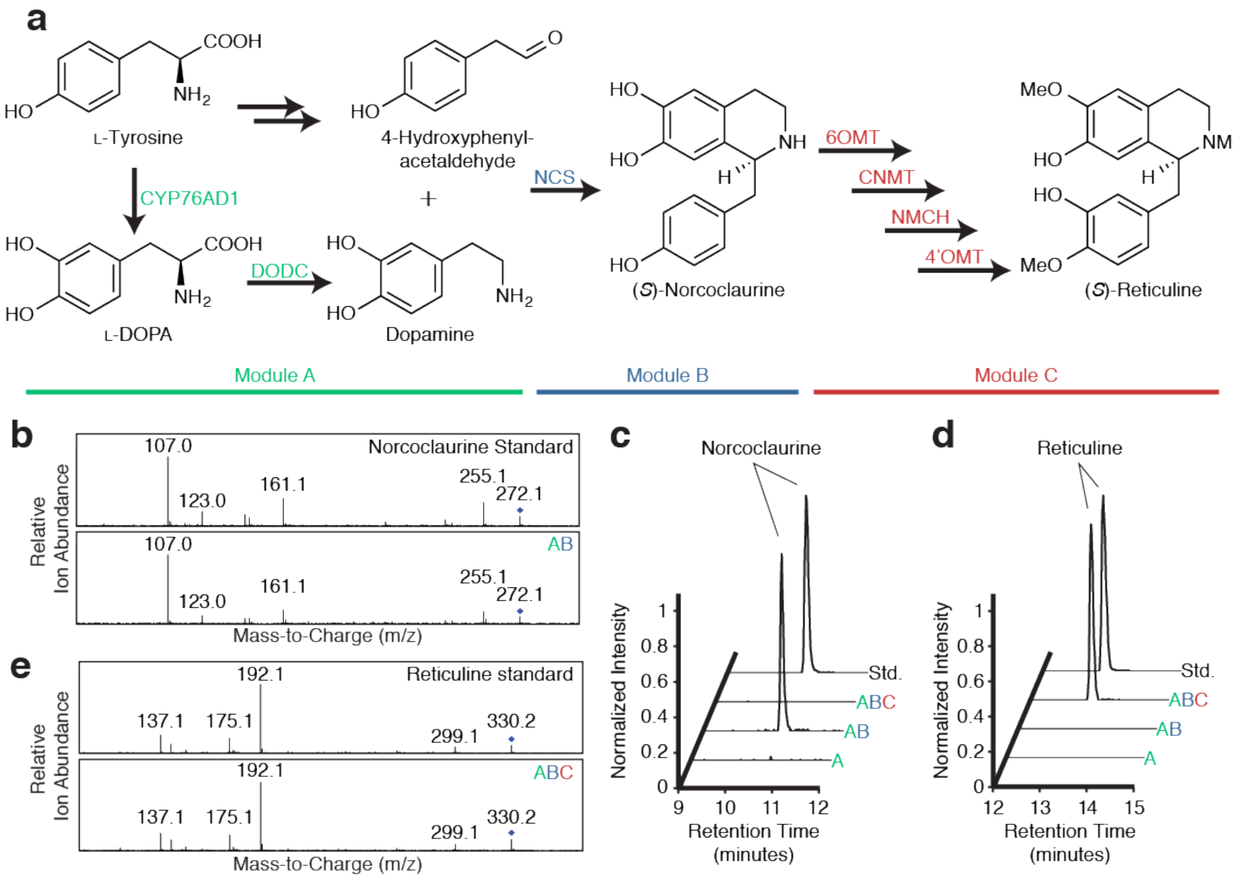


Figure 2-22. Production of (*S*)-reticuline from glucose. (**a**) The (*S*)-reticuline pathway was divided into three modules to facilitate analysis. Because 4-hydroxyphenylacetaldehyde (4-HPAA) is produced endogenously in *S. cerevisiae* (unlabeled black arrows), enzymes catalyzing 4-HPAA synthesis from L-tyrosine were not heterologously expressed. CYP76AD1, tyrosine hydroxylase (W13L/F309L double mutant); DODC, DOPA decarboxylase; NCS, norcoclaurine synthase; 6OMT, 6-*O*-methyltransferase; CNMT, coclaurine *N*-methyltransferase; NMCH, *N*-methylcoclaurine hydroxylase (CYP80B1); 4′OMT, 4′-*O*-methyltransferase. See **Figure 2-23** for a more detailed pathway diagram of module C. (**c, d**) LC/MS analysis of norcoclaurine (**c**) and reticuline (**d**) in the supernatant of strains expressing modules A, AB, and ABC after 48 hours of growth in synthetic media with 4% glucose. Strain AB is engineered to produce norcoclaurine, and strain ABC is engineered to produce reticuline. Traces are normalized to the maximum peak height across all three samples. Std. denotes a 5 μM chemical standard, which was normalized separately. (**b, e)** Tandem mass spectra of norcoclaurine from strain AB (**b**) and reticuline from strain ABC (**e**) confirm their identity in comparison to chemical standards. Parent ions are marked with ♦.

From (*S*)-norcoclaurine, four additional enzymatic steps are needed to produce (*S*)-reticuline, the last shared intermediate of the major BIA pathway branches (**Figure 2-23**). Three of these enzymes catalyze methylation reactions: 6-*O*-methyltransferase (6OMT), coclaurine *N*-methyltransferase (CNMT), and 4'-*O*-methyltransferase (4'OMT). Variants of each methyltransferase from *P. somniferum* have been demonstrated to function in *S. cerevisiae*[24]. The remaining enzyme is the cytochrome P450 NMCH, *N*-methylcoclaurine hydroxylase (CYP80B1). Previous work to synthesize reticuline in *S. cerevisiae* circumvented this enzyme by feeding a hydroxylated derivative of norcoclaurine, norlaudanosoline, which is commercially available, albeit expensive. NMCH has, however, been isolated from the California poppy (*Eschscholzia californica*) and successfully expressed and purified heterologously from *Spodoptera frugiperda* Sf9 cells[50].

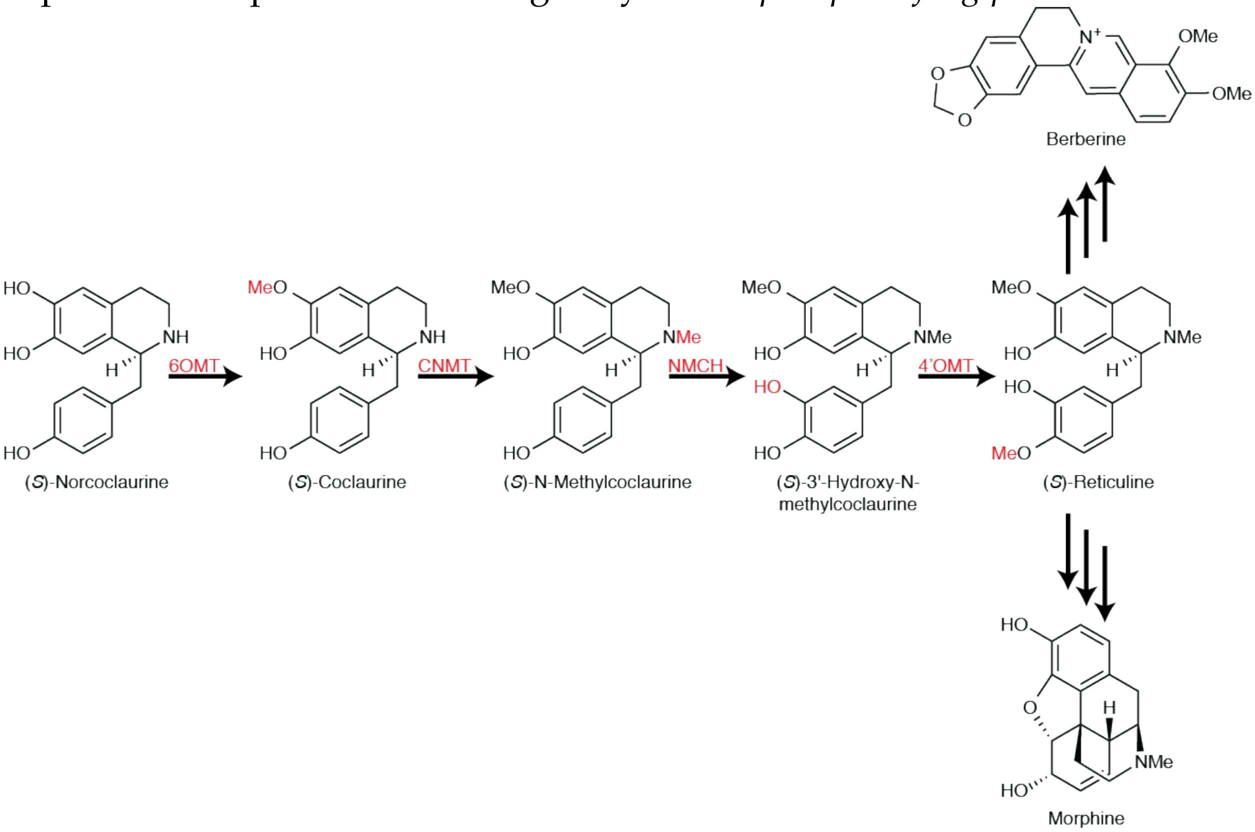


Figure 2-23. Pathway from (*S*)-norcoclaurine to (*S*)-reticuline. (*S*)-reticuline is the last shared intermediate between the morphinan (down) and sanguinarine/berberine (up) branches in BIA biosynthesis. 6OMT, 6-*O*-methyltransferase; CNMT, coclaurine *N*-methyltransferase; NMCH, *N*-methylcoclaurine hydroxylase (CYP80B1); 4'OMT, 4'-*O*-methyltransferase.

We divided the (*S*)-reticuline pathway into three modules (**Figure 2-22a**). Module A combines our best tyrosine hydroxylase mutant (CYP76AD1 W13L/F309L) and DODC to produce dopamine. We relied on endogenous yeast enzymes for 4-HPAA production, although our strains did overexpress feedback-insensitive Aro4p, which increases flux through 4-HPAA in addition to increasing L-tyrosine levels. Module B included an NCS from *P. somniferum* (accession number KP262411), which was selected after screening four NCS variants with homology to a *Thalictrum flavum* NCS (accession number ACO90248.1) for activity in *S. cerevisiae* (**Figure 2-24**)[51]. Module C consists of the four enzymes required for the conversion of (*S*)-norcoclaurine to (*S*)-reticuline, including 6OMT, CNMT, and 4'OMT from *P. somniferum* and NMCH from *E. californica*. We constructed three strains for testing, which we refer to here by the pathway modules that they express (A, AB, and ABC). All heterologous genes in these strains are expressed using high strength promoters and are either integrated into the chromosome (modules A and B) or maintained on a low-copy plasmid (module C).

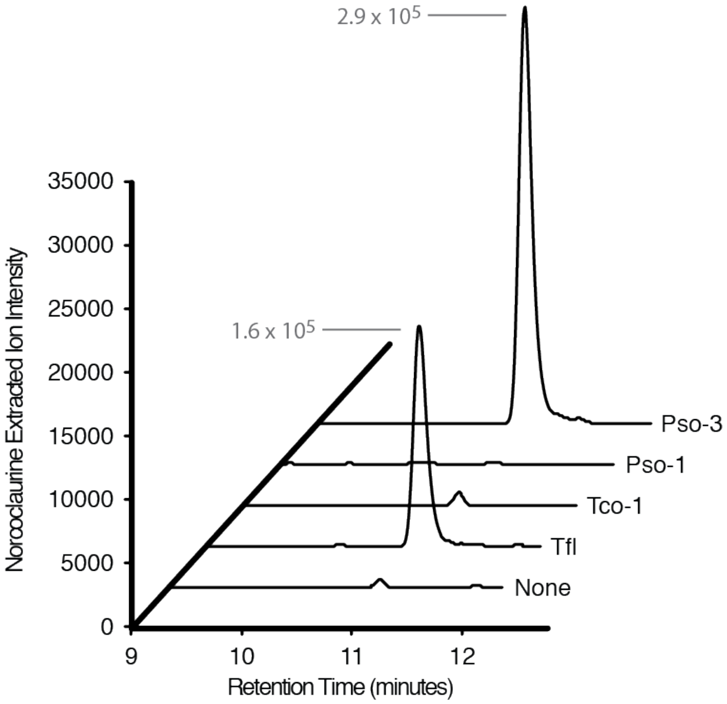


Figure 2-24. Comparison of norcoclaurine production with different norcoclaurine synthases (NCS). Yeast strains expressing CYP76AD1 W13L/F309L, DODC, and Aro4^{FBR} as well as an NCS variant were grown in synthetic dropout medium (Strains 51, 52, 53, 54, 31). Supernatants were collected and analyzed via LC/MS with extraction of norcoclaurine $[M+H]^+$ ion at *m/z* 272.13. If detected, norcoclaurine peak areas are included in grey. The *Papaver somniferum* NCS Pso-3 (Accession number KP262411) showed the greatest production of norcoclaurine and was chosen for further experimentation. (Tfl, *Thalictrum flavum*; Pso, *Papaver somniferum*; Tco, *Tinospora cordifolia*)

Strain AB produced easily detectable levels of norcoclaurine when grown in synthetic complete media with glucose (**Figure 2-22b,c**). Given that this product was not observed in the supernatant of strain A, norcoclaurine synthesis (which can occur spontaneously in other systems[23,52]) appeared to be dependent on NCS expression. While spontaneously formed norcoclaurine is racemic, NCS is known to be stereoselective, producing only (*S*)-norococlaurine[53]. Chiral analysis confirmed that all norcoclaurine produced by strain AB was the (*S*)-enantiomer (**Figure 2-25a-c**). (*S*)-Reticuline was also clearly detected in the supernatant of strain ABC (**Figure 2-22d,e** and **Figure 2-25d-f**). The (*S*)-reticuline-producing strain had no detectable (*S*)-norcoclaurine in the supernatant, indicating that the conversion efficiency from (*S*)-norcoclaurine to (*S*)-reticuline was high. To quantify titers, a 96-hour fermentation was performed in shake flasks for strains AB and ABC (**Figure 2-26**). Maximum titers for (*S*)-norcoclaurine and (*S*)-reticuline were 104.6 μg/l and 80.6 μg/l respectively and did not show substantial increases after cell saturation. Intracellular product levels were also measured using extraction with methanol. Less than 3% of total (*S*)-norcoclaurine/(*S*)-reticuline was observed in the cell pellet, indicating that these products readily escape into the media.

Thus, by combining our engineered tyrosine hydroxylase with a newly identified NCS, we demonstrated biosynthesis of the major BIA branch point (*S*)-reticuline in the industrial workhorse *S. cerevisiae*. This advance establishes a microbial platform in which all BIAs can be synthesized, accelerating the discovery of new enzymes/pathways and the commercial deployment of this technology for industrial scale production.

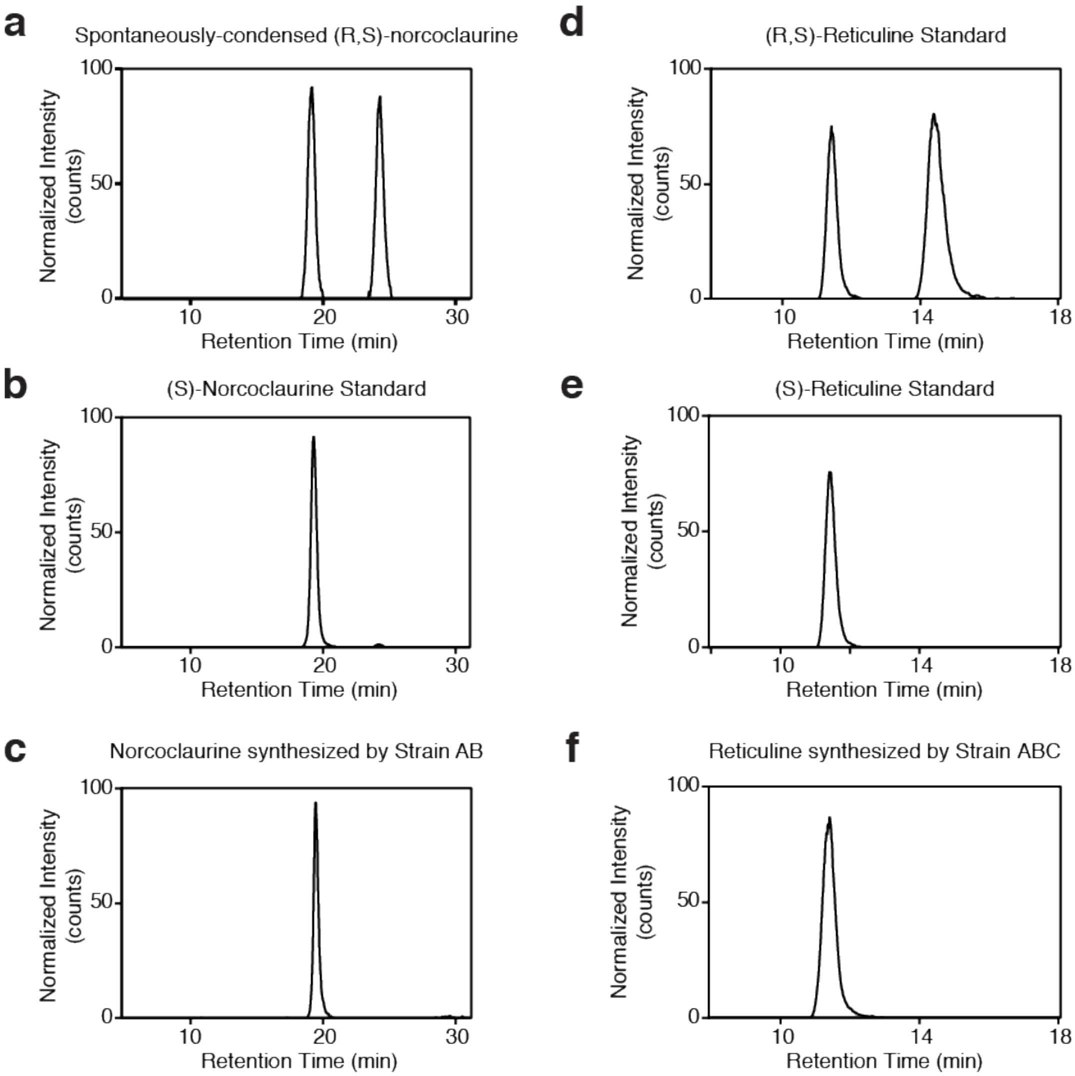


Figure 2-25. Chiral analysis of norcoclaurine produced by Strain AB and reticuline produced by Strain ABC. LC/MS traces of (**a**) (*R*, *S*)-norcoclaurine from spontaneously-condensed dopamine and 4-HPAA, (**b**) an authentic (*S*)-norcoclaurine standard, (**c**) supernatant from the norcoclaurine-producing strain AB, (**d**) a racemic (*R*,*S*)-reticuline standard, (**e**) an authentic (*S*)-reticuline standard, and (**f**) supernatant from the reticuline-producing strain ABC. A chiral column was used to separate (*R*)- and (*S*)-enantiomers, demonstrating that strain AB produces only (*S*)-norcoclaurine and strain ABC produces only (*S*)-reticuline. Strains used were 32, 33.

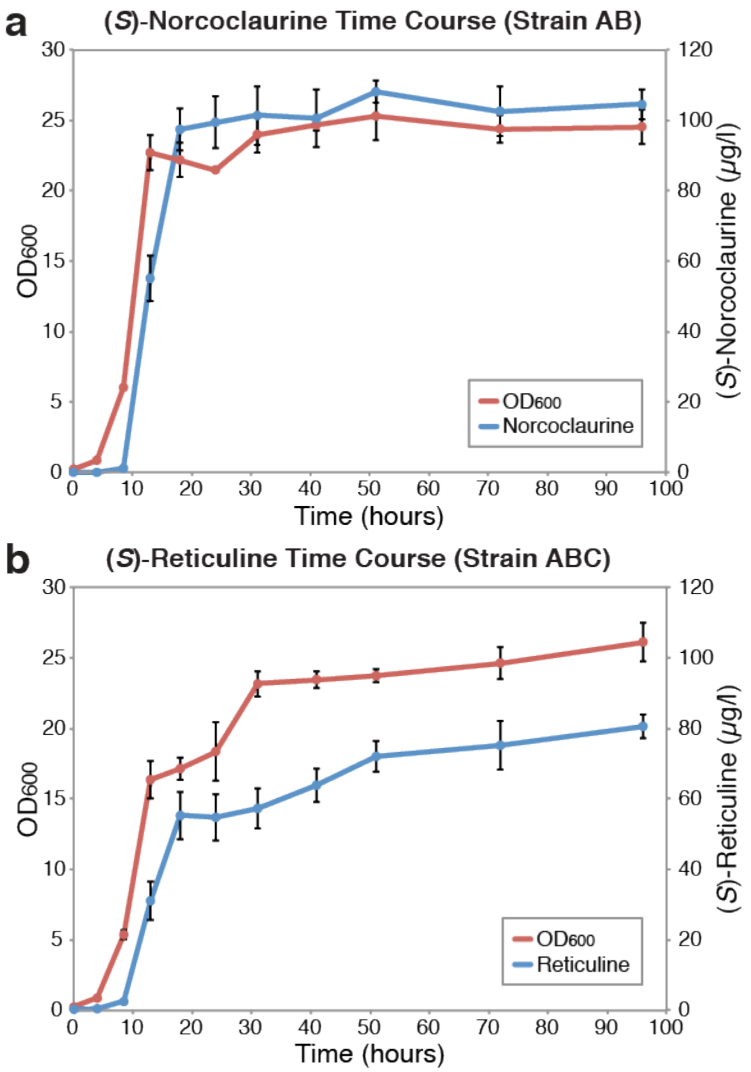


Figure 2-26. Time courses for norcoclaurine and reticuline production. Strains producing (*S*)-norcoclaurine (**a**) or (*S*)-reticuline (**b**) were grown for 96-hours in shake flasks with 2X synthetic dropout media and 4% glucose. Measurements of culture optical density (OD600, red lines) and product titer (blue lines) were taken periodically. Product titer was measured by LC/MS of culture supernatants and comparison to a standard curve. Final (*S*)-norcoclaurine and (*S*)-reticuline titers were 104.6 μg/l and 80.6 μg/l respectively. Data points represent mean values ± s.d. of three biological replicates. Strains used were 32, 33.

2.3. Discussion

Microbial production of plant-derived therapeutics promises to transform both pharmaceutical manufacturing and discovery. BIAs are an attractive engineering target given their widespread use and potential to exhibit new pharmacological properties. The length and complexity of the BIA pathway also presents a formidable metabolic engineering challenge that is likely to spur the development of new tools and techniques for controlling the metabolism of microorganisms[26]. We report a significant breakthrough in microbial BIA production: the development of a yeast strain synthesizing the key intermediate (*S*)-reticuline from glucose. While initial titers are low (80.6 μg/l), this first generation strain marks a starting point for future optimization, which we expect to eventually lead to an alternative, cost-effective method for producing high-value BIAs, thereby dispensing with the need for large-scale dedicated crop cultivation.

(*S*)-Reticuline production from central metabolites was enabled by a novel strategy for intracellular monitoring of L-DOPA that utilized a plant enzyme to convert L-DOPA into a highly fluorescent small molecule—betaxanthin—that is easy to detect. This tool helped us identify a previously unknown activity in a cytochrome P450 from beets (CYP76AD1) that filled a longstanding gap in early BIA biosynthesis in yeast. While wildtype CYP76AD1 had a second activity (DOPA oxidase) that introduces an undesirable side pathway towards melanin, we used our L-DOPA biosensor to identify a point mutation that suppresses DOPA oxidase activity and increases flux towards (*S*)-reticuline.

While we utilized the betaxanthin pathway as an enzyme-coupled L-DOPA biosensor to optimize tyrosine hydroxylase activity, we believe that this pathway could prove more generally useful. Because of its fluorescent, water-soluble product, the betaxanthin pathway could be a valuable tool for studying subcellular enzyme localization and metabolite transport, or as a model pathway for the development of new high-throughput screening techniques that are amenable to FACS. By linking betaxanthin fluorescence to the production of other metabolites of interest (e.g. L-tyrosine), it might be possible to develop high-throughput screens for additional molecules. This work is an example of the rapid progress that is enabled by robust metabolite biosensors.

Further optimization of (*S*)-reticuline titer will be required before production of downstream BIAs is feasible. While we achieved a dopamine titer of 23.8 mg/l, we observed 228-fold lower production of (*S*)-norcoclaurine (104.6 μg/l). It is likely that yeast's endogenous enzymes are largely to blame for inefficiencies at this step. More than twenty potential Ehrlich pathway enzymes could be contributing to the breakdown of 4-HPAA into the side products tyrosol and 4-HPA[54]. Additionally, the majority of dopamine produced by our strain was found in the supernatant, suggesting that this intermediate is secreted by a yet to be identified transporter. Finally, carefully balancing the production of dopamine and 4-HPAA, both of which are derived from L-tyrosine, may lead to increased yields of norcoclaurine. Addressing these and other endogenous factors that limit flux will be important for future strain optimization.

Here we have described a foundational advance in microbial BIA synthesis by linking production of the key intermediate (*S*)-reticuline to yeast central metabolism. As metabolic engineering and synthetic biology tools continue to improve, progress towards a strain capable of high-yielding BIA fermentation from readily available feedstocks will accelerate. Due to the potential for illicit use of these products, including morphine and its derivatives, it is critical that appropriate policies for controlling such strains be established so that we can garner the considerable benefits while minimizing the potential for misuse.

2.4. Extensions and Future Directions

2.4.1. Identification of PABA-betaxanthin

Initially, when we added L-DOPA-supplemented synthetic media to a yeast strain expressing DOD, we observed a stark, fluorescent-orange product where we only expected to see yellow betaxanthin or, with CYP76AD1, red-violet betanidin. This product was unexpected, because, although it is known that betalamic acid can conjugate with many primary and secondary amines, no orange products have been reported in the literature[34,36,37]. We took a sample of the supernatant and analyzed it with LC/MS, finding a mass that corresponded to PABA-betaxanthin around the time of a particularly strong 540 nm absorbance (**Figure 2-27**).

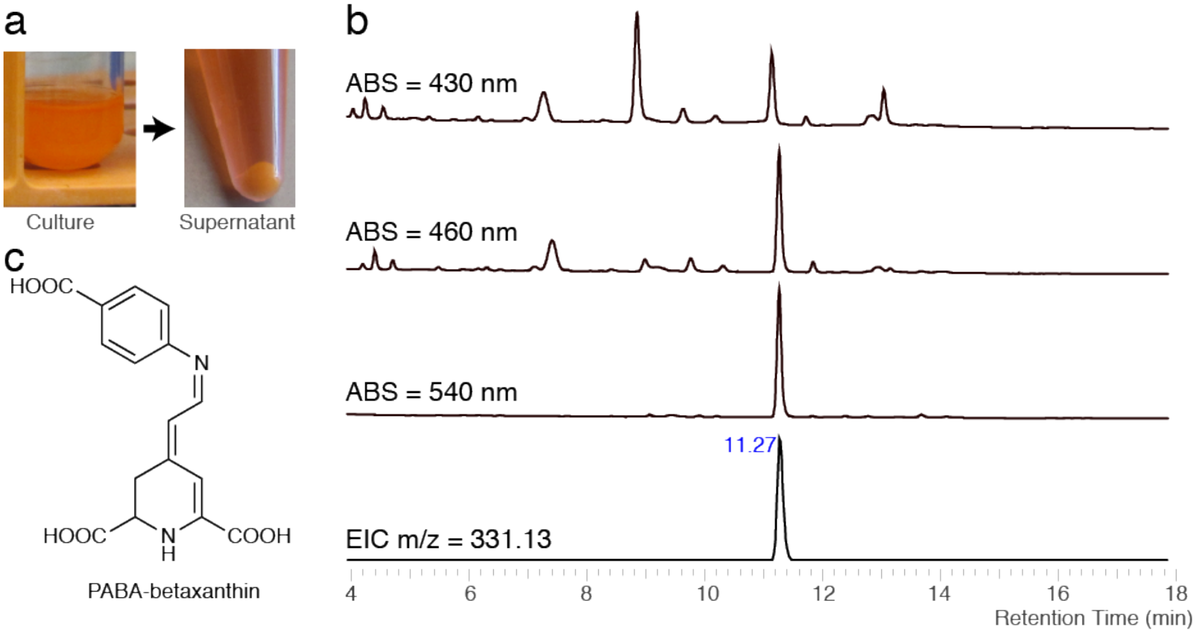


Figure 2-27. Discovery and detection of PABA-betaxanthin in culture supernatant. (**a**) DOD-expressing Strain 3 was grown with 1 mM L-DOPA overnight. The culture was separated by centrifugation, and the supernatant transferred to LC/MS. (**b**) A peak showing broad absorbance, especially at 540 nm, corresponded to a [M+H]+ of 331.13. (**c**) Molecular formula calculations suggested PABA-betaxanthin as a potential match.

For validation, we generated roughly 1 mM betalamic acid *in vitro* using purified DOD following established protocols[35]. We then added 1 mM of different amines to the betalamic acid to see which would generate new colors, pictured in **Figure 2-28** and further characterized in **Figure 2-29** and **Figure 2-30**.

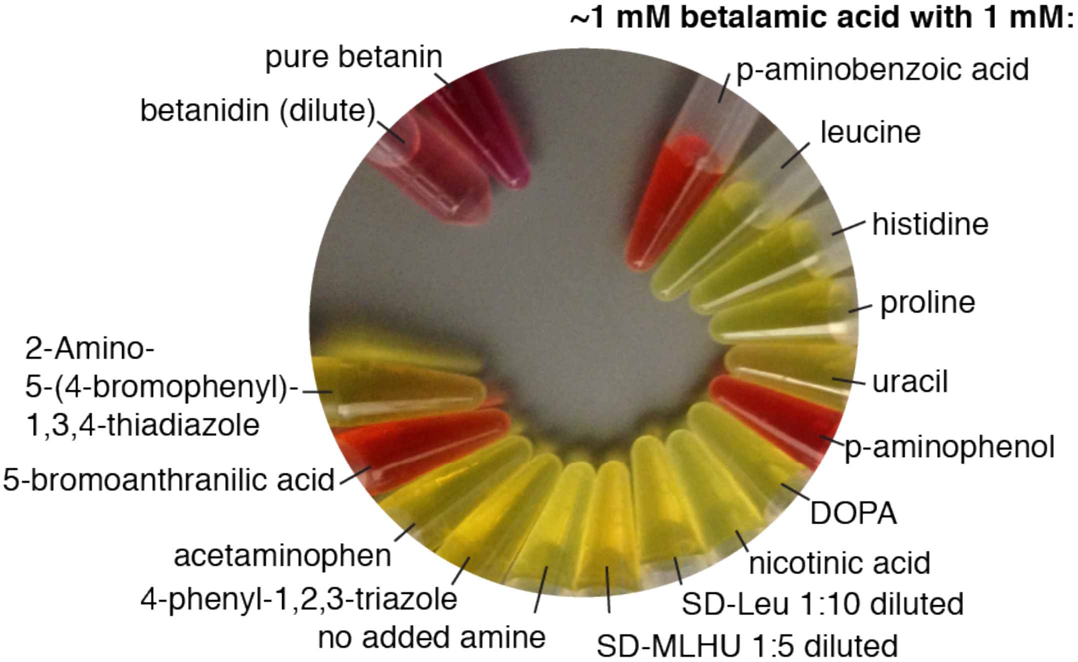


Figure 2-28. Betalamic acid with potential conjugation partners. Approximately 1 mM of betalamic acid was generated *in vitro* using purified DOD and 1 mM of a nitrogen-containing potential conjugation partner was added, with the exception of SD-Leu and SD-MLHU, which were fresh culture media added to 10% and 20% of the final volume, respectively. The betanin standard was pure (no added betalamic acid) and betanidin was generated from betanin using a β-glucosidase.

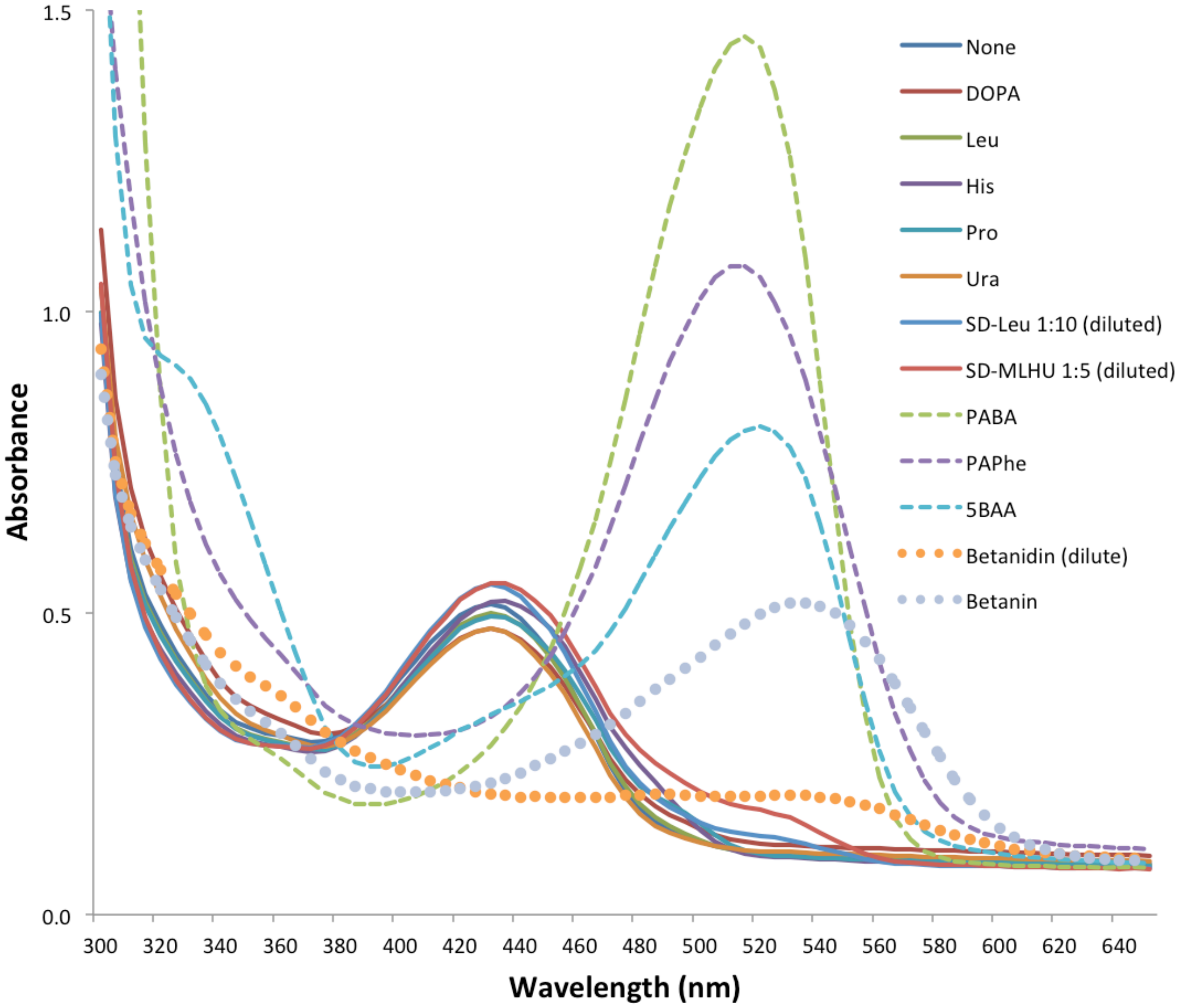


Figure 2-29. Absorbance scan of new betalains. Note that not necessarily all potential partners, e.g. uracil (Ura) are able to form betalain conjugates.

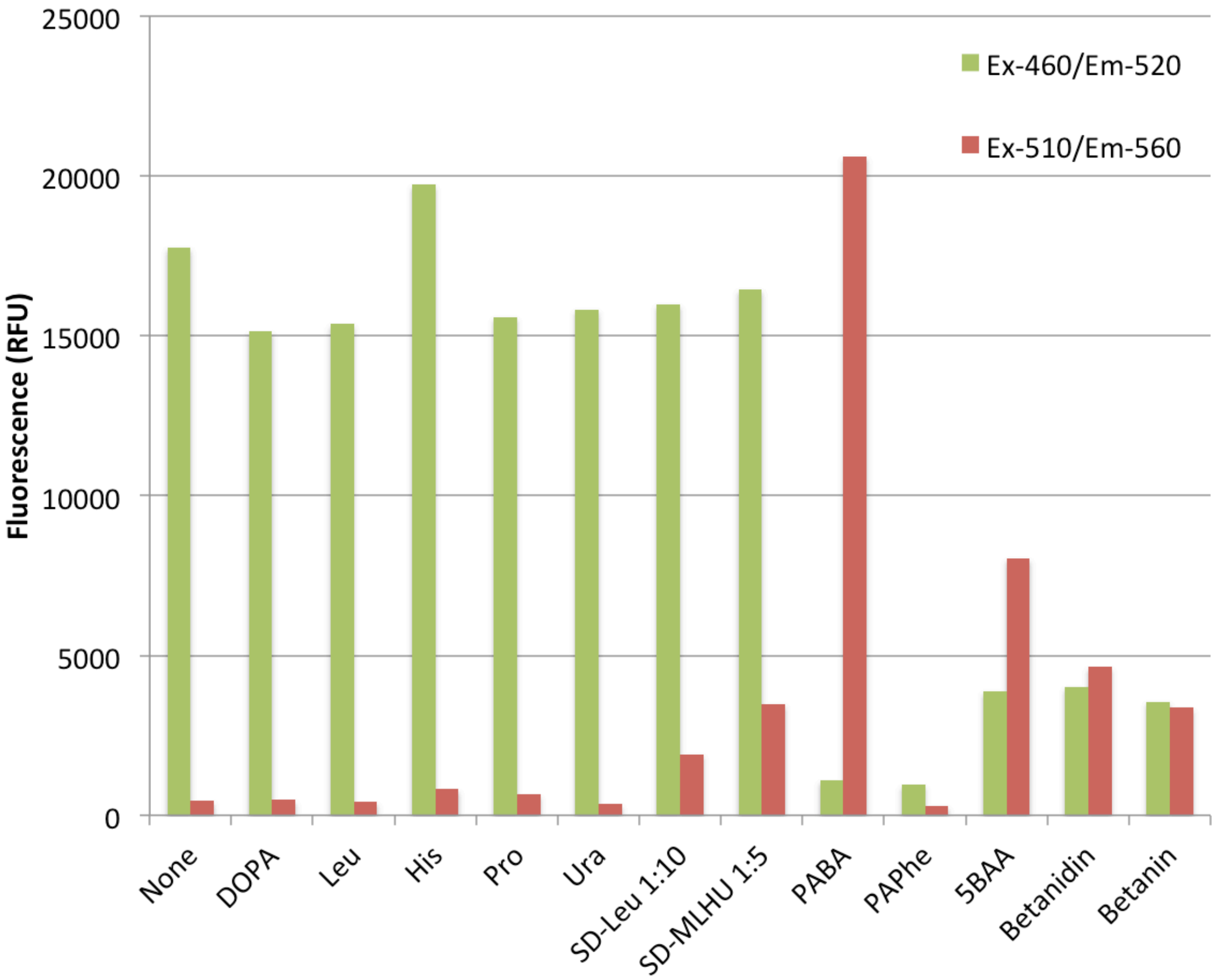


Figure 2-30. Aromatic amine-derived betalains show red fluorescence. Note that not necessarily all potential partners, e.g. uracil (Ura) are able to form betalain conjugates.

Interestingly, orange products were generated with aromatic amines, such as *p*-aminobenzoic acid (PABA, product was orange and fluorescent), *p*-aminophenol (PAPhe, product was orange and apparently nonfluorescent), and 5-bromo-anthranilic acid (5BAA, product was orange and fluorescent). Taken together with betanidin's reddish color and aromatic amine, it suggests that a range of red and orange hues can be achieved with betalains by offering an sp^2-adjacent or aromatic amine for betalamic acid to conjugate with. Since the constituent molecules (betalamic acid and PABA) are naturally occurring and food- or skin-safe, these new molecules may offer an exciting new range of natural alternatives to synthetic dyes.

2.4.2. Confirmation of natural tyrosine hydroxylases

Previous knockout studies in *Beta vulgaris* suggested that CYP76AD1 couldn't be the primary tyrosine hydroxylase that beets use to generate color[42], but our results, especially the F309L reactivity switch, suggested that there was a good chance the native tyrosine hydroxylase was a P450 as well, likely a paralog of CYP76AD1. We found several possible P450s in the *Beta vulgaris* genome[47] with the F309L mutation (**Figure 2-20**), and one of them, CYP76AD6, was recently confirmed by another group to have native tyrosine hydroxylase activity[55,56]. Our copy of CYP76AD6 (called "yqeq") from the genome sequencing project, however, had four amino acid mutations relative to the reported copy[55] – T28S, K164E, A262S, S325T – and we saw no activity with our database-derived sequence. To validate this result, we tested our engineered CYP76AD1 W13L/F309L against the CYP76AD6 reported in the literature and our database-derived CYP76AD6 in 2% glucose PABA-free minimal media supplemented with 2 mM tyrosine and 10 mM ascorbate. As **Figure 2-31** shows, the reported CYP76AD6 is indeed far more efficient than CYP76AD1 at converting tyrosine to L-DOPA while avoiding DOPA oxidase activity. Thus, CYP76AD6 offers a promising option for boosting microbial yields of BIAs.

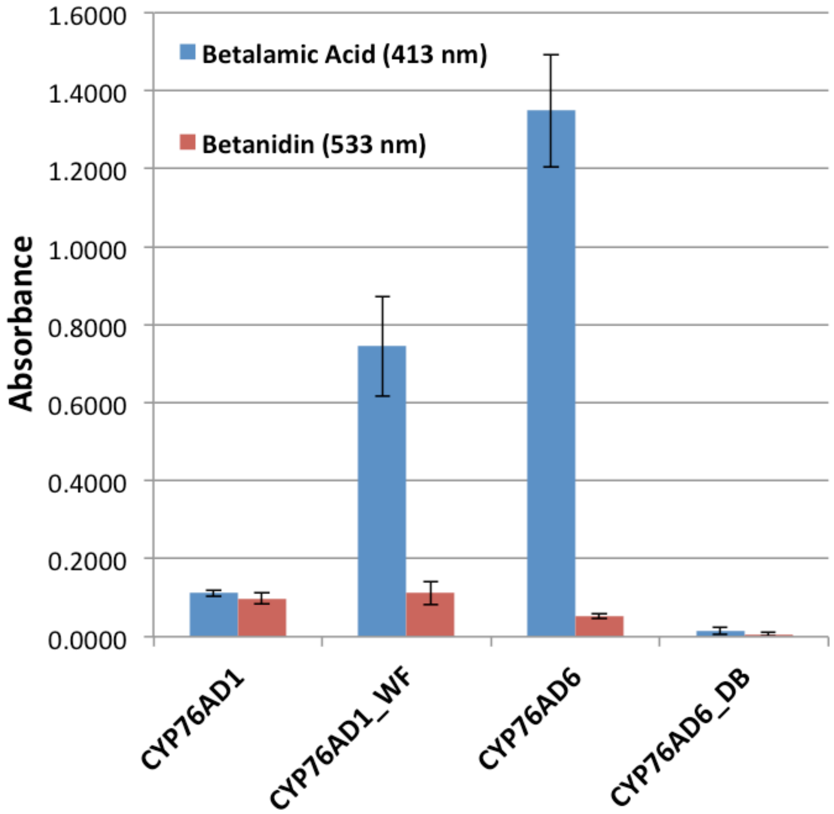


Figure 2-31. CYP76AD6 outperforms even the CYP76AD1 double mutant. Cultures of yeast expressing DOD and either CYP76AD1, CYP76AD1 W13L/F309L (WF), CYP76AD6, or CYP76AD6 T28S/K164E/A262S/S325T from a database (DB) were grown in 2% glucose PABA-free minimal media with 2 mM tyrosine and 10 mM ascorbate for 24 hours. Error bars indicate mean ± s.d. of four biological replicates.

2.4.3. Elucidating the relationship between CYP76AD1/6 and reductase

One of the questions that remained after our manuscript was how CYP76AD1 (and CYP76AD6) is receiving the reducing equivalents necessary for it to complete its catalytic cycle. Cytochromes P450 typically receive reducing equivalents from NADPH-dependent P450 reductase (CPR or POR, in yeast NCP1; in *Beta vulgaris* ATR1 and ATR2) and NADH-dependent cytochrome *b5* (CYB5, in yeast CYB5)[57,58].

The first step in determining which reductases are responsible for CYP76AD1's activity is determining whether the P450 performs identically in a knockout of the reductase of interest. Knockouts of NCP1 proved hard to work with, however, as the membrane defects of sterol deficiency left the yeast extremely difficult to transform, even with sterol supplementation – a problem that has been reported previously[59]. However, the cytochrome *b5* knockouts were more workable, though neither CYP76AD1, its PD1-solubilized variant (covered in the next section), or CYP76AD6 seemed to be affected (**Figure 2-32**). This suggests that CYB5 is not responsible for delivering electrons to CYP76AD1.

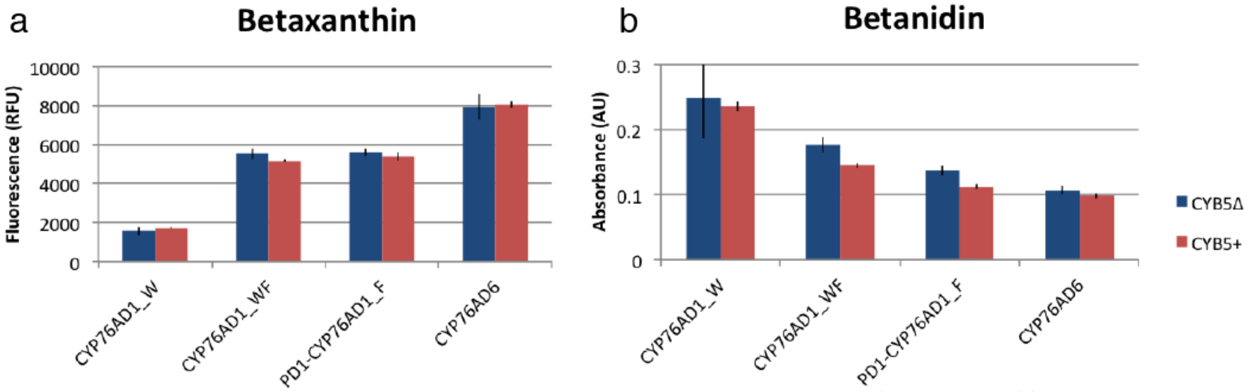


Figure 2-32. Cytochrome b5 (CYB5) deletion does not significantly affect CYP76AD1 or CYP76AD6. Cultures of yeast expressing DOD and either CYP76AD1 W13L (W), CYP76AD1 W13L/F309L (WF), PD1-CYP76AD1 F309L or CYP76AD6 in a background of either CYB5 deletion (CYB5Δ) or GAL2 deletion (CYB5+) were grown in 2% glucose PABA-free minimal media with 2 mM tyrosine and 10 mM ascorbate for 24 hours. (**a**) Betaxanthin fluorescence indicates tyrosine hydroxylase activity. (**b**) Betanidin absorbance indicates DOPA oxidase activity. Error bars indicate mean ± s.d. of six biological replicates.

The next step was to test CYP76AD1 responded to overexpression of possible reductase partners. Once again, cells were tested in 2% glucose PABA-free minimal media supplemented with 2 mM tyrosine and 10 mM ascorbate, this time with pCCW12-level overexpression of either CYB5 or NCP1 (**Figure 2-35**).

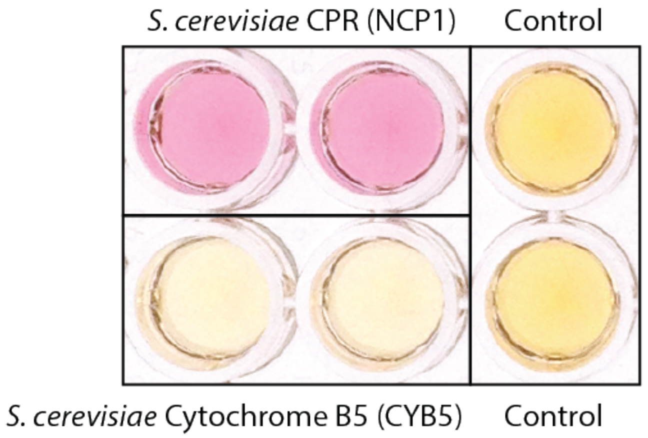


Figure 2-33. Overexpression of NCP1 stimulates DOPA oxidase activity while CYB5 suppresses all activity in CYP76AD1 W13L/F309L. Cell supernatants taken after 24 hours growth in minimal media are shown. Increased yellow color indicates tyrosine hydroxylase activity (betaxanthin) while increased red color indicates DOPA oxidase activity (betanidin). The control strain has wildtype levels of NCP1 and CYB5.

The suppression of both tyrosine hydroxylase and DOPA oxidase activity upon expression of yeast CYB5 has some precedent in the literature – CYB5 can competitively block CPR binding and prevent P450 catalysis if overexpressed[60–62]. The dramatic effect of expressing excess yeast CPR NCP1 overexpression suggested that NCP1 (and its homologs in *B. vulgaris*, ATR1 and ATR2) merited additional study. Strains expressing a mutant of CYP76AD1 or CYP76AD6 were transformed with NCP1, ATR1, or ATR2 under the control of the galactose-inducible promoter pGAL1 in the background of a GAL2Δ mutant to ensure titratability. This titration of CPR expression, shown in **Figure 2-34** and **Figure 2-35**, highlights several key points.

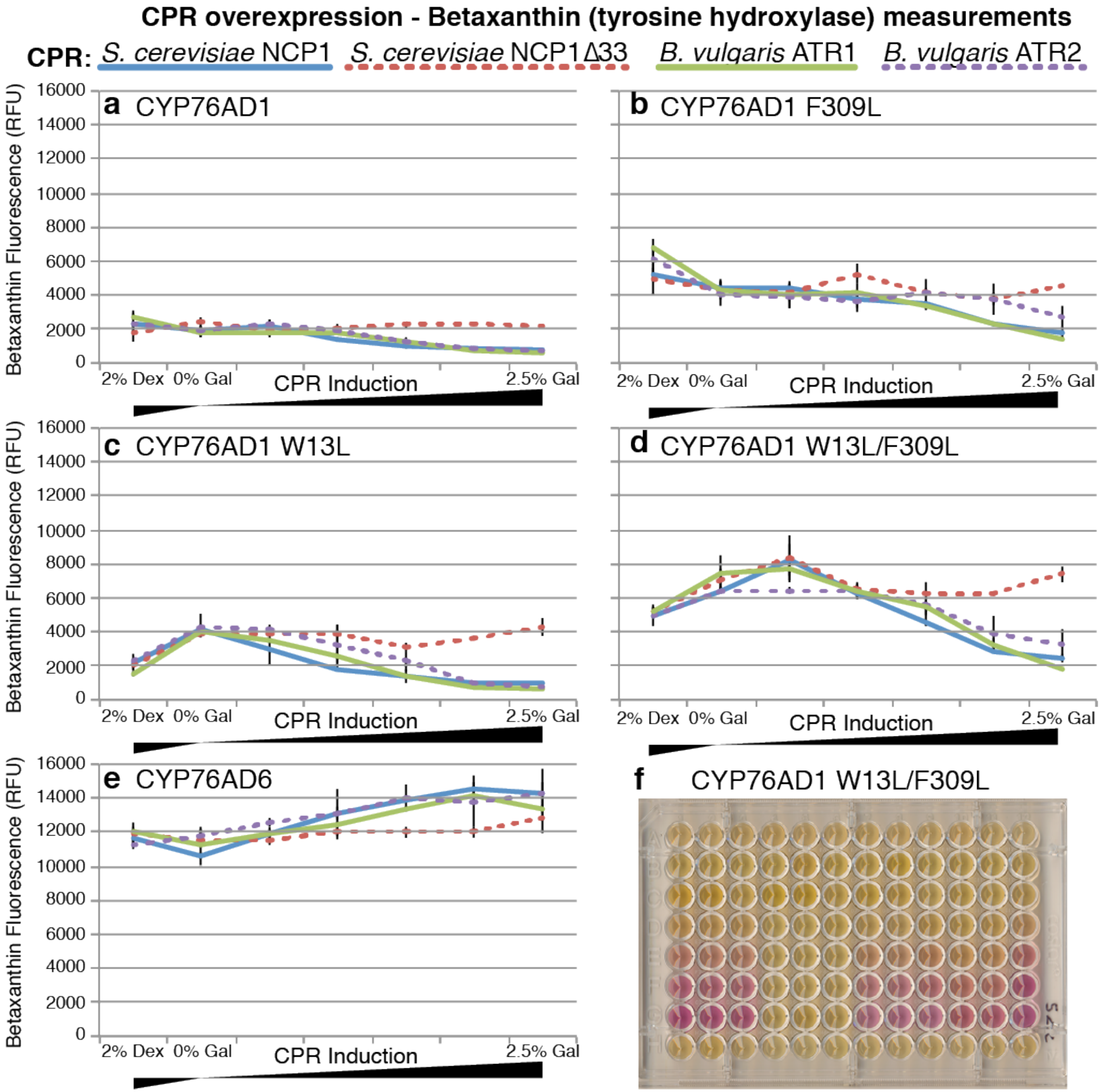


Figure 2-34. Titration of CPRs can push CYP76AD variants to make less betaxanthin. Yeast constitutively expressing the indicated CYP were transferred to 2% raffinose minimal media, 2 mM tyrosine, 10 mM ascorbate, with 0%, 0.01%, 0.04%, 0.16%, 0.625% or 2.5% galactose to induce production of a CPR. These cells were grown for 66 hours to account for slower growth on raffinose. Supernatants were taken, and fluorescence was measured; betaxanthin fluorescence increases as a product of tyrosine hydroxylase activity and is decreased by DOPA oxidase activity. (**a-e**) Induction curves of different CYP and CPR combinations. Error bars indicate mean ± s.d. of three biological replicates. (**f**) Supernatants measured in (**d**) transferred to a screening plate for visualization. Columns are organized in triplets: from left, NCP1, NCP1Δ33, ATR1, ATR2. Galactose concentration increases top to bottom.

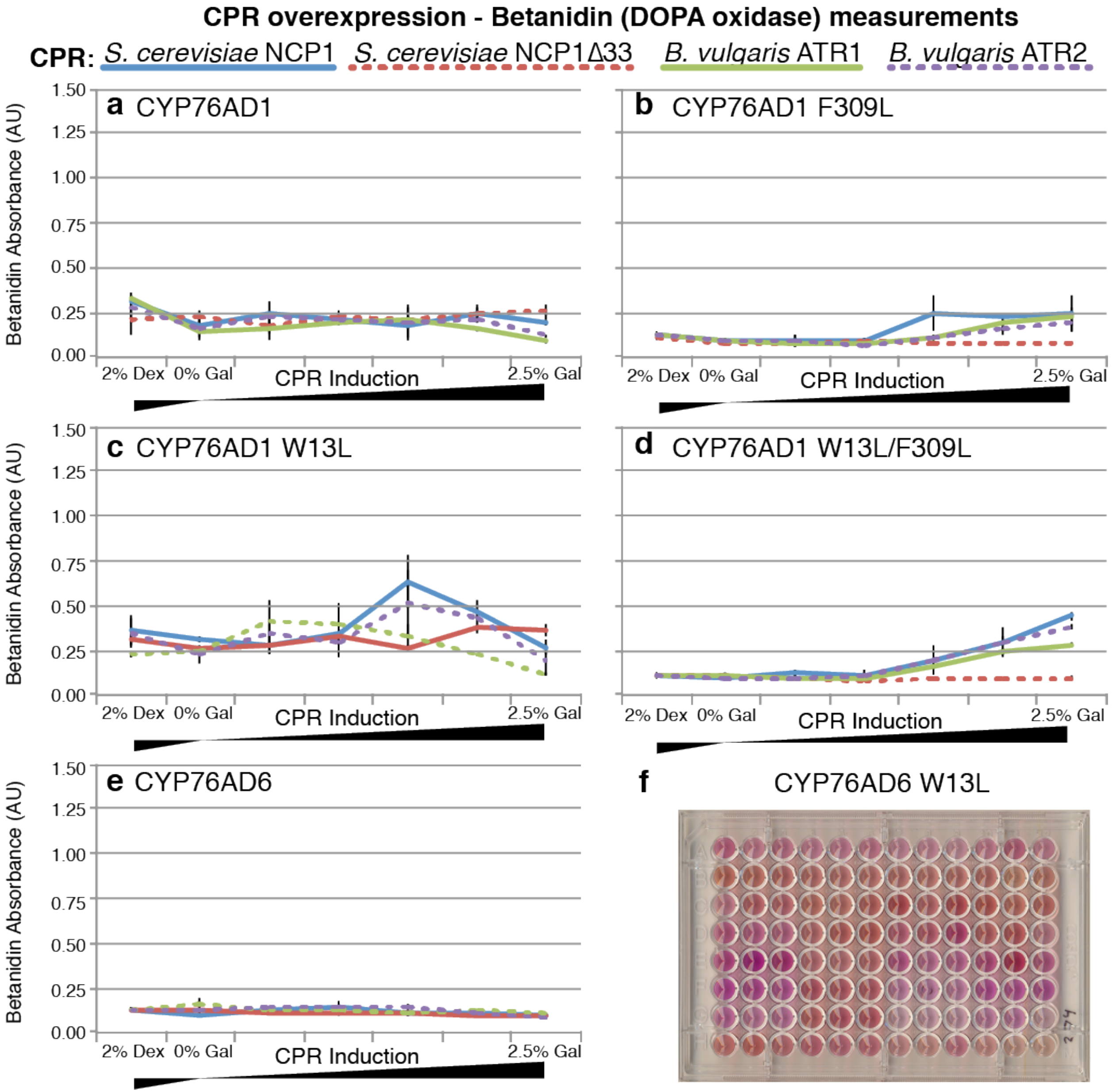


Figure 2-35. Titration of CPRs can push CYP76AD variants to make more betanidin. Yeast constitutively expressing the indicated CYP were transferred to 2% raffinose minimal media, 2 mM tyrosine, 10 mM ascorbate, with 0-2.5% galactose in fourfold increments to induce production of a CPR. These cells were grown for 66 hours to account for slower growth on raffinose. Supernatants were taken, and fluorescence was measured; betanidin absorbance increases as a product of DOPA oxidase activity, but also requires tyrosine hydroxylase product (L-DOPA) to form. Too much DOPA oxidase activity can deplete L-DOPA, resulting in a loss of color. (**a-e**) Induction curves of different CYP and CPR combinations. Error bars indicate mean ± s.d. of three biological replicates. (**f**) Supernatants measured in (**c**) transferred to a screening plate for visualization. Columns are organized in triplets: from left, NCP1, NCP1Δ33, ATR1, ATR2. Galactose concentration increases top to bottom.

First, all CYP76AD1 variants and even CYP76AD6 respond to CPRs by increasing DOPA oxidase activity, resulting in more betanidin. The response to CPRs seems to be a little stronger with NCP1, possibly because the yeast reductase is better adapted to work in yeast than ATR1 and ATR2 from beets. Secondly, for CYP76AD1 W13L/F309L and CYP76AD6, there appears to be a slight enhancement in tyrosine hydroxylase activity, which, for CYP76AD1 W13L/F309L, results in a brief window where additional CPR enhances production. This second point may require knockdown or knockout of endogenous NCP1 to fully investigate. Finally, for many of the mutants tested, there is a clear optimal reductase level for betanidin production, as betanidin requires both L-DOPA and dopaquinone in a 1:1 stoichiometry; overproduction of dopaquinone can deplete L-DOPA and result in loss of both colors, betaxanthin and betanidin. A final interesting point – although CYP76AD6 appears to have virtually no DOPA oxidase activity, as no betanidin can be detected via LC/MS, full induction of CPR expression (ATR2 in this case) can force CYP76AD6 to produce almost as much betanidin as CYP76AD1 W13L/F309L (**Figure 2-36**).

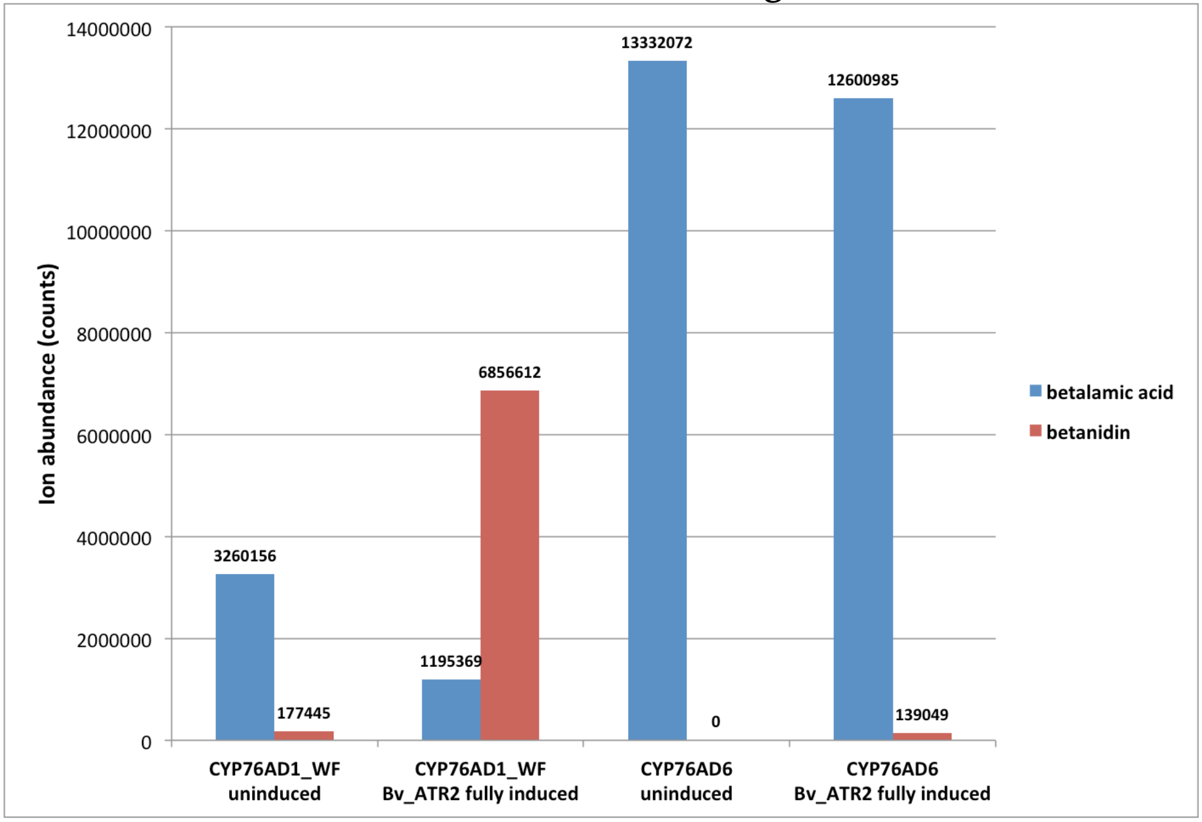


Figure 2-36. CPR activity can restore DOPA oxidase functionality in CYP76AD6. Supernatants from Figure 2-35 and Figure 2-36 were analyzed using LC/MS. With full induction of a CPR (ATR2), CYP76AD6 produces almost as much betanidin as CYP76AD6 W13L/F309L (WF) under normal conditions. No error bars are shown.

Though there is some precedent for electron chain partners influencing P450 activity and apparent substrate specificity[63,64], this is to my knowledge the first case of a P450 with multiple sequential activities on a substrate where both mutations and CPR expression can influence whether those activities happen and whether they go to completion. Although we do know that CYB5 has only an inhibitory role and CPR is likely necessary and sufficient for both tyrosine hydroxylase and DOPA oxidase activities, the exact mechanism of the F309L mutation remains ambiguous. If the homology model is correct, the F309L mutation lies on the I helix, sitting between the L and E helices and facing the D helix which is further away. Study of the bacterial P450 BM3 suggests that the I helix may serve as an important water channel or lever for a conformational shift on reductase binding[65], but other studies suggest that the C helix, which is very distal, is where the reductase actually binds[66]. What we know about CYP76AD1's catalytic cycle is summarized in **Figure 2-37**.

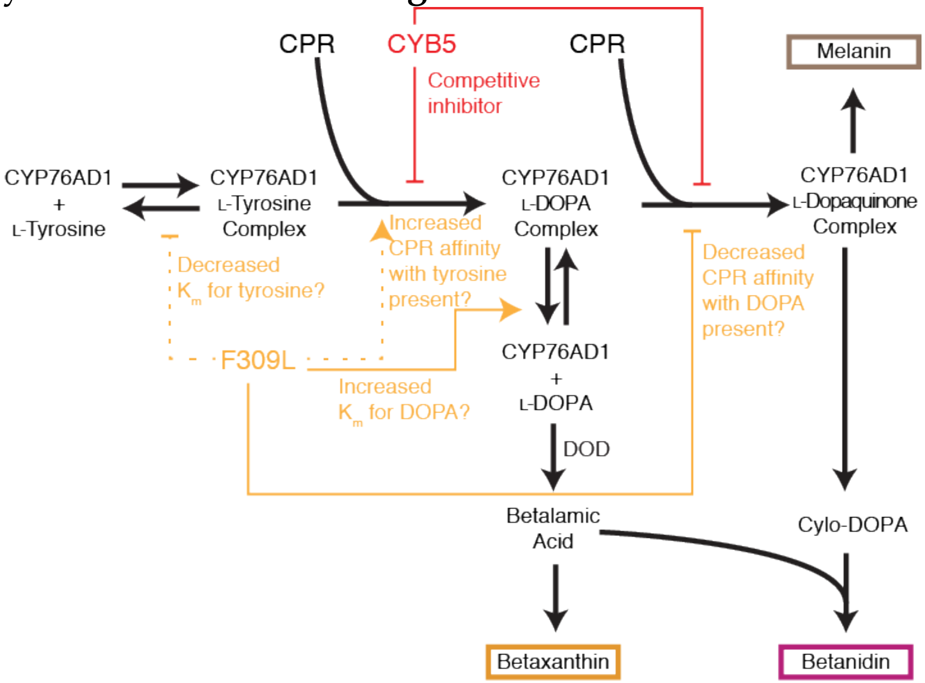


Figure 2-37. Possible mechanisms of F309L's effect on betaxanthin and betanidin production. Supernatants F309L's possible mechanisms are highlighted in gold, while CYB5's activity is highlighted in red. It seems less likely that improved tyrosine binding or CPR binding with tyrosine present are responsible for the increase in betaxanthin production with F309L. Additional CPR is able to drive F309L mutants to make more betanidin, but it remains less than their non-F309L peers, unlike the W13L mutants. More likely is that L-DOPA is being released before the DOPA oxidase activity can take place, either due to poor affinity for L-DOPA, or poor affinity for CPR with L-DOPA bound.

2.4.4. Solubilization of CYP76AD1 and reductases

Since the native yeast reductase NCP1 can strengthen undesirable activities in CYP76AD proteins but is itself critical for healthy yeast, we have a particular need for a CYP76AD1-CPR pair orthogonal from the rest of the host system. Typically this is achieved by directly fusing the enzymes[67,68], but a 1:1 expression ratio of CYP:CPR is not always desirable, especially in this case, where a 1:1 CYP:CPR ratio was approximated by the maximum induction level and showed significant product loss due to DOPA oxidase activity. Being a protein-protein interaction, solubilizing and sequestering the enzymes in a compartment in the cell is a viable alternative.

As further motivation, though microsomal preparations have worked for other cytochromes P450[69], we were unable to achieve an active microsomal preparation of CYP76AD1 as tested by NADPH consumption in the presence of tyrosine or L-DOPA. Active, purified protein is necessary for crystallography and kinetic assays, so efforts to solubilize CYP76AD1 for enhanced expression were expanded. In the literature, strategies to solubilize cytochromes P450 revolve around deleting the N-terminal transmembrane domain[70] or replacing it with a detergent peptide called PD1[71,72]. Also of interest was the solubilization of potential reductase partners NADPH-dependent P450 reductase (in yeast NCP1) and cytochrome *b5* (CYB5). Solubilizing NCP1 by deleting its sole transmembrane domain has already been demonstrated in the literature[73,74] and the same strategy could likely be applied to *B. vulgaris* ATR1 and ATR2. As for CYB5, in humans, a soluble alternate isoform lacking the C-terminal transmembrane domain exists as erythrocyte cytochrome *b5* (CYB5A)[75], and using this protein or copying a similar truncation into yeast could be successful. These strategies are summarized in **Figure 2-38**.

```
MAKKT strategy (Wester, 2004)
CYP2C9          ---------MDSLVVLVLCLSCLLLLSLWRQSSGRGKLPPGPTPLPVIGNILQIGIKDISKSLTNLS 58
MAKKT-CYP2C9    ---------------------------MAKKTSGRGKLPPGPTPLPVIGNILQIGIKDISKSLTNLS 40
CYP76AD1        ------MDHATLAMILAIWFISFHFIKLLFSQQTTKLLPPGPKPLPIIGNILEVGKKP-HRSFANLA 60
MAKKT-CYP76AD1  -------------------------MAKKTSQQTTKLLPPGPKPLPIIGNILEVGKKP-HRSFANLA 41

PD1 strategy (Schoch, 2003)
CYP73A1         -----MDLLLIEKTLVALFAAIIGAILISKLRGKKFKLPPGPIPVPIFGNWLQVGDDLNHRNLTDLA 62
PD1-CYP73A1     MEELLKQALQQAQQLLQQAQELAKKILISKLRGKKFKLPPGPIPVPIFGNWLQVGDDLNHRNLTDLA 67
CYP76AD1        ------MDHATLAMILAIWFISFHFIKLLFSQQTTKLLPPGPKPLPIIGNILEVGKKP-HRSFANLA 60
PD1-CYP76AD1    MEELLKQALQQAQQLLQQAQELAKKIKLLFSQQTTKLLPPGPKPLPIIGNILEVGKKP-HRSFANLA 66
CYP76AD6        ------MDNATLAVILSILFVFYHIFKSFFTNSSSRRLPPGPKPVPIFGNIFDLGEKP-HRSFANLS 60
PD1-CYP76AD6    MEELLKQALQQAQQLLQQAQELAKKFKSFFTNSSSRRLPPGPKPVPIFGNIFDLGEKP-HRSFANLS 66

NCP1 strategy (Lamb, 2001)
NCP1            MPFGIDNTDFTVLAGLVLAVLLYVKRNSIKELLMSDDGDITAVSSGNRDIAQVVTENNKNYLVLYAS 67
NCP1Δ33         ---------------------------------MSDDGDITAVSSGNRDIAQVVTENNKNYLVLYAS 34

Hs_CYB5      MAEQSDEAVKYYTLEEIQKHNHSKSTWLILHHKVYDLTKFLEEHPGGEEVLREQAGGDATENFEDVGHST 70
Hs_CYB5A     MAEQSDEAVKYYTLEEIQKHNHSKSTWLILHHKVYDLTKFLEEHPGGEEVLREQAGGDATENFEDVGHST 70
Sc_CYB5      -------MPKVYSYQEVAEHNGPENFWIIIDDKVYDVSQFKDEHPGGDEIIMDLGGQDATESFVDIGHSD 63
Sc_CYB5Δ23   -------MPKVYSYQEVAEHNGPENFWIIIDDKVYDVSQFKDEHPGGDEIIMDLGGQDATESFVDIGHSD 63
Sc_CYB5Δ42   -------MPKVYSYQEVAEHNGPENFWIIIDDKVYDVSQFKDEHPGGDEIIMDLGGQDATESFVDIGHSD 63

Hs_CYB5      DAREMSKTFIIGELHPDDRPKLNKPPETLITTIDSSSSWWTNWVIPAISAVAVALMYRLYMAED       134
Hs_CYB5A     DAREMSKTFIIGELHPDDRPKLNKPPEP------------------------------------       98
Sc_CYB5      EALRLLKGLYIGDVDKTSERVSVEKVSTSENQSKGSGTLVVILAILMLGVAYYLLNE-------       120
Sc_CYB5Δ23   EALRLLKGLYIGDVDKTSERVSVEKVSTSENQSK------------------------------       97
Sc_CYB5Δ42   EALRLLKGLYIGDVD-------------------------------------------------       78
```

Figure 2-38. Multiple sequence alignment of different solubilization strategies. Identical residues are highlighted in black, similar residues are highlighted in grey, and added solubility peptides are highlighted in cyan. “Hs” indicates human, “Sc” is yeast.

The "solubilized" versions of CYP76AD1 provided enhanced activity on a level comparable to the W13L/F309L double mutant (MAKKT data not shown), and the PD1-variant showed roughly 5-fold greater expression by microscopy (**Figure 2-39**) although PD1-CYP76AD1 still showed association with membranes. We suspect this is due to interaction with its reductase partner NCP1, or possibly nonspecific interactions due to the hydrophobic F-G patch present on many P450s[76,77].

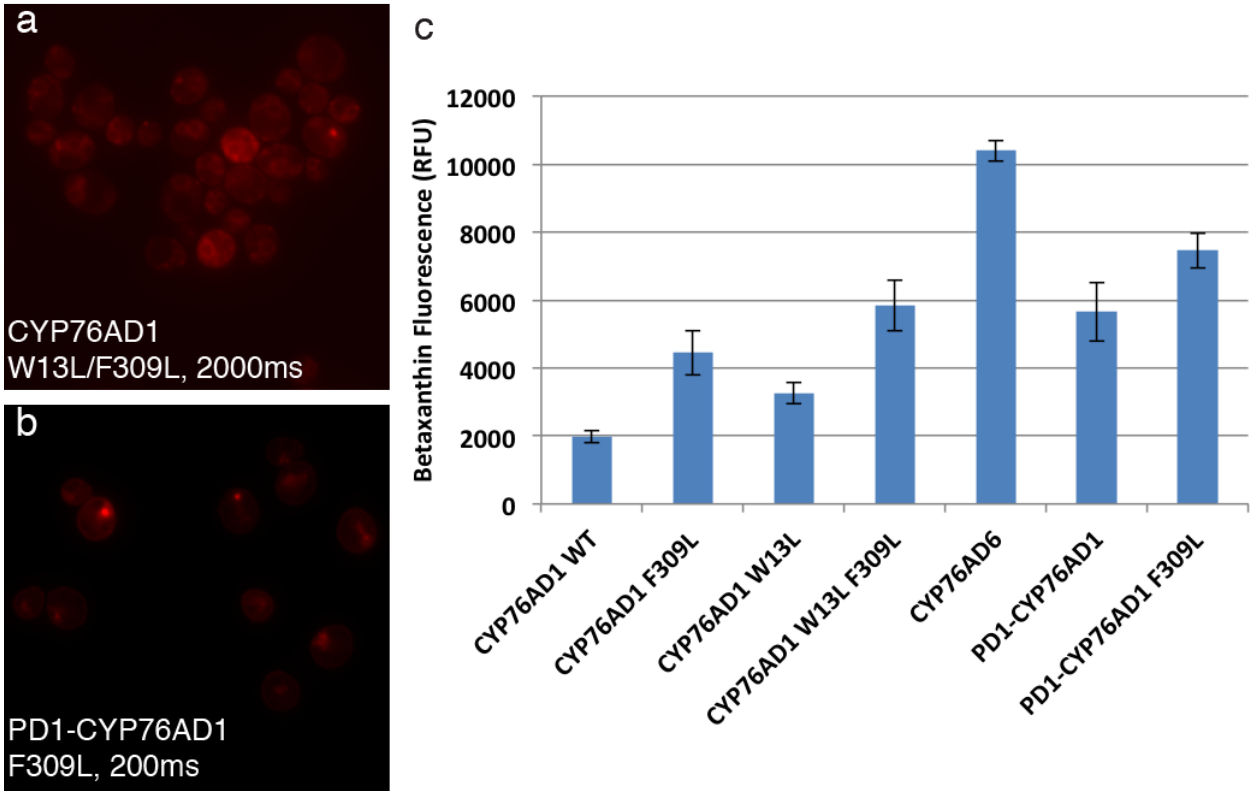


Figure 2-39. The PD1 solubilization peptide enhances CYP76AD1 expression and performance. (**a,b**) Microscopy of CYP76AD1 F309L – mRuby2 with either the W13L mutation or the PD1 solubility modification, expressed using pTDH3. (**a**) The ratio of cellular fluorescence to background fluorescence with W13L is roughly 2.2:1. (**b**) The ratio of cellular to background fluorescence with PD1 is 10.7:1, giving an approximate 5-fold increase in expression. (**c**) Measurements of betaxanthin fluorescence in the supernatants of cells grown on 2% raffinose, 2 mM tyrosine, 10 mM ascorbate in minimal media. Error bars indicate mean ± s.d. of six biological replicates.

Although there were sizeable gains in productivity, a more interesting trend was seen when PD1-CYP76AD1 ± F309L was combined with NCP1Δ33, the N-terminally truncated solubilized yeast CPR (**Figure 2-40** and **Figure 2-41**). NCP1Δ33 showed no effect on CYP76AD1 or CYP76AD6, but when combined with PD1-CYP76AD1 ± F309L, the reductase substantially increased betaxanthin production / tyrosine hydroxylase and took much more expression to reach the same increase in DOPA oxidase activity that comes with overexpression of non-solubilized reductases like NCP1, ATR1, or ATR2.

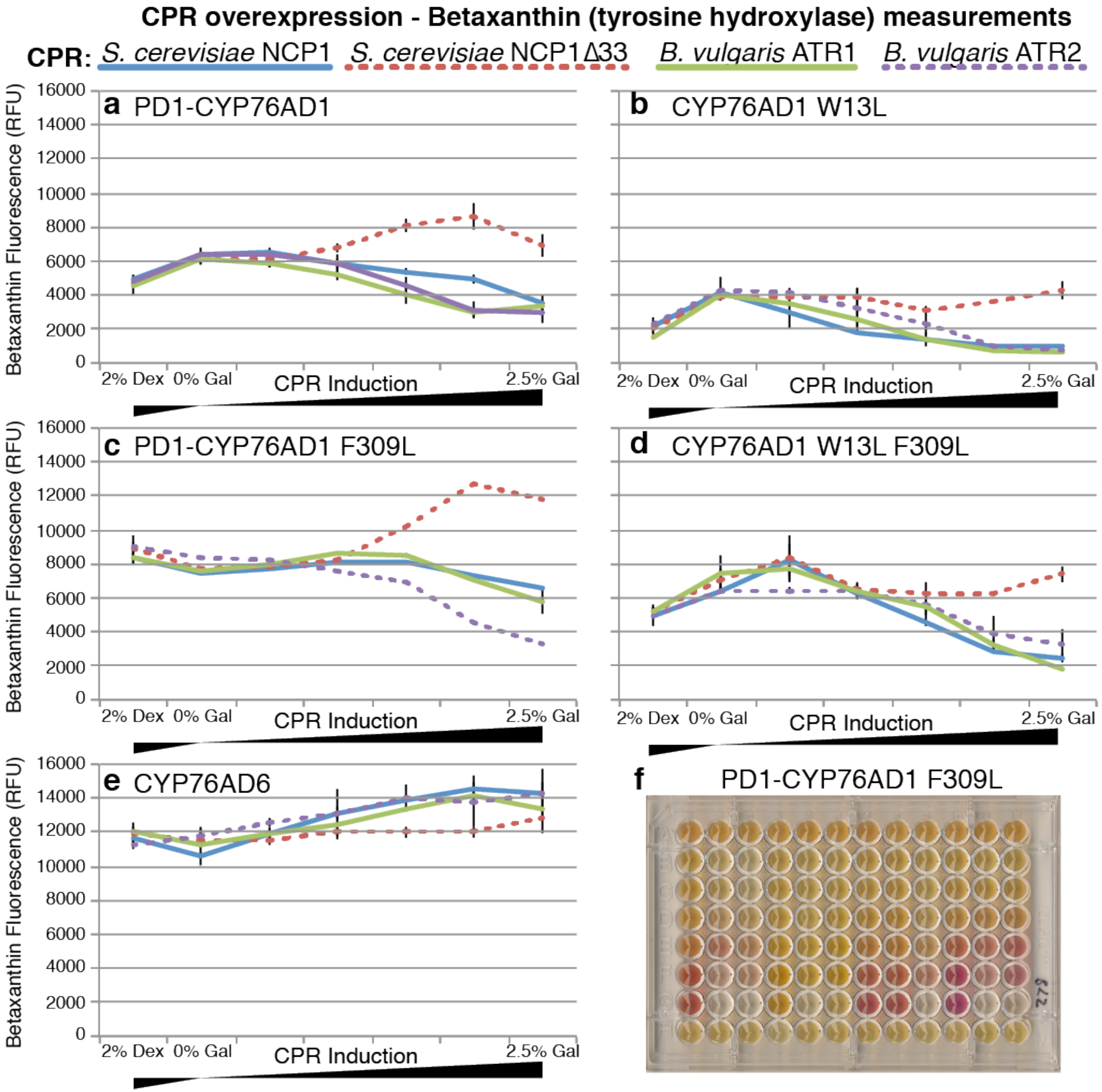


Figure 2-40. Only PD1-CYP76AD1 mutants respond to NCP1Δ33. Yeast constitutively expressing the indicated CYP were transferred to 2% raffinose minimal media, 2 mM tyrosine, 10 mM ascorbate, with 0-2.5% galactose in fourfold increments to induce production of a CPR. These cells were grown for 66 hours to account for slower growth on raffinose. Betaxanthin fluorescence increases as a product of tyrosine hydroxylase activity and is decreased by DOPA oxidase activity. (**a-e**) Induction curves of different CYP and CPR combinations. Error bars indicate mean ± s.d. of three biological replicates, except for (**c**), where error bars are missing from three of the CPRs because some cells failed to grow in the time allotted. Note that PD1-CYP76AD1 F309L approaches CYP76AD6 levels of betaxanthin production. (**f**) Supernatants transferred to a screening plate for visualization. Columns are organized in triplets: from left, NCP1, NCP1Δ33, ATR1, ATR2. Galactose concentration increases top to bottom.

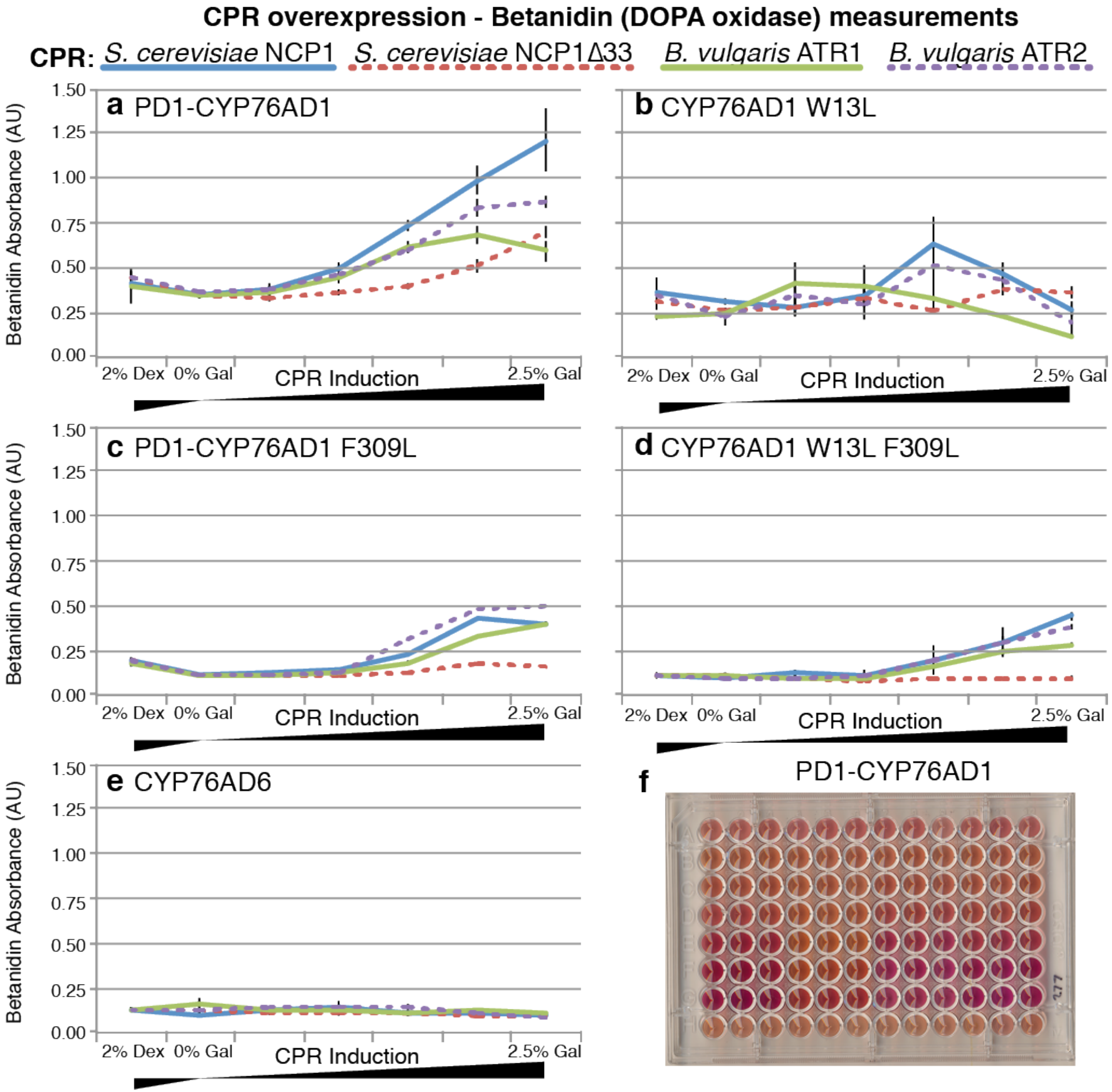


Figure 2-41. Only PD1-CYP76AD1 truncation mutants respond to NCP1Δ33. Yeast constitutively expressing the indicated CYP were transferred to 2% raffinose minimal media, 2 mM tyrosine, 10 mM ascorbate, with 0-2.5% galactose in fourfold increments to induce production of a CPR. These cells were grown for 66 hours to account for slower growth on raffinose. Betanidin absorbance increases as a product of DOPA oxidase activity, but also requires tyrosine hydroxylase product (L-DOPA) to form. Too much DOPA oxidase activity can deplete L-DOPA, resulting in a loss of color. (**a-e**) Induction curves of different CYP and CPR combinations. Error bars indicate mean ± s.d. of three biological replicates, except for (**c**), where error bars are missing from three of the CPRs because some cells failed to grow in the time allotted. Note that PD1-CYP76AD1 F309L approaches CYP76AD6 levels of betaxanthin production. (**f**) Supernatants transferred to a screening plate for visualization. Columns are organized in triplets: from left, NCP1, NCP1Δ33, ATR1, ATR2. Galactose concentration increases top to bottom.

This exciting result suggests we are well on our way to making an orthogonal P450-CPR pair and reaching new heights of L-DOPA / betaxanthin productivity on the way. **Figure 2-42** highlights how the different P450-CPR pairs appear to work together, but there are still many unanswered questions. For example, where are PD1-CYP76AD1 and NCP1Δ33 encountering each other? While it is possible that they are purely cytosolic, it may be equally likely that both proteins ended up on the surface of a different membrane from the ones that CYP76AD1 normally occupies. Also, is the smaller increase in DOPA oxidase activity per unit of induction with NCP1Δ33 due to the protein being spread over a greater volume, as opposed to being tethered to a membrane with two degrees of freedom? What is the ratio of P450:CPR in the membrane-tethered system and the PD1-solubilized system? Even with all these questions, the solubilized P450-CPR strategy is extremely promising.

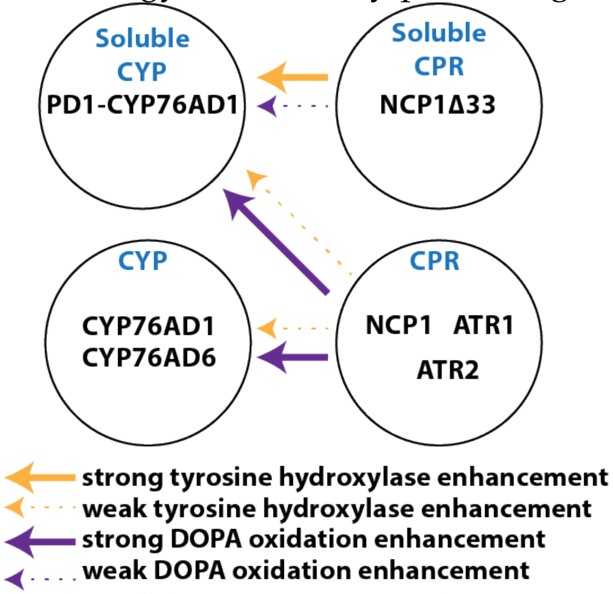


Figure 2-42. Interaction network between CYPs and CPRs. How can we suppress the CPR → soluble CYP interaction?

2.4.5. Synergy with peroxisomal compartmentalization

As mentioned above, a minimal level of native yeast CPR activity is necessary for maintaining healthy cells with good transformability. This presents a challenge when that same CPR can interact with a heterologous cytochrome P450 in a pathway of interest, such as CYP76AD1. The ability to isolate P450 – CPR pairs from the rest of the cell would help, and control over protein-protein interactions via compartmentalization is one thing the peroxisome excels at. An initial investigation suggests that peroxisomally-tagged CYPs, whether solubilized or not, go to the peroxisome and possibly the vacuole (**Figure 2-43d-f**). However, this investigation used

Venus-tagged constructs, and, for unknown reasons, soluble (MAKKT- or PD1-) constructs are highly toxic regardless of peroxisomal import, while mRuby-tagged varieties are nontoxic. Thus, the data must be interpreted skeptically. The toxicity might be caused by misfolding and degradation (with possible loss of essential NCP1), which, considering the vacuolar appearance and loss of fluorescence of solubilized fusion proteins in *pex5Δ* backgrounds (**Figure 2-43g-i**), seems likely. Toxicity also prevented testing of whether CYP-ePTS1 constructs produce L-DOPA. Future formulations of this experiment with mRuby or mKate tags should be more informative.

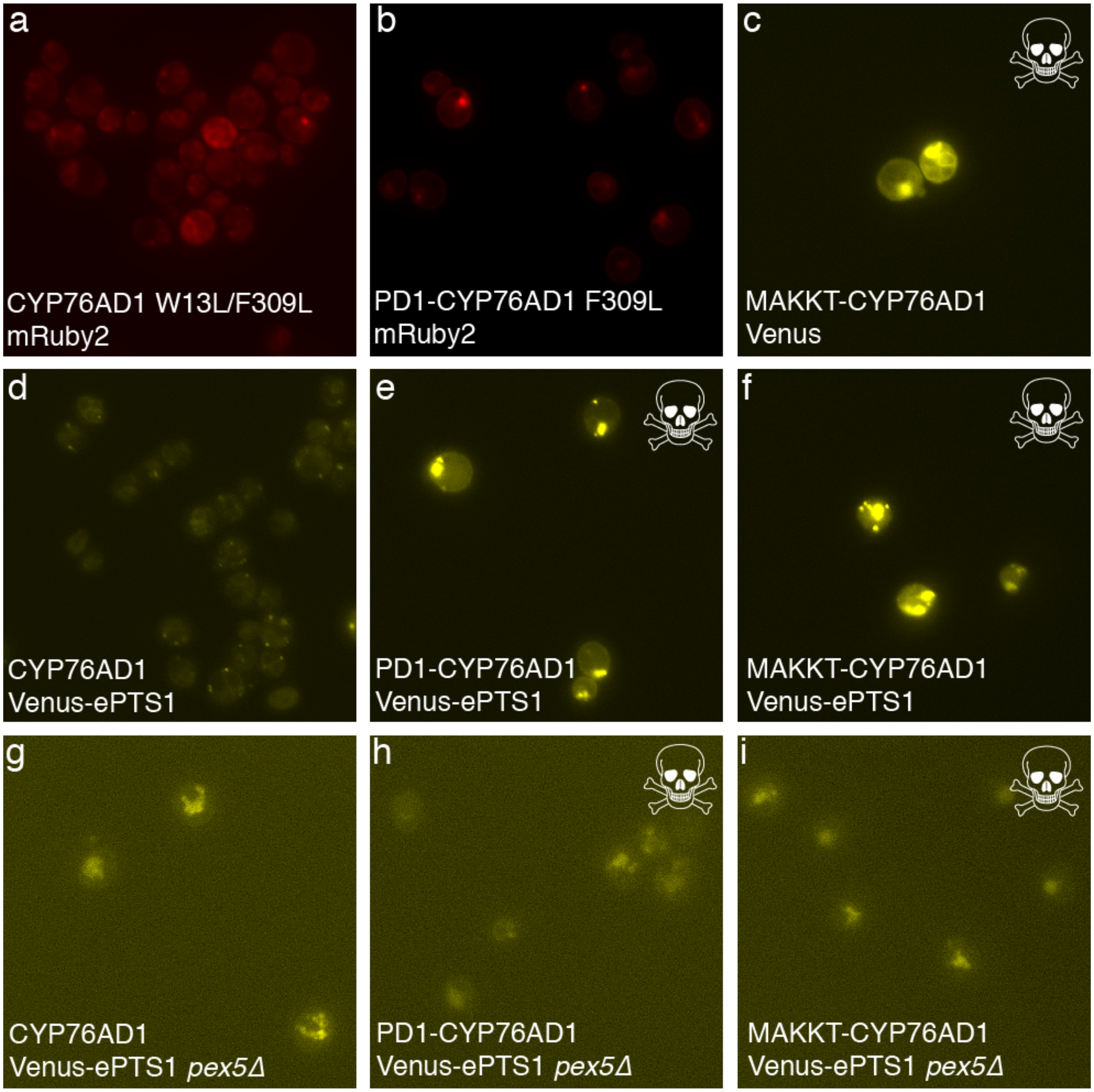


Figure 2-43. CYP76AD1 can be placed in the peroxisome. Fluorescence microscopy of cells expressing CYPs with (**d-i**) or without (**a-c**) a peroxisomal targeting sequence in a wildtype (**a-f**) or pex5Δ peroxisomal import deficient (**g-i**) background. Skull = toxic.

Though these results are encouraging, it remains to be tested whether the powerful combination of PD1-CYP76AD1 and NCP1Δ33 can be sequestered together and remain productive. Although NCP1Δ33 didn't significantly affect CYP76AD1 (non-solubilized), it could interfere with other CYPs and has been known to do so in the literature[74]. It may also be possible to draw on separate reducing equivalent (NADPH) pools by moving the P450-CPR pair into the peroxisome. See **Chapter 3** for more details on engineering peroxisomes to sequester protein cargo of interest.

2.4.6. Synergy with biological protecting groups

One of the major features of the BIA pathway and DOPA-derived molecules in general is the presence of reactive catechols. This functional group is unstable under mild oxidizing conditions, even air, especially above pH 8-9 when the catechol becomes deprotonated[78]. However, adding a chemical group to these unstable centers will block most degradation reactions. Enzymes that add biological protecting groups, such as glucose or sulfate, to DOPA and DOPA-derived compounds have already been identified and characterized[79–82]. In fact, betanidin is one such compound, where glucose is added to preserve the molecule in the form of betanin[83,84]! Furthermore, these biological protecting groups can be added in a stereoselective manner[81], which can aid chemists in their work with medicinal BIAs. Finally, not only are BIAs a rich medicinal field, but with biological protecting groups added, these molecules, such as morphine-6-glucuronide[85], can gain altered or increased bioactivity. As an alternative, morphine-3-glucuronide remains inactive[85] unless deprotected by gut bacteria, limiting the route of administration of a common opiate to avoid the negative outcomes associated with injection misuse[86]. See **Chapter 4** for more details on using biological protecting groups to stabilize reactive molecules, aid in chemical syntheses, and make novel products.

2.5. Materials and Methods

2.5.1. Strains and growth media

The base *S. cerevisiae* strain for all experiments in this article was BY4741 (*MAT*a *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*). Wildtype yeast cultures were grown in YPD (10 g/l Bacto Yeast Extract; 20 g/l Bacto Peptone; 20 g/l Dextrose). Selection of auxotrophic markers (URA3, LEU2, and/or HIS3) was performed in synthetic complete media (6.7 g/l Difco Yeast Nitrogen Base without amino acids; 2 g/l Drop-out Mix Synthetic Minus appropriate amino acids, without Yeast Nitrogen Base (US Biological); 20 g/l Dextrose). All strains constructed in this work are listed in **Table 2-1**.

Golden gate assembly reactions were transformed in TG1 chemically competent *E. coli*. Error-prone PCR libraries were transformed in TransforMax EPI300 (Epicentre) electrocompetent *E. coli*. Transformed cells were selected on Lysogeny Broth (LB) containing the antibiotics ampicillin or kanamycin.

2.5.2. Yeast expression vectors

Yeast expression vectors were built using Golden Gate Assembly[87]. Vector sequences were derived from the pRS series of plasmids. Promoter and terminator sequences for heterologous enzyme expression were derived from the yeast genome. Unique restriction sites (BsaI and BsmBI) were removed to facilitate plasmid construction.

Error-prone PCR library plasmids included a CEN6/ARS4 low copy number origin of replication to enable high-efficiency library transformations as did pWCD2353 for the expression of reticuline pathway enzymes (module C). All other plasmids contained no yeast origin of replication and were designed for direct integration into the yeast genome via homologous recombination at the URA3 or LEU2 locus. All strains and plasmids used in this work are listed in **Table 2-1** and **Table 2-2**. All sequences are available on Genbank at the accession numbers listed, and the accession number of the *Papaver somniferum* NCS used is KP262411.

2.5.3. Yeast strain construction

Aside from mutant libraries and the reticuline production strain ABC (Strain 33, **Table 1-1**), all genetic modifications to yeast were made via genomic integration. Yeast integration plasmids (YIPs) were constructed that lacked a yeast origin of replication but included regions of homology to either the URA3 or LEU2 locus. Five hundred μg of plasmid was linearized by digestion with NotI and transformed directly into yeast using a standard LiOAc transformation. Cells were plated onto dropout plates corresponding to the YIP's auxotrophic marker. Replicate colonies were picked directly from this transformation plate for further analysis.

Table 2-1. List of yeast strains used in this work. Strains were made by chromosomal integration of plasmids directly into the parent strain as indicated (except for Strain 33 which contains a CEN6/ARS4 plasmid). Strains 31, 32, and 33 are referred to as strains A, AB, and ABC in the text respectively. Annotated plasmid sequences can be found on GenBank with the accession numbers provided.

#	Strain Name	Strain Parent	Accession Number	Description	Yeast Marker
1	yWCD098	BY4741	KR232335	Empty Vector	URA3
2	yWCD303	BY4741	KR232334	pRPL18B-mKate2-DOD-tADH1	URA3
3	yWCD538	BY4741	KR232333	pTDH3-DOD-tADH1	URA3
4	yWCD683	BY4741	KR232325	pCCW12-DOD-tADH1	URA3
5	yWCD684	BY4741	KR232324	pTDH3-CYP76AD1-tTDH1-pCCW12-DOD-tADH1	URA3
6	yWCD732	BY4741	KR232311	pTDH3-AbPPO2-tTDH1-pCCW12-DOD-tADH1	URA3
7	yWCD688	BY4741	KR232320	pTDH3-CYP76AD1_F309L-tTDH1-pCCW12-DOD-tADH1	URA3
8	yWCD735	BY4741	KR232310	pTDH3-CYP76AD1_W13L-tTDH1-pCCW12-DOD-tADH1	URA3
9	yWCD736	BY4741	KR232309	pTDH3-CYP76AD1_W13L_F309L-tTDH1-pCCW12-DOD-tADH1	URA3
10	yWCD852	BY4741	KR232299	pTDH3-CYP76AD1-tTDH1	URA3
11	yWCD853	BY4741	KR232298	pTDH3-CYP76AD1_W13L-tTDH1	URA3
12	yWCD854	BY4741	KR232297	pTDH3-CYP76AD1_F309L-tTDH1	URA3
13	yWCD855	BY4741	KR232296	pTDH3-CYP76AD1_W13L_F309L-tTDH1	URA3
14	yWCD856	BY4741	KR232295	pTDH3-AbPPO2-tTDH1	URA3
15	yWCD694	BY4741	KR232316	pCCW12-DODC-tADH1	URA3
16	yWCD695	BY4741	KR232315	pTDH3-CYP76AD1-tTDH1-pCCW12-DODC-tADH1	URA3
17	yWCD699	BY4741	KR232314	pTDH3-CYP76AD1_F309L-tTDH1-pCCW12-DODC-tADH1	URA3
18	yWCD742	BY4741	KR232308	pTDH3-CYP76AD1_W13L-tTDH1-pCCW12-DODC-tADH1	URA3
19	yWCD743	BY4741	KR232307	pTDH3-CYP76AD1_W13L_F309L-tTDH1-pCCW12-DODC-tADH1	URA3
20	yWCD748	BY4741	KR232305	pTDH3-CYP76AD1-tTDH1-pCCW12-DODC-tADH1-pPGK1-ARO4_FBR-tPGK1	URA3
21	yWCD745	BY4741	KR232306	pTDH3-CYP76AD1_W13L_F309L-tTDH1-pCCW12-DODC-tADH1-pPGK1-ARO4_FBR-tPGK1	URA3
22	yWCD872	BY4741	KR232286	pTDH3-CYP76AD1-3XFlag_6XHis-tTDH1	URA3
23	yWCD873	BY4741	KR232285	pTDH3-CYP76AD1_W13L-3XFlag_6XHis-tTDH1	URA3
24	yWCD874	BY4741	KR232284	pTDH3-CYP76AD1_F309L-3XFlag_6XHis-tTDH1	URA3
25	yWCD875	BY4741	KR232283	pTDH3-CYP76AD1_W13L_F309L-3XFlag_6XHis-tTDH1	URA3
26	yWCD783	BY4741	KR232313	pPGK1-ARO4_FBR-tPGK1	LEU2
27	yWCD857	BY4741	KR232294	pTDH3-CYP76AD1-tTDH1-pPGK1-ARO4_FBR-tPGK1	URA3

28	yWCD858	BY4741	KR232293	pTDH3-CYP76AD1_W13L-tTDH1-pPGK1-ARO4_FBR-tPGK1	URA3
29	yWCD859	BY4741	KR232292	pTDH3-CYP76AD1_F309L-tTDH1-pPGK1-ARO4_FBR-tPGK1	URA3
30	yWCD860	BY4741	KR232291	pTDH3-CYP76AD1_W13L_F309L-tTDH1-pPGK1-ARO4_FBR-tPGK1	URA3
31 (A)	yWCD714	BY4741	KR232312	pTDH3-CYP76AD1_W13L_F309L-tTDH1-pCCW12-DODC-tADH1-pPGK1-ARO4_FBR-tPGK1	URA3
32 (AB)	yWCD758	31	KR232302	pTDH3-Ps_NCS-3-tENO2	URA3, LEU2
33 (ABC)	yWCD782	32	KR232300	pTDH3-Ps_6OMT-tADH1-pCCW12-Ps_CNMT-tPGK1-pPGK1-Ps_4'OMT2-tENO2-pTEF1-Ec_CYP80B1-tTDH1	URA3, LEU2, HIS3
34	yWCD619	3	KR232326	pTDH3-CYP76AD1-mKate2-tADH1	URA3, LEU2
35	yWCD620	3	KR232332	pTDH3-CYP76AD1_mut1-mKate2-tADH1	URA3, LEU2
36	yWCD621	3	KR232331	pTDH3-CYP76AD1_mut2-mKate2-tADH1	URA3, LEU2
37	yWCD622	3	KR232330	pTDH3-CYP76AD1_mut3-mKate2-tADH1	URA3, LEU2
38	yWCD623	3	KR232329	pTDH3-CYP76AD1_mut4-mKate2-tADH1	URA3, LEU2
39	yWCD624	3	KR232328	pTDH3-CYP76AD1_mut5-mKate2-tADH1	URA3, LEU2
40	yWCD625	3	KR232327	pTDH3-CYP76AD1_mut6-mKate2-tADH1	URA3, LEU2
41	yWCD868	BY4741	KR232290	pTDH3-CYP76AD1-Venus-tTDH1	URA3
42	yWCD869	BY4741	KR232289	pTDH3-CYP76AD1_W13L-Venus-tTDH1	URA3
43	yWCD870	BY4741	KR232288	pTDH3-CYP76AD1_F309L-Venus-tTDH1	URA3
44	yWCD871	BY4741	KR232287	pTDH3-CYP76AD1_W13L_F309L-Venus-tTDH1	URA3
45	yWCD685	BY4741	KR232323	pTDH3-CYP76AD2-tTDH1-pCCW12-DOD-tADH1	URA3
46	yWCD686	BY4741	KR232322	pTDH3-CYP76AD3-tTDH1-pCCW12-DOD-tADH1	URA3
47	yWCD687	BY4741	KR232321	pTDH3-CYP76AD4-tTDH1-pCCW12-DOD-tADH1	URA3
48	yWCD689	BY4741	KR232319	pTDH3-CYP76AD2_F308L-tTDH1-pCCW12-DOD-tADH1	URA3
49	yWCD690	BY4741	KR232318	pTDH3-CYP76AD3_F307L-tTDH1-pCCW12-DOD-tADH1	URA3
50	yWCD691	BY4741	KR232317	pTDH3-CYP76AD4_F306L-tTDH1-pCCW12-DOD-tADH1	URA3
51	yWCD757	31	KR232336	Empty Vector	URA3, LEU2
52	yWCD759	31	KR232301	pTDH3-NCS-Tfl-tENO2	URA3, LEU2
53	yWCD760	31	KR232304	pTDH3-NCS-Tco-1-tENO2	URA3, LEU2
54	yWCD761	31	KR232303	pTDH3-NCS-Pso-1-tENO2	URA3, LEU2

Table 2-2. List of additional plasmids used in this work. Plasmids shown are those not included in **Table 2-1**.

#	Accession Number	Description	Yeast Marker	Yeast Origin	Purpose
1	KR232281	CYP76AD1	N/A	N/A	Error-prone PCR template
2	KR232280	pRPL18B-(dropout)-mKate2-tADH1	LEU2	CEN6/ARS4	CYP76AD1 error-prone PCR library vector
3	KR232279	pRPL18B-CYP76AD1-mKate2-tADH1	LEU2	CEN6/ARS4	CYP76AD1 library wildtype control
4	KR262888	pRPL18B-CYP76AD1_F309L-mKate2-tADH1	LEU2	CEN6/ARS4	CYP76AD1 library winner
5	KR232282	pRPL18B-CYP76AD1_W13L_F309L-mKate2-tADH1	LEU2	CEN6/ARS4	CYP76AD1 library winner

2.5.4. L-DOPA titrations for betaxanthin production

Strains 1 and 3 were grown overnight in synthetic complete media (minus uracil) with 2% glucose. Saturated cultures were back-diluted 100X into fresh media supplemented with L-DOPA (Sigma #D9628) and grown in 24-well or 96-well blocks with shaking at 30°C for 24 hours. L-DOPA concentrations ranged from 10,000 μM to 0.15 μM in two-fold increments. Cells were spun down, washed in PBS (pH=7.4), and betaxanthin was quantified on either a microplate fluorometer or a flow cytometer (see below).

2.5.5. Flow cytometry

A Guava easyCyte flow cytometer was used to quantify betaxanthin fluorescence at an excitation of 488 nm and emission of 510 nm with a gain setting of 20. Cells were gated for forward and side scatter as shown in **Figure 2-4**. Betaxanthin fluorescence was normalized to forward scatter to account for differences in cell size. Histograms were prepared and data analysis performed using the software package FlowJo (http://flowjo.com).

2.5.6. Fluorescence microscopy

For analysis of intracellular betaxanthin (**Figure 2-2b**), Strain 2 was grown with or without 1 mM L-DOPA in synthetic complete media (minus uracil) with 2% glucose overnight. Cultures were concentrated by centrifugation, washed in PBS (pH 7.4), and spotted onto plain glass slides to be examined with a Zeiss Observer D1 microscope using a 100x DIC objective. Images were captured using a Hamamatsu Orca-flash 4.0 (C11440) camera using auto-exposure. Fluorescence images were taken using an X-Cite Series 120 lamp, Zeiss filter set 45 (excitation at 560/40 nm and emission at 630/75 nm) for mKate2, and Zeiss filter set 46 (excitation at 500/20 nm and emission at 535/30 nm) for betaxanthin[34].

Intracellular CYP76AD1 localization was analyzed via C-terminal fusion of the yellow fluorescent protein Venus to each enzyme variant and visualization on a spinning disk confocal microscope. Strains 41-44 were grown to saturation in SD-uracil and back diluted 50X into fresh media. After 6 hours of growth at 30˚C, cells were concentrated by centrifugation, washed in PBS (pH 7.4), and spotted onto plain glass slides to be examined with a spinning disk confocal microscope (Yokogawa CSU-X1, Tokyo, Japan) using a 100x brightfield objective and a 488 nm excitation laser. Images were taken using a QuantEM 512SC EMCCD camera (Photometrics, Tucson, AZ) and analyzed using Fiji (http://fiji.sc).

2.5.7. Error-prone PCR library construction

Error-prone PCR was performed using the GeneMorph II Random Mutagenesis Kit (Agilent Technologies). Oligos 1 and 2 (**Table 2-3**) were used to amplify the CYP76AD1 coding sequence off of the template for PCR (Plasmid 1, **Table 2-2**). Plasmid template was added to the PCR reaction at a concentration of 40 ng/μl to achieve the desired error rate. The PCR reaction was run as suggested in the GeneMorph II user manual, using an annealing temperature of 45˚C.

The PCR was incubated with DpnI for 1 hour at 37˚C before being cleaned up using a Zymo DNA Clean and Concentrator kit. A BsaI Golden Gate Assembly reaction was run using 40 fmol of both the PCR product and Plasmid 2. This reaction was again cleaned up, eluted in 10 μl of water, and transformed in its entirety into TransforMax EPI300 (Epicentre) electrocompetent cells. After a 1-hour rescue, the cells were grown to saturation in 500 ml of LB+kanamycin overnight. Two ml of saturated culture was miniprepped and all resulting DNA was transformed into Strain 3 for screening using a standard LiOAc transformation.

Table 2-3. List of oligonucleotide primers used in this work. Oligos 13 and 14 were used in place of Oligo 5 for amplifying mutants 3 and 4 (**Figure 2-8**) respectively since mutations were present in the primer binding region of these mutants.

#	Oligo Name	Description	Sequence
1	CN40	Error-prone PCR – forward	TTGGTAGTCGGTCTCCTATG
2	DC48	Error-prone PCR - reverse	TTTTTATTGGTCTGGTCTCAGGAT
3	S16	Forward sequencing pZNR0943	CAAAACTACCTGTTTCACCAAAGG
4	AA05	Reverse sequencing pZNR0943	ACTTGAAGTGGTGGTTGTTC
5	CW77	CYP76AD1 shuffling 1F	GCATGGTCTCATATGGATCATGCAACATTAGC
6	CW78	CYP76AD1 shuffling 1R	ATGCGGTCTCAACAGCAGTTATCTTTCTGA
7	CW79	CYP76AD1 shuffling 2F	GCATGGTCTCACTGTTCATTTATTGTCTCCACA
8	CW80	CYP76AD1 shuffling 2R	ATGCGGTCTCAAGTTGTAGTAGTTGTAGTTGAT
9	CW81	CYP76AD1 shuffling 3F	GCATGGTCTCAAACTACAGATGATGTCTTGGA
10	CX01	CYP76AD1 shuffling 3R	ATGCGGTCTCAGGCAACAAGAAAACTGTAGG
11	CX02	CYP76AD1 shuffling 4F	GCATGGTCTCATGCCTAGAAAAGCCGAC
12	CX03	CYP76AD1 shuffling 4R	ATGCGGTCTCAGGATCCGTATCTTGGAATTG
13	CW76	CYP76AD1 shuffling 1F – alt1	GCATCGTCTCATCGGTCTCCTATGGAACATGCAACATTAGCTATGAT
14	CW61	CYP76AD1 shuffling 1F – alt2	GCATCGTCTCATCGGTCTCCTATGGATCACGCAACATTAGCTATG

2.5.8. DNA shuffling library construction

CYP76AD1 mutants were shuffled using Golden Gate Shuffling[88]. The coding sequence was divided into four regions that equally distributed the set of mutations being shuffled. Oligos 5-14 were used to PCR amplify fragments using the mutant plasmids as templates. These PCR's were mixed in a Golden Gate Assembly reaction and prepped for library screening as described in the above section on error-prone PCR library construction.

2.5.9. Library screening for improved betaxanthin production

For screening of tyrosine hydroxylase mutants, yeast transformations were plated on multiple 500 cm^2 agar plates (with synthetic dropout media) at a density of approximately 50,000 colonies per plate. After 72 hours of growth at 30°C, the most intensely yellow colonies were isolated for further analysis. To aid in the selection of high-producing colonies, 470 nm blue LED light *epi*-illumination was applied to the plates and viewed through an amber high-pass emission filter.

2.5.10. Betaxanthin fluorescence quantification

Colonies were picked into synthetic complete media with 2% glucose and grown overnight. Saturated cultures were back-diluted 50X into 500 μL of fresh media in 96-deep-well blocks. Cultures were grown for 48 hours in a Multitron ATR shaker at 30°C. Cells were pelleted, washed, and resuspended in PBS (pH=7.4). The cells were transferred to glass-bottomed microplates and measured for fluorescence in a TECAN Safire2 (excitation: 485/5 nm, emission: 505/5 nm, gain: 120)[34]. Fold over background fluorescence measurements were obtained by normalizing to the average fluorescence observed in control Strain 4.

2.5.11. Production assay for L-DOPA, dopamine, dopaquinone, and melanin

Colonies were picked into 2.5 ml of 2X synthetic complete media (minus uracil) with 4% glucose. After overnight growth, saturated cultures were back-diluted 50X into 24-deep-well blocks with fresh media. The cultures were grown in a Multitron ATR shaker for 48 hours at 30°C. Cultures were pelleted and culture supernatants were measured for titer via LC/MS in the case of L-DOPA, dopamine, and dopaquinone. For dopaquinone production, 10 mM ascorbic acid was added to the growth media to prevent oxidation of this unstable product. In the case of melanin production, pellets were photographed directly.

2.5.12. Western blot

Strains 1 and 22-25 were grown to saturation in synthetic complete media (minus uracil) with 2% glucose overnight and back diluted 50X into fresh media. After 6 hours of growth with shaking at 30°C, 2.5ODs of cells were pelleted, washed in water, resuspended in 200 mM NaOH and incubated at room temperature for 5 minutes. Cells were again pelleted, resuspended in 50 μl of PAGE sample buffer and boiled at 95°C for 5 minutes. Samples were diluted 10-fold in PAGE sample buffer and 8 μl was loaded onto a NuPAGE Novex 4-12% Bis-Tris gel (Life Technologies) and run for 2.5 hours at 80 volts. Proteins were transferred onto PVDF transfer membrane in NuPAGE transfer buffer (Life Technologies) and blocked overnight in TBST (0.05% Tween20) with 5% milk. The membrane was washed twice with TBST for 5 minutes and incubated for 1 hour with HRP conjugated anti-FLAG (Sigma #A8592) and anti-GAPDH (Fisher Scientific # MA515738HRP) monoclonal antibodies at dilutions of 1:5,000 and 1:10,000 respectively. After six 5-minute washes in TBST, the HRP antibodies were detected by chemiluminescence using a ChemiDoc XRS imager (Bio-Rad) and the blots was analyzed using Fiji (http://fiji.sc).

2.5.13. Expression level quantification via fluorescent protein fusion

Colonies of Strains 1 and 41-44 were grown in 96-deep-well blocks in synthetic complete media minus uracil to saturation and back diluted into fresh media. After 24 hours of growth with shaking at 30°C, 100 μl of each culture was transferred to glass-bottomed microplate and measured for fluorescence in a TECAN Safire2 (excitation 516/5nm, emission 530/5nm, gain 100). Fold over background measurements were obtained by normalizing all fluorescence values to the average fluorescence observed in control Strain 1.

2.5.14. Betalain production assay

Special media was required to observe the production of betanidin in culture supernatants. We discovered that the standard media component *para*-aminobenzoic acid (PABA) was capable of spontaneous condensation with betalamic acid to produce PABA-betaxanthin. PABA-betaxanthin had similar absorbance properties to betanidin, making it difficult to visualize changes in betanidin formation.

We therefore prepared minimal media lacking PABA (7.6 g/l Difco Yeast Nitrogen Base without amino acids or vitamins; 76 mg/l Histidine; 76 mg/l Methionine; 76 mg/l Tryptophan; 380 mg/l Leucine; 2 mg/l Biotin; 400 mg/l Calcium Pantothenate; 2 g/l Inositol; 400 mg/l Pyridoxin HCL; 400 mg/l Thiamine HCL; 20 g/l Dextrose).

Cells were first grown to saturation in synthetic complete media (minus uracil) with 2% glucose. They were then back-diluted 50X into our custom minimal media which was supplemented with 10 mM ascorbic acid to prevent betanidin oxidation and 1 mM tyrosine. Cultures were grown in 24-deep-well blocks at 30°C in a Multitron ATR shaker for 24 hours. Cells were pelleted and supernatants were analyzed. Absorbance spectra of culture supernatants were acquired using a TECAN Safire2 plate reader. LC/MS analysis was performed as described below.

2.5.15. Shake-flask fermentations for norcoclaurine and reticuline

Colonies were picked into 2.5 ml of selective synthetic media with 2% glucose. After overnight growth, saturated cultures were back-diluted 50X into 250 ml baffled shake flasks containing 50 ml of fresh 2X selective synthetic media with 4% glucose. The cultures were grown at 30˚C with shaking at 220 rpm for 96 hours in an Innova model 44 incubator. At designated time points, aliquots were taken from the cultures to measure OD600, and media was stored at -20˚C for later analysis by LC/MS.

2.5.16. LC/MS analysis - Pellet extraction using acetonitrile

One ml of each cell culture was centrifuged at 3000 RCF for five minutes to collect cells. The cells were then washed with 1 ml of PBS (pH 7.4), centrifuged again and resuspended in 500 μl of PBS containing 35 U of zymolyase (Zymo Research, Irvine, CA). This was then incubated at 37 °C for 1 hour to spheroplast the yeast. After incubation, 500 μl of acetonitrile was added to disrupt cells and precipitate protein. The mixture was vortexed for 1 minute and centrifuged at 14000 RCF to remove precipitate. Finally, the supernatant was concentrated under vacuum to one tenth its original volume.

2.5.17. LC/MS analysis - HPLC absorbance traces

Ten μl of each culture supernatant or extract was separated on an Agilent 1260 Infinity Quaternary LC System with Zorbax Eclipse Plus C18 4.6 x 100 mm - 3.5 μm reversed-phase column (Agilent) at ~20°C using a 0.5 ml / min flow rate. Samples were eluted with a linear gradient from 100% water / 0% acetonitrile + 0.1% formic acid to 65% water / 35% acetonitrile + 0.1% formic acid over the course of 15 minutes. Absorbance was measured using a diode array detector at 405 nm (betalamic acid), 480 nm (betaxanthin) and 536 nm (betanidin) with 4 nm bandwidth. Mass spectrometry was carried out using an attached Agilent 6120 Quadrupole LC/MS, in atmospheric pressure ionization - positive electrospray (API-ES positive) mode at 100V fragmentor voltage, scanning masses between 100 and 700 Da. Peak identification was additionally assisted by ion extraction of betalamic acid (m/z 212.055 and R_t 11.0 minutes), tyrosine-betaxanthin/portulacaxanthin II (m/z 375.119 and R_t 10.9 minutes), betanidin (m/z 389.098 and R_t 10.3 minutes), as well as proline-betaxanthin (m/z 309.110 and R_t 8.8 minutes), valine-betaxanthin (m/z 311.125 and R_t 11.3 minutes), (Iso)leucine-betaxanthin (m/z 325.139 and R_t 13.1 or 13.4 minutes), phenylalanine-betaxanthin (m/z 359.123 and R_t 13.7 minutes), and tryptophan-betaxanthin (m/z 398.1357 and R_t 14.5 minutes). Traces were collected and analyzed using ChemStation (Agilent Technologies, Santa Clara, CA).

2.5.18. LC/MS analysis - quantification of L-DOPA and derivatives

Quantitative and tandem mass spectrometry (MS/MS) data were collected with the same column, pump, and gradient as above, but using an Agilent 6520 Accurate-Mass Q-TOF LC/MS for fragmentation and mass detection. The system was run in positive electrospray (ESI+) mode, with a 100V fragmentor voltage and, for MS/MS, a 23V collision energy. For quantification of betalamic acid (m/z 212.055 and R_t 11.0 minutes), tyrosine-betaxanthin/portulacaxanthin II (m/z 375.119 and R_t 10.9 minutes), and betanidin (m/z 389.098 and R_t 10.3 minutes), extracted ion counts were integrated and normalized against 8-point relative standard curves generated by repeated twofold dilutions of threefold concentrated supernatant from Strain 8 (betanidin) or Strain 9 (tyrosine-betaxanthin). Dopaquinone was quantified by ion extraction at m/z 196.06 and R_t 2.8 minutes; peaks were integrated and reported. For quantification of dopamine, ion counts with m/z of 154.086 $[M+H]^+$ and a retention time of 3.0 minutes were extracted, integrated, and quantified against an 8-point dopamine calibration curve ranging from 0.004 mM to 1 mM in twofold steps. Other species targeted were norcoclaurine (m/z 272.121 $[M+H]^+$ and R_t 10.7 minutes) and reticuline (m/z 330.17 $[M+H]^+$ and R_t 13.3 minutes). For quantification purposes, these extracted ion counts were integrated and compared against six-point calibration curves covering 0.078 – 2.50 μM and 0.039 – 1.25 μM in two-fold steps for norcoclaurine and reticuline, respectively. These calibration curves were generated using (*S*)-norcoclaurine and (*R*)-reticuline authentic standards (TRC Inc., North York, Ontario, Canada). Traces were analyzed, extracted, and quantified using MassHunter (Agilent Technologies, Santa Clara, CA) and MzMine2 (http://mzmine.sourceforge.net).

2.5.19. LC/MS analysis - Norcoclaurine chiral analysis

Culture supernatant from Strain 32 was concentrated 10X by SpeedVac, resuspended 1:2 in MeOH and centrifuged at 21,000 g for 30 sec. Norcoclaurine enantiomers were analyzed by reverse-phase HPLC using a Perkin-Elmer Series 200 Micropump (Perkin Elmer, Norfolk, CT, USA) equipped with a Shodex ORpak CDBS-453 column (FUTECS Chromatography, Daejeon, South Korea). Ten μl of supernatant was loaded on the column and separated using the following gradient: 0-35 min 5% B, 35-38 min 95% B, 38-45 min 5% B at a flow rate of 250 μL/min, where Solvent A was 0.1% acetic acid and Solvent B was 80:20:1 acetonitrile:methanol:acetic acid. Following separation, eluent was injected into a 7T-LTQ FT ICR instrument (Thermo Scientific, Bermen, Germany).

Norcoclaurine was identified by exact mass (m/z 272.128 [M+H]$^+$). Spontaneously-condensed (*R,S*)-norcoclaurine, generated through resuspending an equal ratio of L-dopamine (Sigma) and 4-HPAA (gift from Peter Facchini) in H_2O, and (*S*)-norcoclaurine standard (TRC Inc, North York, Ontario, Canada) were used to determine retention time of each enantiomer. Curves were smoothed with a 7-point boxcar average.

2.5.20. LC/MS analysis - Reticuline chiral analysis

Determination of reticuline enantiomers in Strain 33 supernatant was performed by HPLC-FT-MS. Yeast supernatant was concentrated 200X using a SepPak Light C18 cartridge (Waters, Milford, Massachusetts, USA), then resuspended 1:7 in MeOH and centrifuged at 21,000 g for 30 sec. Reticuline enantiomers were analyzed by standard-phase HPLC using a Perkin-Elmer Series 200 Micropump (Perkin Elmer, Norfolk, CT, USA) equipped with a 4.6x250 mm CHIRALCEL OD-H column (Daicel Chemical Industries, Osaka, Japan) heated to 40°C. Five μl of supernatant was loaded on the column and separated using an isocratic mixture of 72:28:0.001 hexane:isopropanol:diethylamine at a flow rate of 0.55 ml / min[22]. Following separation, eluent was injected into a 7T-LTQ FT ICR instrument (Thermo Scientific, Bermen, Germany). Reticuline was identified by exact ion mass (m/z 330.17 [M+H]$^+$). Authentic (*R*)-reticuline (TRC Inc, North York, Ontario, Canada) and (*S*)-reticuline (gift from Peter Facchini) were used as standards to determine retention time of each enantiomer. Curves were smoothed with a 7-point boxcar average.

Chapter 3. Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways

3.1. Introduction

Metabolic engineering of microorganisms promises to enable the environmentally friendly production of fuels, bulk chemicals, and therapeutics[89]. A chief consideration for optimizing production in microbial hosts is limiting crosstalk between high-flux engineered metabolic pathways and the native cellular processes of the production host[90]. Eukaryotes address the problem of metabolic crosstalk by partitioning proteins and metabolites in membrane-bound organelles to sequester toxic compounds[91], direct the activity of enzymes towards specific substrates[92], and establish distinct chemical environments (e.g. altered pH or redox state[93,94]). Recently, metabolic engineers have begun to harness the subcellular organelles of *Saccharomyces cerevisiae* to take advantage of these properties: the mitochondrion was used to enhance the production of isobutanol[95], and the vacuole was used as a site for methyl halide synthesis[96]. These successes relied on substrates that naturally accumulate in the mitochondria and vacuole, limiting the applicability to new pathways. We sought to extend the benefits of compartmentalization to any pathway of interest by laying the foundation for a flexible synthetic organelle.

A generalizable organelle compartmentalization strategy would feature several important traits: 1) the organelle should be orthogonal to native cellular processes, 2) import of heterologous enzymes should be rapid, efficient, and modular, and 3) transport of metabolites across the organelle membrane should be characterized and ultimately controllable. With the first criterion as a prerequisite for further engineering, we identified the yeast peroxisome as a promising starting point for the construction of such a synthetic organelle. The peroxisome of *S. cerevisiae* is primarily involved in β-oxidation of long-chain fatty acids, and previous characterization has shown that, unlike other organelles, peroxisome biogenesis can be completely disrupted without adversely affecting cell growth in common glucose-fed fermentation conditions[97]. Thus, the peroxisome can be cleared of its endogenous matrix proteins to establish an orthogonal subcellular compartment. We therefore focused on exploring the design rules for both protein and metabolite transport across the peroxisomal membrane as a step towards repurposing this organelle for heterologous pathway compartmentalization.

The wide diversity in peroxisomal form and function observed in nature provides further evidence for peroxisomes' versatility and suitability as a synthetic organelle. Methylotrophic yeasts such as *Pichia pastoris* and *Ogataea polymorpha* possess peroxisomes that can expand up to 80% of the total volume of the cell[98], suggesting that enormous protein capacity is achievable. Furthermore, the peroxisome has found use in nature not only as a catabolic compartment, but also as a site for unusual biochemistry, such as penicillin biosynthesis in *Penicillium chrysogenum*[99] and light generation in the firefly *Photinus pyralis*[100]. Even diversity in peroxisomal structure has been observed, as import of HEX proteins transforms peroxisomes into Woronin bodies, which help plug leaks in damaged ascomycetes[101]. Peroxisomes have already attracted some interest from metabolic engineers, who have placed lycopene and polyhydroxyalkanoate biosynthetic pathways in the peroxisomal lumen[102,103]. However, these efforts left many open challenges, including validation and optimization of protein import, measurements of peroxisomal permeability, and verification that intermediates are trapped within the peroxisome. It is these open challenges that we hope to address in this work.

Though the optimal parameters for peroxisomal protein import remain an open question, the biological mechanism has been extensively studied[104–106]. Peroxisomal matrix proteins generally contain one of two peroxisomal targeting signals (PTS1 or PTS2) that are recognized in the cytosol by corresponding receptor proteins, Pex5p and the Pex7p/Pex18p/Pex21p complex, respectively[104]. Upon recognition, these proteins are recruited to the Pex13p/Pex14p/Pex17p import pore complex and translocated into the peroxisomal lumen while remaining in the folded state[104]. The majority of native cargo proteins enter via some variant of the PTS1 tag, which canonically consists of Ser-Lys-Leu (SKL) at the carboxy-terminus[107]. Numerous studies have demonstrated that fusion of this three-amino acid tag is sufficient to redirect cytosolic proteins into the peroxisome, however recent work suggests that additional upstream residues also contribute to recognition by Pex5p[107–109]. A systematic experimental analysis of their effect on import has yet to be performed.

While the participants involved in peroxisomal protein transport are well established, transport of metabolites into and out of the peroxisome remains a matter for debate. To date, few peroxisomal membrane transporters have been identified in yeast, namely Pxa1p/Pxa2p for fatty acids[110] and Ant1p for ATP/AMP[111]. These transporters cannot account for the many metabolites consumed and released by the peroxisome. Early *in vitro* work suggested broad peroxisomal permeability to metabolites[112], potentially accounting for the dearth of identified transporters. This view was challenged by *in vivo* studies that suggested fatty acyl-Coenzyme A molecules do not freely cross the peroxisomal membrane[113]. To reconcile these observations, a nonspecific, aqueous pore was proposed that allows for the free diffusion of molecules below a certain size[114]. While recent *in vitro* results support this conclusion[111], the picture is complicated by the possibility that the *in vitro* assay protocols themselves generate different pore sizes depending on peroxisome pretreatment[104,115,116]. *In vivo* validation is lacking, and the hypothesized protein responsible for this porosity remains unidentified in yeast.

Here, we extend the current understanding of peroxisome biology to uncover key design rules for compartmentalizing heterologous pathways in the yeast peroxisome. We develop a novel enzyme-based strategy for evaluating the efficiency of targeting non-native cargo to the peroxisome. Our assay, which we find to be more sensitive than fluorescence microscopy for assaying peroxisomal import, indicates that the canonical PTS1 tag is context-dependent and can result in slow import of cargo if the residues immediately upstream of the tag are not optimized. By screening a library of randomized PTS1 peptide linker sequences, we identify linker charge as an important factor and define a modular PTS1 tag with enhanced import performance. This enhanced tag allows efficient compartmentalization of several non-native enzymes and as well as the first *in vivo* demonstration that the peroxisomal membrane is indeed permeable to small metabolites while presenting a barrier to larger solutes. We take advantage of this size-dependent permeability to compartmentalize in the peroxisome a two-enzyme model pathway that consists of a small, permeant substrate and a larger, impermeant intermediate. Finally, we assess the effects of peroxisomal compartmentalization on pathway performance and identify future avenues for developing the peroxisome as a generalizable metabolic engineering tool.

3.2. Results

3.2.1. Development of a sensitive peroxisomal protein import assay

The first step in expressing a heterologous pathway in the peroxisome is ensuring that each enzyme is effectively compartmentalized. While protein import is commonly monitored using fluorescence microscopy, we sought a method that would directly test the primary metric of interest for compartmentalizing new enzymatic pathways—cytosolic activity of an enzyme prior to peroxisomal import. For this purpose, we turned to a three-enzyme pathway from the bacterium *Chromobacterium violaceum* that converts tryptophan to the green pigment prodeoxyviolacein (PDV)[117] (**Figure 3-1** and **Figure 3-2a**). We reasoned that by expressing the first two pathway enzymes (VioA and VioB) in the cytosol but targeting the final enzyme (VioE) to the peroxisome, we would isolate VioE from its substrate, indole-3-pyruvic acid (IPA) imine dimer, and measurably reduce pathway flux. A large difference in PDV production between strains with VioE in the cytosol versus localized to the peroxisome would be indicative of efficient protein import (**Figure 3-2b**).

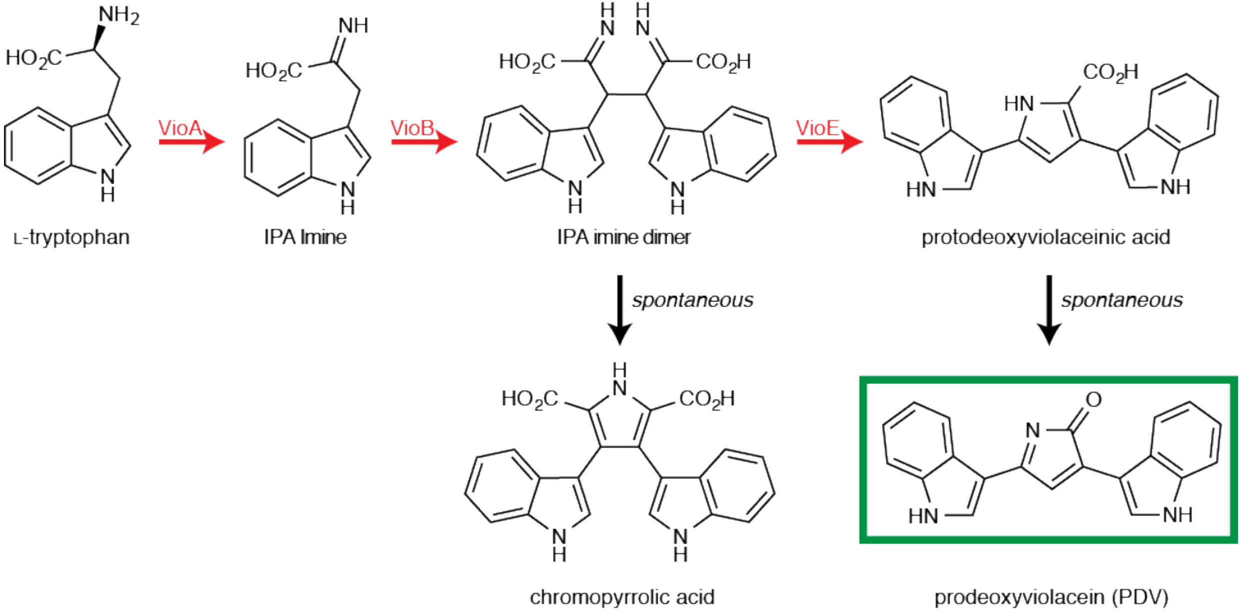


Figure 3-1. The prodeoxyviolacein (PDV) pathway and its byproducts. Three enzymes, VioA, VioB, and VioE, are responsible for transforming L-tryptophan into prodeoxyviolacein (PDV). PDV alone has a visible green color; the other intemediates are colorless. The product of VioB, IPA imine dimer, is unstable and will spontaneously convert into chromopyrrolic acid (CPA) if no VioE is encountered. Red arrows denote enzymatic reactions; black arrows denote spontaneous reactions.

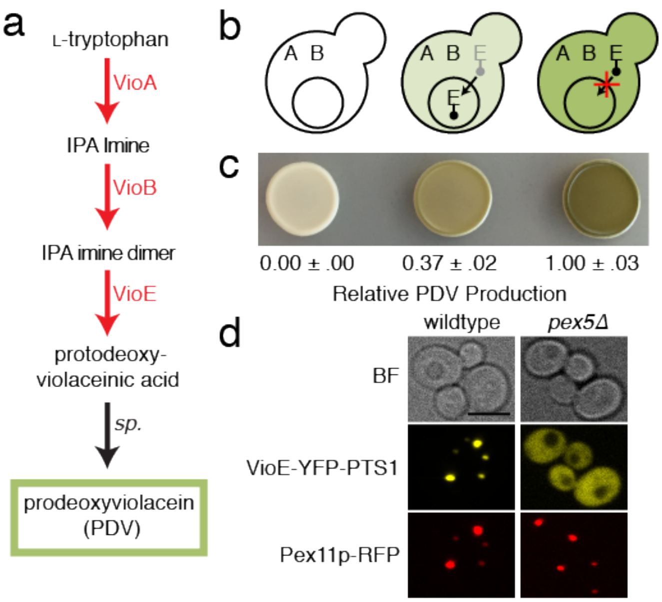


Figure 3-2. An enzyme-linked assay for peroxisomal import efficiency reveals cytosolic activity of PTS1-targeted cargo. (**a**) The green pigment PDV is produced from L-tryptophan as the result of three enzymatic steps (red arrows) and one spontaneous reaction (black arrow). (**b**) The precursors of PDV are colorless, so VioA and VioB can be freely expressed in the cytosol of yeast without any color change (left). Efficient PTS1 targeting of VioE will result in enzyme sequestration in the peroxisome and considerably reduced PDV production in a wildtype strain (middle) compared with a *pex5Δ* strain that is deficient for PTS1 import (right). (**c**) Agar plate spots corresponding to the strains depicted above after 36 hours of growth. VioE included a C-terminal YFP fusion and the canonical PTS1 tag (SKL). HPLC quantification of relative PDV production is included below each spot. (**d**) Fluorescence microscopy of yeast cells coexpressing the peroxisomal marker Pex11p-RFP and the VioE-YFP-PTS1 fusion protein in either a wildtype or *pex5Δ* strain background. Scale bar, 10 μm.

To achieve peroxisomal targeting, we fused the canonical PTS1 tag (Ser-Lys-Leu) to the C-terminus of VioE, preceded by a yellow fluorescent protein (YFP) so that proper localization could be confirmed by fluorescence microscopy. We expressed all three enzymes in either a wildtype strain or a *pex5Δ* strain that is deficient for PTS1 protein import (**Figure 3-2c**). Proper targeting of VioE to the peroxisome was confirmed by colocalization of VioE-YFP-PTS1 with a red fluorescent protein (RFP) peroxisomal marker, Pex11p-RFP (**Figure 3-2d**). While there was no cytosolic fluorescence observed via microscopy when VioE-YFP-PTS1 was expressed in wildtype cells, we measured only a 63% reduction in PDV production relative to the *pex5Δ* cytosolic control strain. This result suggested that, despite observing clearance of protein fluorescence from the cytosol, VioE remained in the cytosol long enough to confer significant activity prior to peroxisomal import. Reducing the occurrence of preimport cytosolic activity is critical for ensuring that pathway activity is completely confined to the peroxisome.

3.2.2. Positively charged residues enhance PTS1 import

Based on previous work showing that the residues adjacent to the PTS1 tag can affect protein import[108], we attempted to improve compartmentalization of VioE by modifying a peptide linker sequence between the enzyme and the PTS1 tag. We took advantage of the high-throughput nature of our import assay to screen a library of randomized linker sequences inserted immediately preceding the carboxy-terminal SKL residues. This library was made up of six DNK degenerate codon repeats designed to balance the frequency of positively and negatively charged residues and to reduce the frequency of stop codons (**Figure 3-3**). When this VioE-YFP-6X_DNK-SKL library was transformed into a strain expressing VioA and VioB in the cytosol, we observed colonies that ranged in color from white to dark green—presumably the result of differences in VioE peroxisomal import efficiency (**Figure 3-4**).

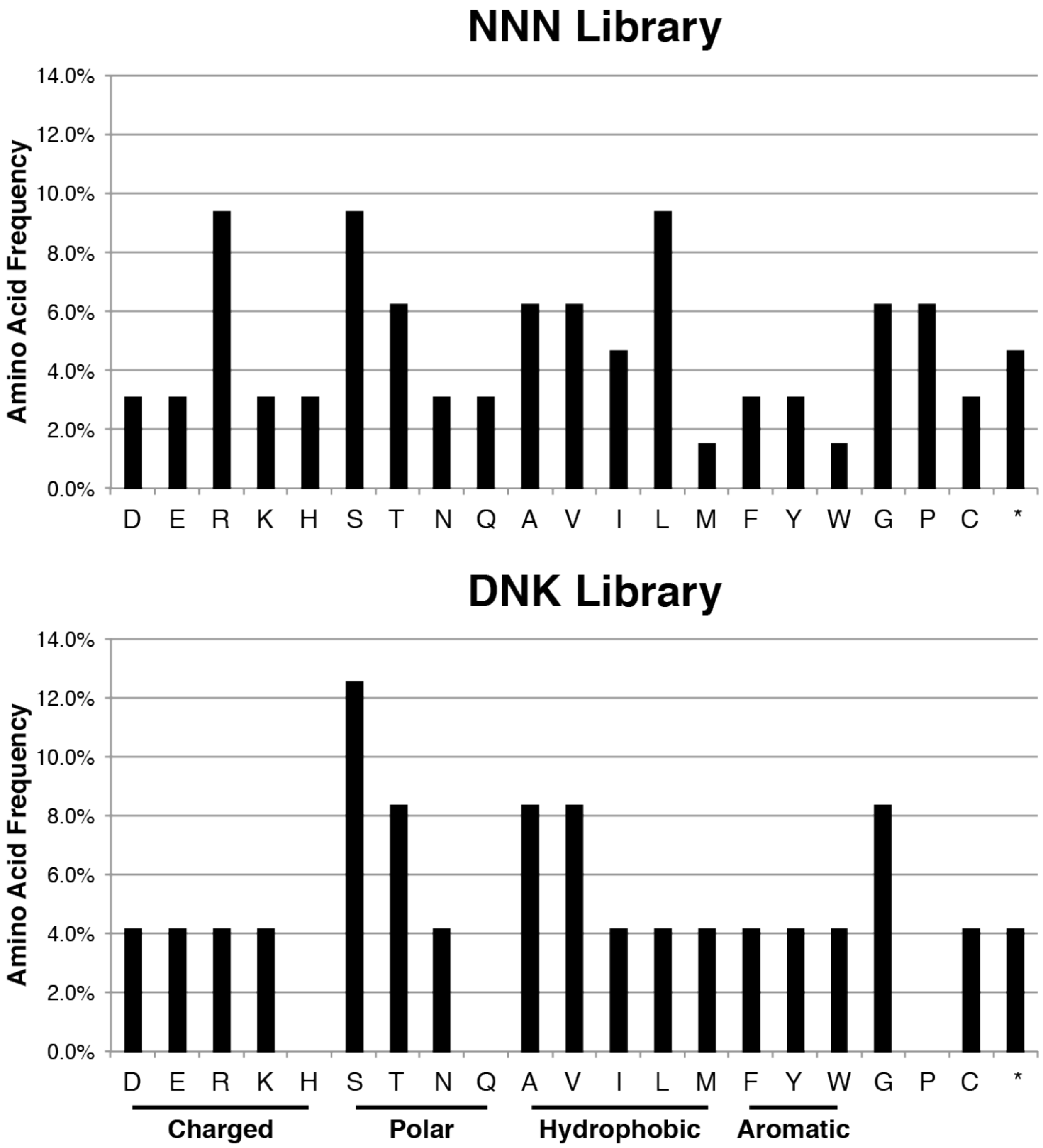


Figure 3-3. Relative frequency of amino acids encoded by the degenerate codons NNN and DNK. Used in the linker library of Fig. 2A, DNK permits the following 24 different nucleotide combinations: (A,G,T)-(A,C,G,T)-(G,T). This codon provides an equal 1/12 likelihood of obtaining a positively charged basic residue (R,K) or a negatively charged acidic residue (D,E), ensuring minimal bias with regard to net charge.

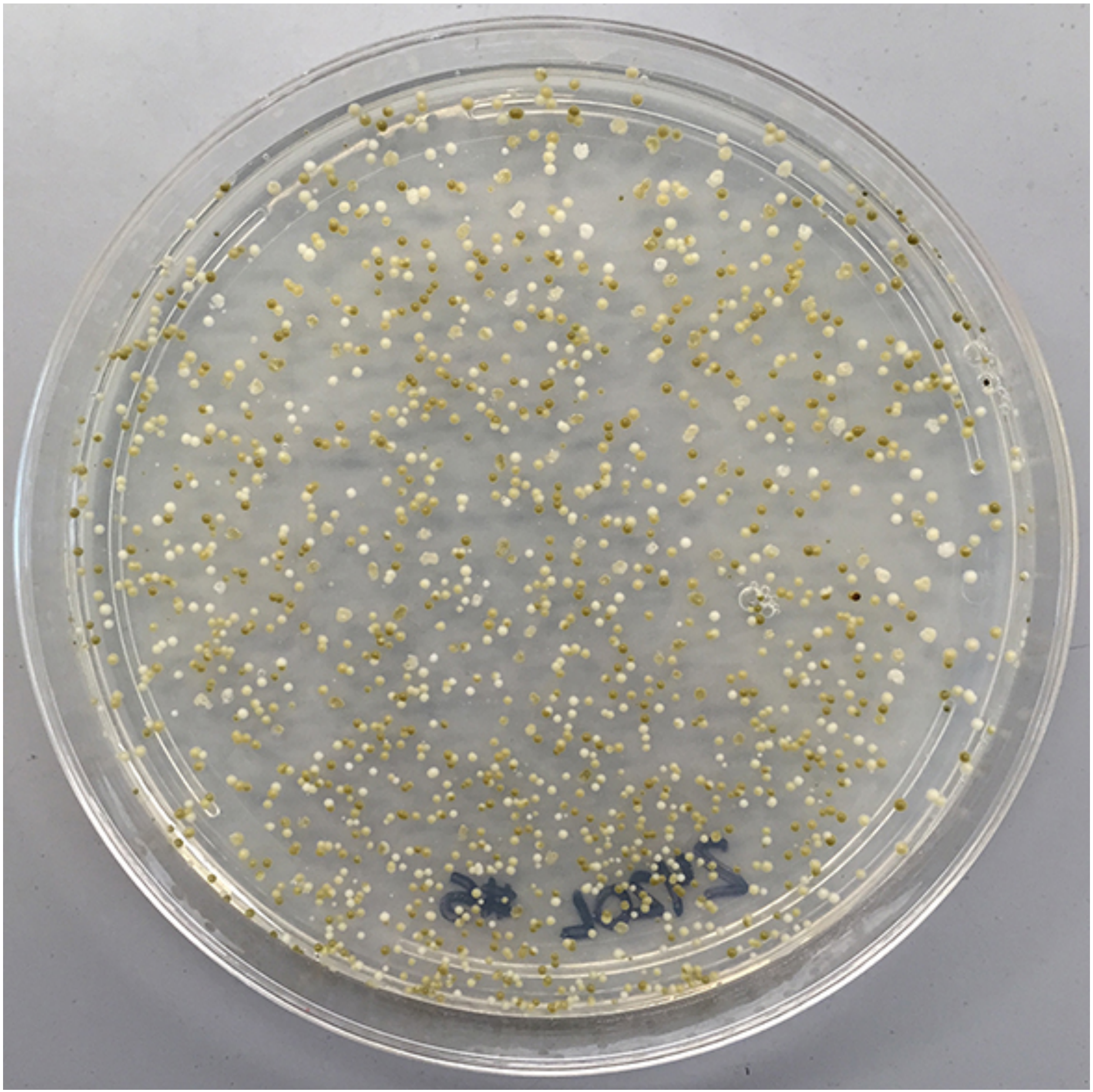


Figure 3-4. Plate-based screening of linker influence on peroxisomal import using a VioE-YFP-linker-PTS1 library. Agar plate with yeast colonies coexpressing cytosolic VioA and VioB and peroxisomally-targeted VioE-YFP-linker-PTS1 after 72 hours of growth. White colonies are indicative of decreased PDV production due to peroxisomal VioE sequestration. Library members from this plate were grown and extracted to generate the data shown in **Figure 3-7a**.

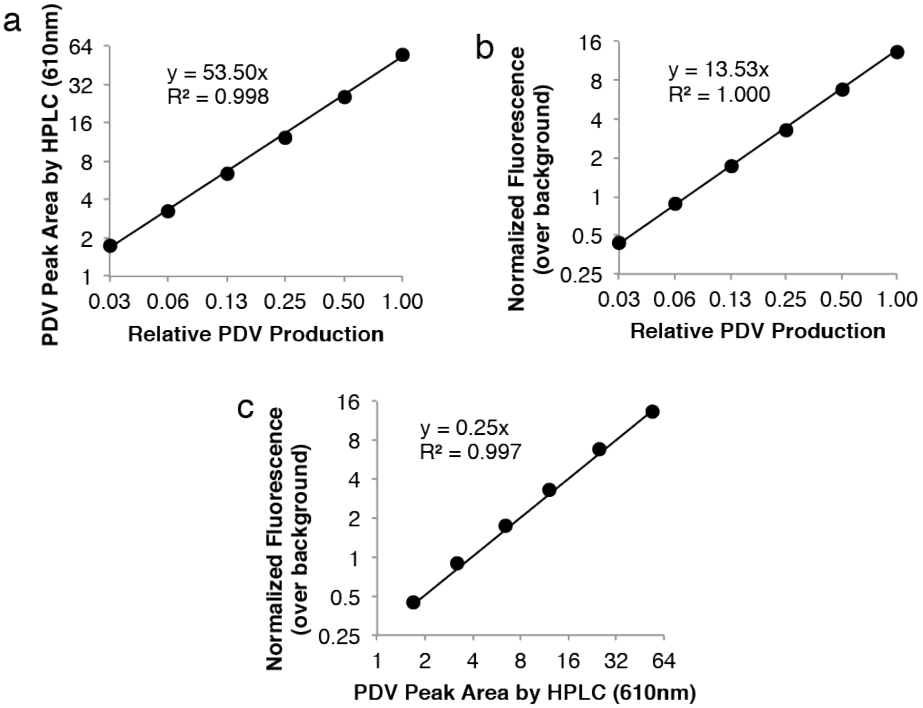
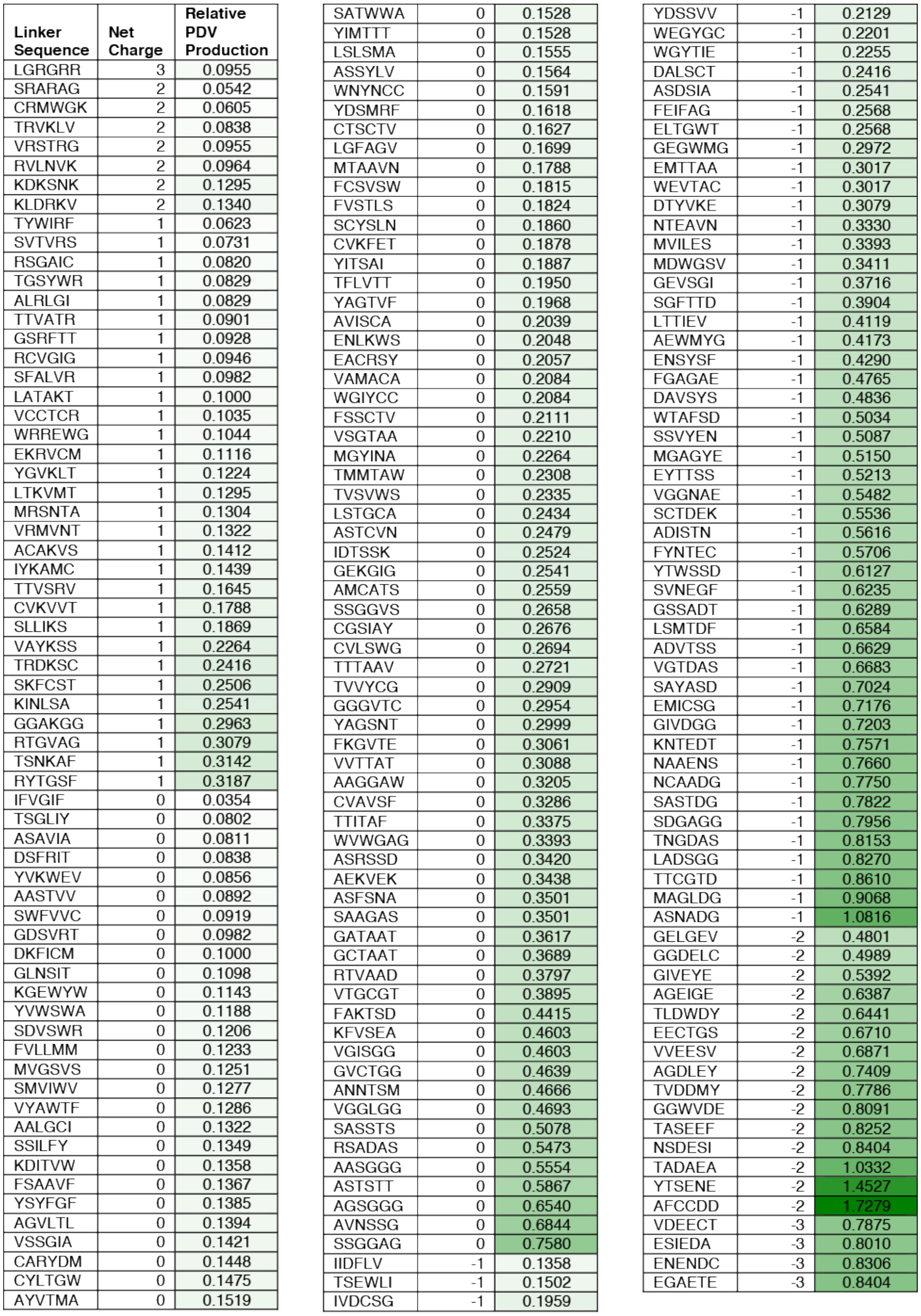


Figure 3-5. HPLC absorbance and bulk fluorescence measurements of PDV show a strong linear correlation with each other and actual PDV production. Cells of high PDV-producing Strain 6 were serially diluted twofold with the cells of non PDV-producing Strain 4. Each dilution was extracted with acetic acid and measured via HPLC and bulk fluorescence to generate six-point calibration curves ranging from 1X to 0.03125X.



Linker Sequence	Net Charge	Relative PDV Production
LGRGRR	3	0.0955
SRARAG	2	0.0542
CRMWGK	2	0.0605
TRVKLV	2	0.0838
VRSTRG	2	0.0955
RVLNVK	2	0.0964
KDKSNK	2	0.1295
KLDRKV	2	0.1340
TYWIRF	1	0.0623
SVTVRS	1	0.0731
RSGAIC	1	0.0820
TGSYWR	1	0.0829
ALRLGI	1	0.0829
TTVATR	1	0.0901
GSRFTT	1	0.0928
RCVGIG	1	0.0946
SFALVR	1	0.0982
LATAKT	1	0.1000
VCCTCR	1	0.1035
WRREWG	1	0.1044
EKRVCM	1	0.1116
YGVKLT	1	0.1224
LTKVMT	1	0.1295
MRSNTA	1	0.1304
VRMVNT	1	0.1322
ACAKVS	1	0.1412
IYKAMC	1	0.1439
TTVSRV	1	0.1645
CVKVVT	1	0.1788
SLLIKS	1	0.1869
VAYKSS	1	0.2264
TRDKSC	1	0.2416
SKFCST	1	0.2506
KINLSA	1	0.2541
GGAKGG	1	0.2963
RTGVAG	1	0.3079
TSNKAF	1	0.3142
RYTGSF	1	0.3187
IFVGIF	0	0.0354
TSGLIY	0	0.0802
ASAVIA	0	0.0811
DSFRIT	0	0.0838
YVKWEV	0	0.0856
AASTVV	0	0.0892
SWFVVC	0	0.0919
GDSVRT	0	0.0982
DKFICM	0	0.1000
GLNSIT	0	0.1098
KGEWYW	0	0.1143
YVWSWA	0	0.1188
SDVSWR	0	0.1206
FVLLMM	0	0.1233
MVGSVS	0	0.1251
SMVIWV	0	0.1277
VYAWTF	0	0.1286
AALGCI	0	0.1322
SSILFY	0	0.1349
KDITVW	0	0.1358
FSAAVF	0	0.1367
YSYFGF	0	0.1385
AGVLTL	0	0.1394
VSSGIA	0	0.1421
CARYDM	0	0.1448
CYLTGW	0	0.1475
AYVTMA	0	0.1519
SATWWA	0	0.1528
YIMTTT	0	0.1528
LSLSMA	0	0.1555
ASSYLV	0	0.1564
WNYNCC	0	0.1591
YDSMRF	0	0.1618
CTSCTV	0	0.1627
LGFAGV	0	0.1699
MTAAVN	0	0.1788
FCSVSW	0	0.1815
FVSTLS	0	0.1824
SCYSLN	0	0.1860
CVKFET	0	0.1878
YITSAI	0	0.1887
TFLVTT	0	0.1950
YAGTVF	0	0.1968
AVISCA	0	0.2039
ENLKWS	0	0.2048
EACRSY	0	0.2057
VAMACA	0	0.2084
WGIYCC	0	0.2084
FSSCTV	0	0.2111
VSGTAA	0	0.2210
MGYINA	0	0.2264
TMMTAW	0	0.2308
TVSVWS	0	0.2335
LSTGCA	0	0.2434
ASTCVN	0	0.2479
IDTSSK	0	0.2524
GEKGIG	0	0.2541
AMCATS	0	0.2559
SSGGVS	0	0.2658
CGSIAY	0	0.2676
CVLSWG	0	0.2694
TTTAAV	0	0.2721
TVVYCG	0	0.2909
GGGVTC	0	0.2954
YAGSNT	0	0.2999
FKGVTE	0	0.3061
VVTTAT	0	0.3088
AAGGAW	0	0.3205
CVAVSF	0	0.3286
TTITAF	0	0.3375
WVWGAG	0	0.3393
ASRSSD	0	0.3420
AEKVEK	0	0.3438
ASFSNA	0	0.3501
SAAGAS	0	0.3501
GATAAT	0	0.3617
GCTAAT	0	0.3689
RTVAAD	0	0.3797
VTGCGT	0	0.3895
FAKTSD	0	0.4415
KFVSEA	0	0.4603
VGISGG	0	0.4603
GVCTGG	0	0.4639
ANNTSM	0	0.4666
VGGLGG	0	0.4693
SASSTS	0	0.5078
RSADAS	0	0.5473
AASGGG	0	0.5554
ASTSTT	0	0.5867
AGSGGG	0	0.6540
AVNSSG	0	0.6844
SSGGAG	0	0.7580
IIDFLV	-1	0.1358
TSEWLI	-1	0.1502
IVDCSG	-1	0.1959
YDSSVV	-1	0.2129
WEGYGC	-1	0.2201
WGYTIE	-1	0.2255
DALSCT	-1	0.2416
ASDSIA	-1	0.2541
FEIFAG	-1	0.2568
ELTGWT	-1	0.2568
GEGWMG	-1	0.2972
EMTTAA	-1	0.3017
WEVTAC	-1	0.3017
DTYVKE	-1	0.3079
NTEAVN	-1	0.3330
MVILES	-1	0.3393
MDWGSV	-1	0.3411
GEVSGI	-1	0.3716
SGFTTD	-1	0.3904
LTTIEV	-1	0.4119
AEWMYG	-1	0.4173
ENSYSF	-1	0.4290
FGAGAE	-1	0.4765
DAVSYS	-1	0.4836
WTAFSD	-1	0.5034
SSVYEN	-1	0.5087
MGAGYE	-1	0.5150
EYTTSS	-1	0.5213
VGGNAE	-1	0.5482
SCTDEK	-1	0.5536
ADISTN	-1	0.5616
FYNTEC	-1	0.5706
YTWSSD	-1	0.6127
SVNEGF	-1	0.6235
GSSADT	-1	0.6289
LSMTDF	-1	0.6584
ADVTSS	-1	0.6629
VGTDAS	-1	0.6683
SAYASD	-1	0.7024
EMICSG	-1	0.7176
GIVDGG	-1	0.7203
KNTEDT	-1	0.7571
NAAENS	-1	0.7660
NCAADG	-1	0.7750
SASTDG	-1	0.7822
SDGAGG	-1	0.7956
TNGDAS	-1	0.8153
LADSGG	-1	0.8270
TTCGTD	-1	0.8610
MAGLDG	-1	0.9068
ASNADG	-1	1.0816
GELGEV	-2	0.4801
GGDELC	-2	0.4989
GIVEYE	-2	0.5392
AGEIGE	-2	0.6387
TLDWDY	-2	0.6441
EECTGS	-2	0.6710
VVEESV	-2	0.6871
AGDLEY	-2	0.7409
TVDDMY	-2	0.7786
GGWVDE	-2	0.8091
TASEEF	-2	0.8252
NSDESI	-2	0.8404
TADAEA	-2	1.0332
YTSENE	-2	1.4527
AFCCDD	-2	1.7279
VDEECT	-3	0.7875
ESIEDA	-3	0.8010
ENENDC	-3	0.8306
EGAETE	-3	0.8404

Figure 3-6. PDV production linked to specific sequences from the 6-amino acid linker library. All 200 linker sequences measured in the library are shown here, along with a calculation of their net charge and the cells' PDV production relative to that of a defective peroxisomal import mutant (*pex5Δ*). PDV production was measured by bulk fluorescence at excitation/emission 535/585 nm. The data shown in this figure was used to generate **Figure 3-7a**.

To gain insight into how linker sequences affect VioE targeting, we sequenced the linker region of randomly picked colonies from the library transformation plate, obtaining sequences for 200 unique clones after removal of those with stop codons in the linker. In order to quantify PDV production in these strains, we developed a medium-throughput, microplate-based method for assaying PDV pathway flux. This method takes advantage of our observation that cells producing PDV are highly red fluorescent, presumably due to fluorescence of either PDV or a PDV derivative. After confirming that fluorescence measurements of extracts yielded an excellent linear correlation with HPLC-measured PDV production across a wide range of concentrations (**Figure 3-5**), we used this method to quickly screen our 200-linker library members (**Figure 3-6**). Analysis of these data showed a correlation between a linker's net charge and the amount of PDV produced (R^2 = 0.54 and standard error of regression (*SER*) = 0.1928 when fit to a logistic curve) (**Figure 3-7a**). In general, clones with basic residues (Arg or Lys) in their linker produced low levels of PDV, indicating efficient targeting of VioE to the peroxisome. The correlation improved when we doubled the weight applied to the charge of the three linker residues immediately preceding the canonical PTS1 tag (R^2 = 0.61 and *SER* = 0.1781), establishing charge of the residues closest to SKL as a major determinant of import efficiency.

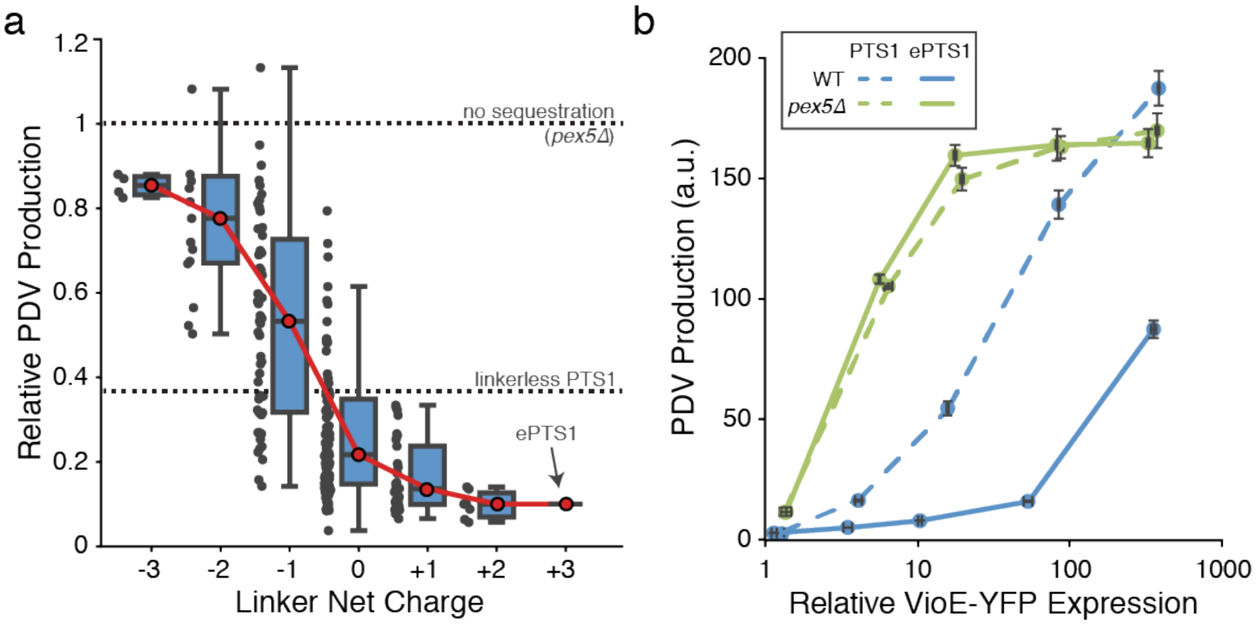


Figure 3-7. Library screening indicates that positively charged linker residues enhance PTS1-based import of VioE. (**a**) A randomized library of six amino acid linkers was inserted between VioE-YFP and the PTS1 tag and coexpressed with cytosolically-localized VioA and VioB. Individual spots represent PDV-linked fluorescence measurements of two hundred arbitrarily selected library members (excluding two outliers with values 1.45 and 1.73 at the -2 charge level for clarity), normalized relative to a *pex5Δ* strain and grouped by net charge of the linker residues. Quartiles and extrema for each net charge level are shown with a box-and-whiskers plot. The median PDV productivity at each net charge is further highlighted with a red line. Efficient peroxisomal import of VioE is correlated with decreased PDV production. We selected the sole +3 charge sequence, LGRGRR-SKL, for further study as ePTS1. (**b**) HPLC quantification of PDV from yeast strains expressing increasing levels of VioE-YFP fused to either naïve PTS1 (dashed lines) or ePTS1 (solid lines). Both wildtype (blue lines) and *pex5Δ* (green lines) strain backgrounds coexpressed VioA and VioB cytosolically. Error bars represent the mean ± s.d. of four biological replicates.

3.2.3. A modular enhanced PTS1 tag enables rapid peroxisome targeting

In search of a generalizable strategy for achieving efficient enzyme targeting, we decided to further characterize a single linker, LGRGRR, which appeared in one of our lowest PDV-producers and contains three basic residues (**Figure 3-7a**). We named the combined linker-PTS1 tag (LGRGRR-SKL) ePTS1 for “enhanced PTS1”. To compare how much VioE could be sequestered with ePTS1 or the canonical PTS1 tag before high levels of cytosolic activity could be detected, we again fused these tags to VioE-YFP and performed enzyme titrations using a set of five promoters that span approximately three orders of magnitude in expression strength[118] (**Figure 3-7b** and **Figure 3-8**). These promoters are, from weakest to strongest, pREV1, pRNR2, pRPL18B, pTEF1, and pTDH3, with the last four promoters representing ~4X, 12X, 45X, and 140X the expression strength of pREV1. At the moderate pRPL18B expression level used earlier (**Figure 3-2**), ePTS1 yielded a 95% reduction in PDV compared to a 63% reduction with PTS1 (**Figure 3-7b**) relative to their corresponding *pex5Δ* control strains. When VioE was expressed under the strongest yeast promoter (pTDH3), the ePTS1 tag still reduced PDV production by half while the PTS1 tag was saturated to the point of being indistinguishable from its *pex5Δ* control. In total, these results indicated that the ePTS1 tag directed VioE to the peroxisome far more efficiently than did the canonical PTS1 tag.

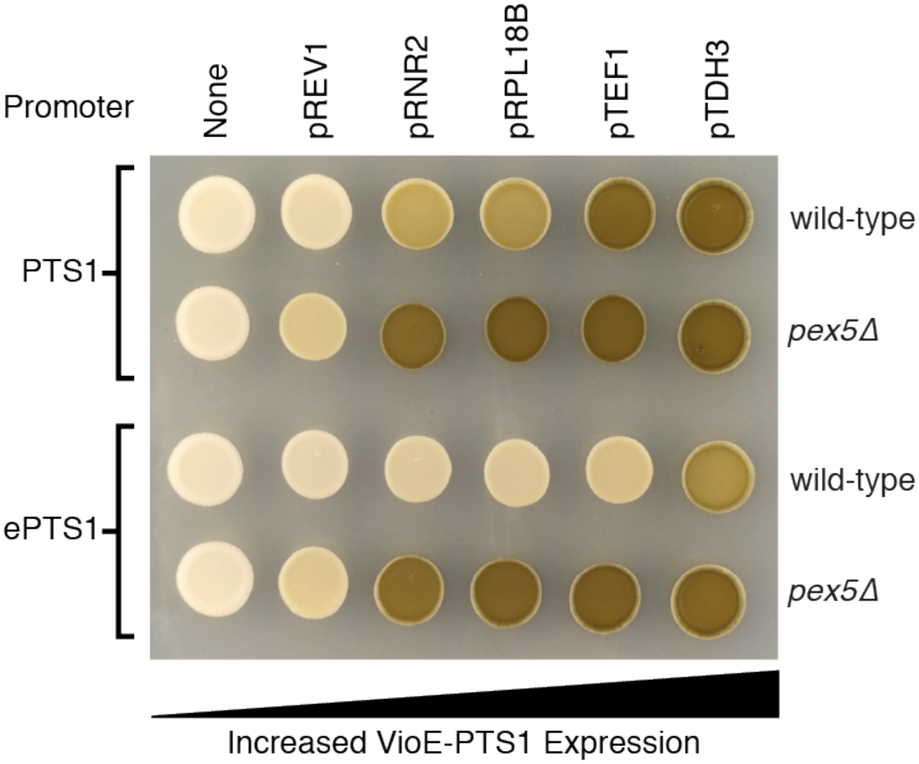


Figure 3-8. Yeast spots comparing PTS1 and ePTS1 sequestration efficiency for VioE-YFP. 10 μL of saturated culture from each of the same strains used to generate **Figure 3-7b** was spotted onto agar plates and grown for 48 hours. As in **Figure 3-7b**, Strains 12-33 coexpressed cytosolic VioA and VioB along with VioE-YFP tagged with either PTS1 or ePTS1. Defective peroxisomal import controls (*pex5Δ*) are also shown.

While this VioE titration validated ePTS1's utility for efficiently compartmentalizing large quantities of heterologous enzyme in the peroxisome, we wondered whether the rate of VioE import was responsible for the improved performance of the ePTS1 tag. To test the rate of cargo import, we designed a competition experiment that pitted peroxisomal import of a cargo protein against proteolysis of that same cargo by a cytosolically expressed protease (**Figure 3-9a**). Our cargo was an

RFP-YFP fusion protein with a tobacco etch virus (TEV) cleavage site inserted between the two fluorescent proteins and a PTS1 or ePTS1 tag fused to the C-terminus. By cytosolically expressing TEV protease (TEVp), we could gauge the speed with which the cargo was imported, since slow import would allow TEV site cleavage and cytosolic accumulation of RFP. Titrating TEV protease expression permitted us to modulate the rate of TEV site cleavage.

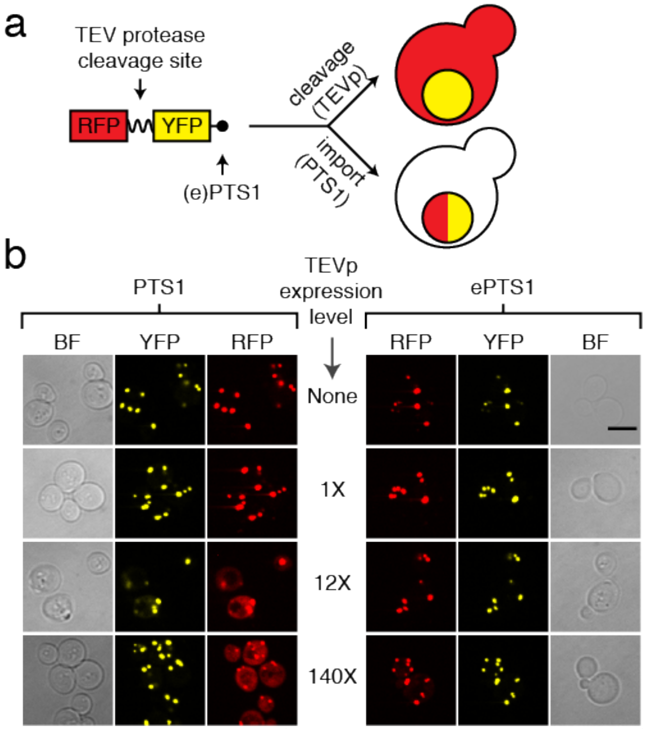


Figure 3-9. ePTS1-mediated peroxisomal import outpaces cytosolic proteolysis. (**a**) An assay for the relative speed of peroxisomal import was devised in which TEV protease-mediated protein cleavage competes with PTS1-mediated peroxisomal protein import. The RFP-YFP fusion protein shown here experiences two possible fates: either the fusion protein is cleaved prior to import, resulting in a diffuse red fluorescence, or peroxisomal import occurs prior to cleavage, resulting in punctate red fluorescence. In either case, yellow fluorescence is punctate, as peroxisomal import of YFP is not affected by proteolysis. (**b**) Brightfield (BF) and confocal fluorescence microscopy of yeast cells constitutively expressing an RFP-YFP-PTS1 or –ePTS1 fusion along with varying levels of TEV protease fused to CFP for quantification. The relative expression of TEV protease includes no TEVp, a baseline level (1X), as well as 12-fold and 140-fold above that baseline. Scale bar, 10 μm.

Our results showed the ePTS1-tagged cargo was imported much faster than cargo targeted via the canonical PTS1 (**Figure 3-9b**). At the highest expression level of TEV protease, the mean RFP brightness in the cytosol (approximating protein concentration) was one-fourth of the peroxisomal brightness with the PTS1 tag, while ePTS1 showed a mean cytosolic brightness that was approximately 50-fold lower than peroxisomal brightness. Based on the volumes of the two compartments[119], approximately 3% of the total RFP was imported before cleavage when using PTS1, versus 30% with ePTS1. Western blot analysis detected no difference between PTS1 and ePTS1 with respect to the fraction of cleaved cargo at any TEVp expression level (**Figure 3-10**), but the assay was likely confounded by TEV protease "piggybacking" into the peroxisome via transient association with its substrate[120,121] and cleaving the RFP-YFP inside the peroxisome.

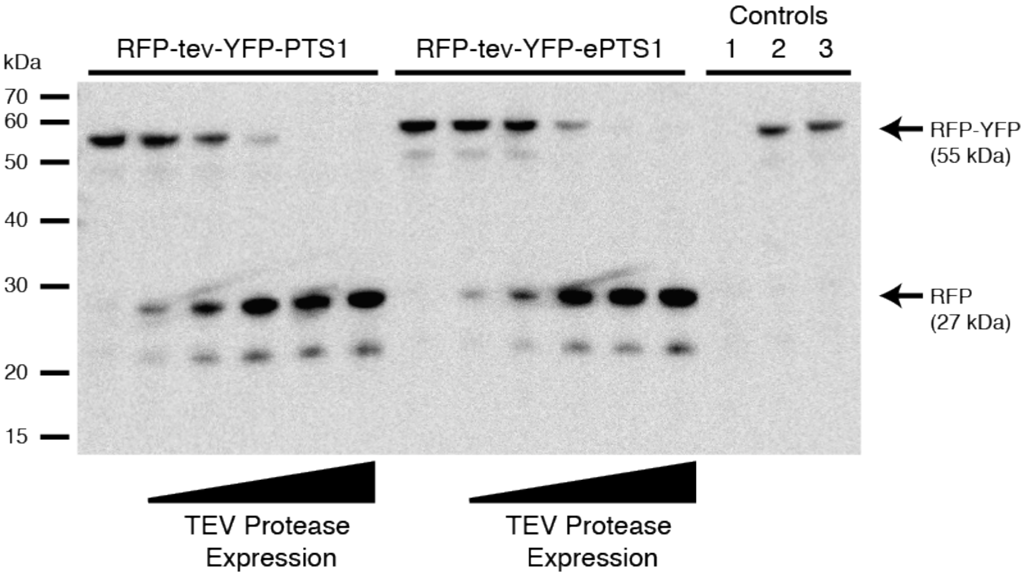


Figure 3-10. Western blot analysis indicates that proteolysis of the RFP-tev-YFP fusion protein may occur in both the peroxisome and cytosol. The strains from **Figure 3-9**, as well as strains expressing no TEV or intermediate levels of TEV (pRNR2 and pTEF1 at 4X and 45X respectively), were lysed and run on a denaturing gel, then blotted and stained with an antibody recognizing RFP. Full length (55 kDa) and proteolyzed (27 kDa) protein bands are indicated with arrows. Three controls were included to verify that TEV cleavage did not occur after cell lysis: 1) Strain 53 expressing pTDH3-TEV (140X) with no RFP-tev-YFP-ePTS1, 2) Strain 40 expressing RFP-tev-YFP-ePTS1 mixed 1:1 with Strain 52, which has only selection markers and 3) a 1:1 mixture of Strain 40 (expressing TEV protease) and Strain 53 (expressing RFP-tev-YFP-ePTS1). No RFP antibody cross-reactivity for the CFP on TEV protease was observed.

We suspected that greater affinity for the import receptor Pex5p played a role in the increased speed of import with ePTS1 versus PTS1. By subjecting a C-terminal truncation of Pex5p and synthetic PTS1 peptides to *in vitro* binding analysis via fluorescence anisotropy, we found a 200-fold difference in affinity between the ePTS1 peptide (LGRGRRSKL) and a peptide of equal length that results from fusion of the canonical PTS1 to the C-terminus of YFP (DELYKGSKL) (**Figure 3-11**). The dissociation constant for the ePTS1 peptide was 0.8 ± 0.1 μM, compared to 175.6 ± 17.3 μM for the PTS1 peptide. These affinities were weaker than expected based on previous measurements[107], but this discrepancy could be attributed to our use of fluorescein-tagged peptides instead of lissamine-tagged or our use of *S. cerevisiae* Pex5p instead of human Pex5p. Regardless, the large difference in PTS1 and ePTS1 affinity for Pex5p provides some evidence for the mechanism behind the improved efficiency of ePTS1.

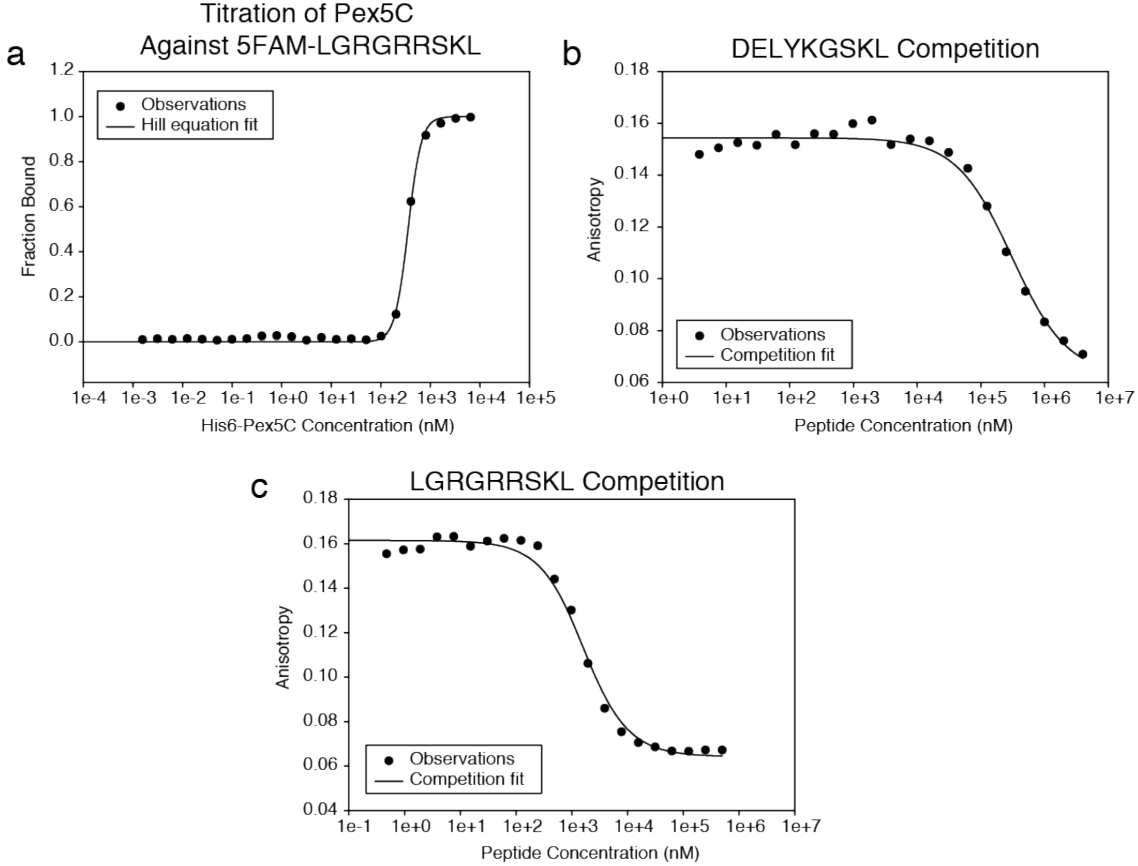


Figure 3-11. Pex5p has greater affinity for ePTS1. (**a**) Direct binding assay of the C-terminal fragment of *S. cerevisiae* Pex5p against 20 nM fluorescein-labeled ePTS1 peptide 5FAM-LGRGRRSKL. Fitting this data to the Hill equation found that the K_D of Pex5p fragment binding to 5FAM-LGRGRRSKL was 355±6.38 nM with a Hill coefficient of n = 3.275±0.176. (**b,c**) Competitive binding of unlabeled PTS1 and ePTS1 peptides against 20 nM of fluorescein-labeled ePTS1 peptide. Peptides were competing for binding to 375 nM of Pex5p C-terminal fragment, and the state of the labeled ePTS1 peptide was tracked via fluorescence anisotropy. A non-cooperative model from Roerhl et al. was fit to the data, allowing estimates of relative affinities of PTS1 and ePTS1. (**b**) The unlabeled PTS1 peptide used was DELYKGSKL, reflecting the last nine amino acids of the VioE-YFP-PTS1 constructs used in Fig. 2b. The K_D found was 176000±17300nM. (**c**) The ePTS1 peptide used was LGRGRRSKL, and the K_D was 801±105 nM.

As a final test of ePTS1's utility, we examined the tag's modularity using an inducible sequestration assay in which the native *pex5* promoter was replaced with the galactose-inducible *gal1* promoter. Upon galactose-induction of Pex5p and thus peroxisomal import, we observed that RFP-ePTS1 was completely compartmentalized in the peroxisome after only one hour in this strain background (**Figure 3-12**). We fused the PTS1 or ePTS1 tag to one of four essential genes—*cdc14*, *cdc28*, *tys1*, and *spc42*—expecting that sequestration of these proteins in the peroxisome would prevent growth. Indeed, we observed that all four ePTS1-tagged constructs yielded strains that were unable to grow under galactose induction (**Figure 3-13**). In contrast, the canonical PTS1 tag arrested growth of only two of four strains, suggesting that Cdc14p-SKL and Cdc28p-SKL were inadequately compartmentalized. We also tested these constructs with *pex14Δ* strains, which are missing a key component of the import pore complex necessary for peroxisomal protein import. As expected, all *pex14Δ* strains grew on galactose-containing media, indicating that the growth-arrest effects are import-dependent and not an artifact of Pex5p binding. These results demonstrate that the ePTS1 tag maintains its rapid import properties when fused to a variety of cargo proteins.

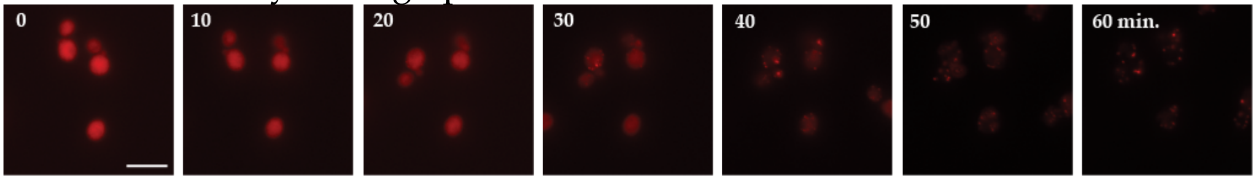


Figure 3-12. Fluorescence tracking of RFP-ePTS1 import upon induction of Pex5p expression. At time 0, Strain 55 was transferred into synthetic complete medium with 2% raffinose and 1% galactose for induction of Pex5p expression. Fluorescence images were taken at ten-minute intervals as the newly created Pex5p shuttled RFP-ePTS1 into the peroxisomes. By 60 minutes, the cells had been virtually cleared of cytosolic RFP-ePTS1. Scale bar, 10 μM.

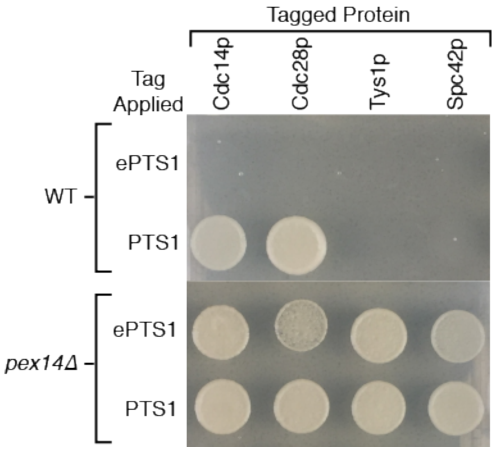


Figure 3-13. The ePTS1 tag modularity allows for rapid sequestration of a variety of proteins. Sequestration of essential proteins and subsequent growth arrest was tested using yeast strains with galactose-inducible control of Pex5p expression. Strains shown were modified so that one of four essential proteins (Cdc14p, Cdc28p, Tys1p, or Spc42p) was individually fused to either ePTS1 or naïve PTS1. Control strains with a permanently defective peroxisomal import system (*pex14Δ*) were also included. Strains were spotted onto agar plates with galactose to induce Pex5p expression and grown for 48 hours.

3.2.4. Small metabolites freely traverse the peroxisomal membrane

Having improved our control over protein targeting, we turned to metabolite transport and applied our new tools to the challenges of measuring *in vivo* transport through subcellular compartments. Using β-glucosidase (BGL), a promiscuous enzyme from *Neurospora crassa*[122] that accepts a wide variety of glycosides for hydrolysis, we designed a novel *in vivo* polymer exclusion assay to measure peroxisomal permeability. We reasoned that if the previously-hypothesized[111] aqueous pore were indeed present on the peroxisomal membrane, peroxisomally-localized BGL would have activity on small substrates but not larger ones (**Figure 3-14a**).

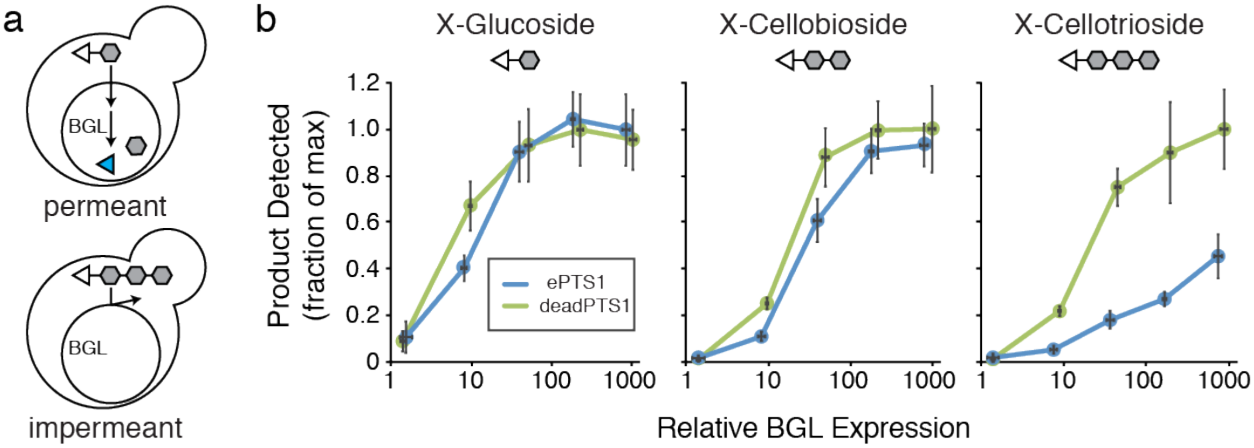


Figure 3-14. Oligosaccharide-based assay reveals size-dependent permeability of the peroxisomal membrane. (**a**) Yeast cells expressing a peroxisomally localized β-

glucosidase (BGL) were fed 5-bromo-4-chloro-indoxyl (X) dye molecules (triangles) conjugated to glucose chains of increasing length (hexagons). BGL can cleave the dye-sugar conjugates, releasing the fluorescent dye molecule (blue triangle), but only if those conjugates are able to cross the peroxisomal membrane to access the BGL. Impermeant substrates should be unable to access the BGL and thus should generate no fluorescence. (**b**) Relative fluorescent product release in yeast strains expressing increasing amounts of peroxisomal (blue lines) or cytosolic (green lines) β-glucosidase. Dye-sugar reactions involving X-glucoside (M_w = 408.63 Da), X-cellobioside (M_w = 570.78 Da), and X-cellotrioside (M_w = 732.93 Da) are described in detail in **Figure 3-15**. Error bars are the mean ± s.d. of six biological replicates.

We selected three structurally-similar, size-diverse substrates for testing, all of which are oligosaccharides conjugated to a dye molecule: X-glucoside (molecular weight M_w = 409 Da), X-cellobioside (M_w = 571 Da), and X-cellotrioside (M_w = 733 Da), where "X" represents 5-bromo-4-chloro-3-indolyl (**Figure 3-15**). When these substrates are hydrolyzed by BGL, they release 5-bromo-4-chloro-indoxyl, a fluorescent molecule that spontaneously oxidizes into the blue pigment 5,5′-dibromo-4,4′-dichloro-indigo. Because the substrates are hydrophilic, they cannot cross phospholipid membranes—including the plasma membrane—without the aid of a protein transporter. To import the substrates into the cytosol, we expressed the cellobiose transporter, CDT1, from *N. crassa*[122], which recognized all three substrates (**Figure 3-16**). Once again, we used a set of constitutive promoters to titrate the expression of BGL, and created cytosolic controls with a deadPTS1 tag featuring a C-terminal leucine to threonine mutation. We also augmented BGL by fusion to VioE and YFP, which enabled us to confirm consistent expression via YFP fluorescence, and verify peroxisomal import via VioE activity measurement. When VioE-YFP-BGL-ePTS1 or –deadPTS1 were expressed in strains with cytosolic VioA and VioB, we observed a decrease in VioE activity upon ePTS1 tagging, confirming BGL was contained in the peroxisome (**Figure 3-17**).

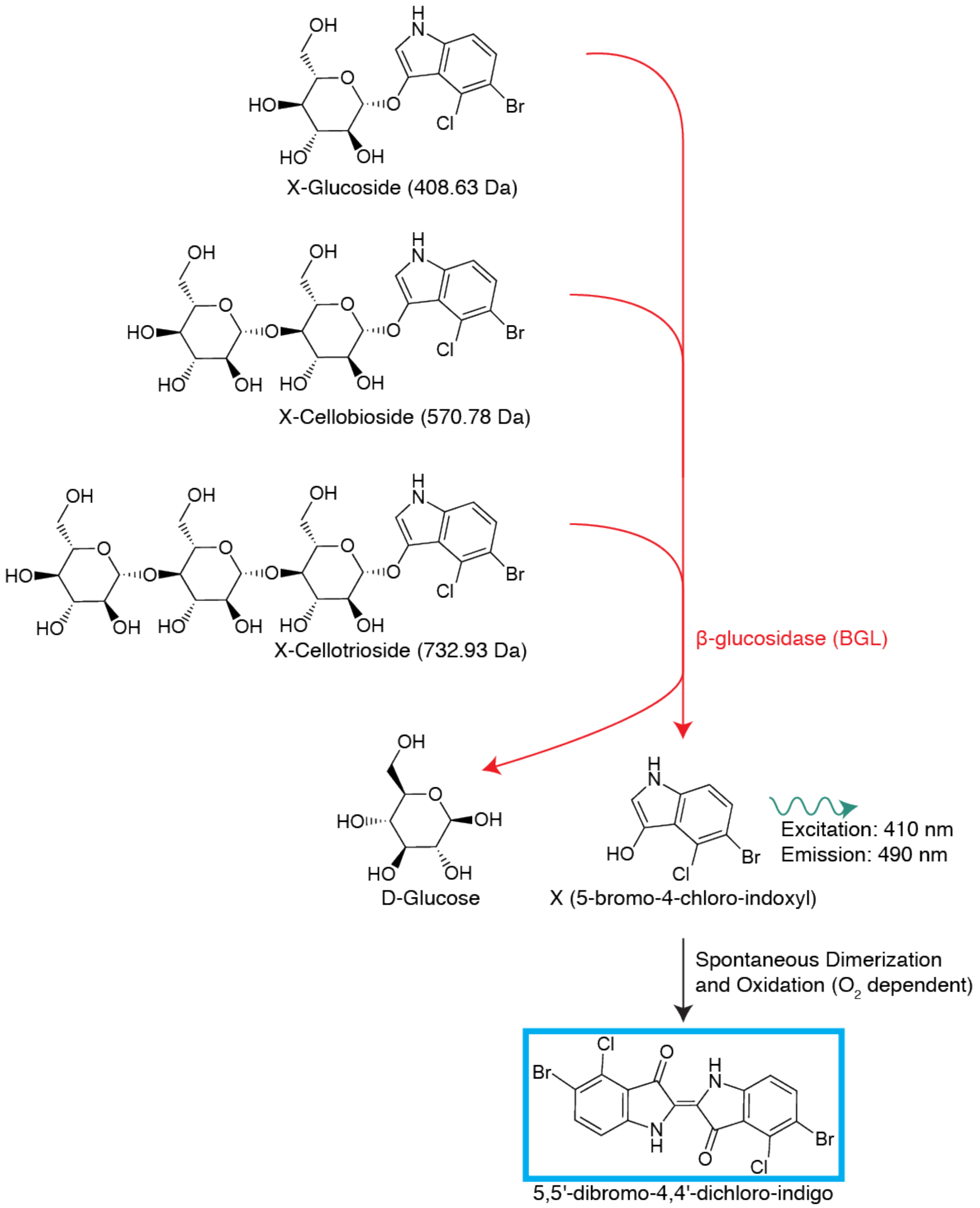


Figure 3-15. Conversion of dye-substrate conjugates into fluorescent product. Upon feeding of X-dyes (also called BCI or 5-bromo-4-chloro-3-indoxyl dyes), the *N. crassa* β-glucosidase (BGL) cleaves off sugar groups, releasing fluorescent BCI and glucose. In anaerobic environments, the spontaneous reaction of BCI to produce 5,5′-dibromo-4,4′-dichloro-indigo is diminished, extending the lifetime of the transient BCI dye and allowing fluorescent quantitation of BGL activity. In yeast, the X-dyes require expression of cellodextrin transporter CDT1 to be able to enter cells, and in the experiments shown in **Figure 3-14**, BGL is expressed in either the peroxisome or cytosol.

We then fed the X-substrates to our CDT1- and BGL-expressing strains and compared the degree to which each substrate was hydrolyzed in strains with cytosolic or peroxisomal BGL (**Figure 3-14b**). For the two smallest substrates, X-glucoside and X-cellobioside, we observed little to no effect of peroxisomal BGL import on hydrolysis. However, a significant sequestration effect was observed with the largest substrate, X-cellotrioside, with a 77% reduction in hydrolysis upon peroxisomal import at moderate pRPL18B levels of BGL. This indicates that peroxisomal transport impeded hydrolysis of the larger metabolite but not smaller ones, and the size cutoff matched well with *in vitro* polymer exclusion experiments on purified peroxisomes[114] (**Figure 3-18**).

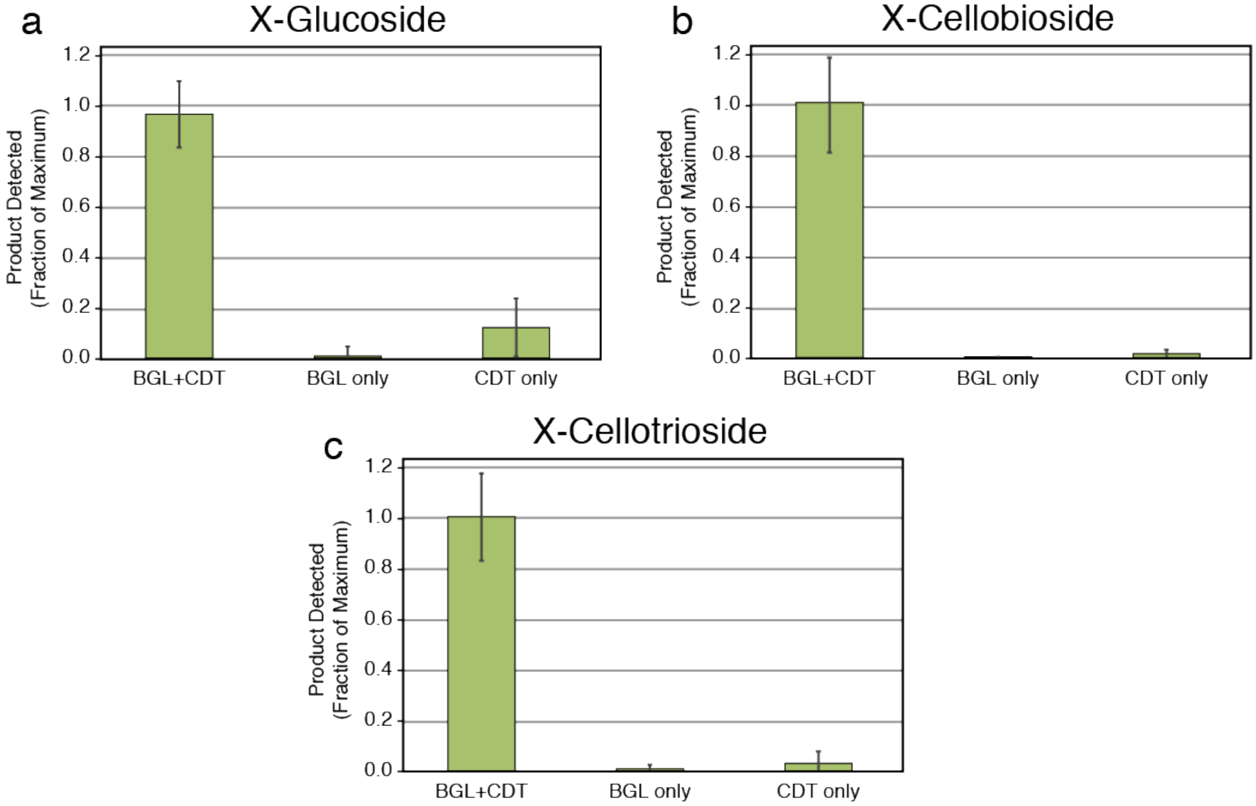


Figure 3-16. Hydrolysis of X-glycosides requires coexpression of CDT1 and BGL. Strains 117, 122, and 123 express CDT1 and cytosolic BGL, BGL only, and CDT1 only, respectively. These cells were treated with X-glycosides as in **Figure 3-14**. Error bars are the mean ± s.d. of six biological replicates.

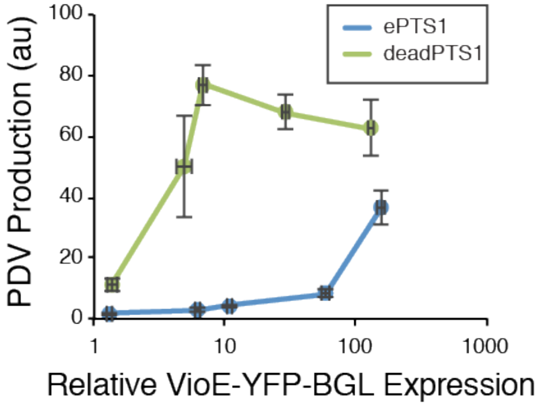


Figure 3-17. Measurement of BGL import into the peroxisomes using the VioE-fusion assay. To confirm efficient BGL targeting to the peroxisome, Strain 2 coexpressing VioA and VioB cytosolically was transformed with expression cassettes for VioE-YFP-BGL-ePTS1 (the same cassettes used to make the strains shown in **Figure 3-14**), generating Strains 124-134. These strains allowed direct testing of VioE activity by the same methods of extraction and fluorescent quantification used in **Figure 3-7a**. The sequestration of VioE provides confirmation that the entire fusion protein is being successfully imported into the peroxisome. Error bars are the mean ± s.d. of eight biological replicates.

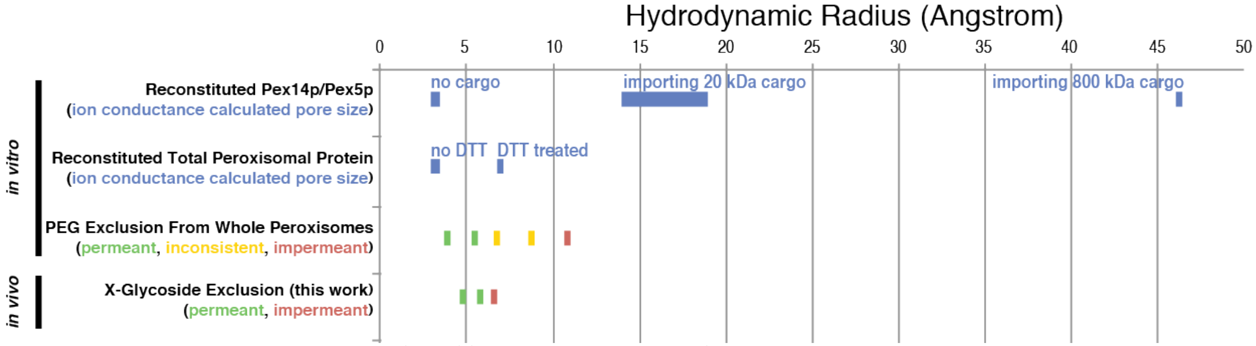


Figure 3-18. Comparison of different methods for calculating peroxisomal pore size. Bars show either peroxisomal pore size estimated from an ion conductance experiment (blue) or polymers tested in exclusion experiments that consistently crossed the peroxisomal membrane (green), inconsistently crossed (yellow), or never crossed (red). Ion conductance experiments used reconstituted protein in lipid bilayers, either from total peroxisomal protein[115,123] or purified Pex14p/Pex5p[104]. Polymer exclusion experiments measured polyethylene glycol (PEG) entering whole purified peroxisomes[114] *in vitro* or X-glycosides entering peroxisomes *in vivo* as described in this work. PEG sizes tested had molecular weights of 200, 400, 600, 1000, and 1500 Da[114] and these were converted to hydrodynamic radius via previously published data[124]. X-glycosides were converted to hydrodynamic radii using experimental data for their cellodextrin counterparts[125,126]. The polymer exclusion experiments with PEG *in vitro* and X-glycosides *in vivo* agree well, placing an estimate of pore radius between 5.7 and 6.5 angstrom.

Given these findings, we were interested in the effect that non-specific permeability would have on our ability to sequester the PDV pathway. We fused the ePTS1 tag directly to the C-terminus of each enzyme and assayed activity in wildtype and *pex5Δ* strains with cytosolic expression of the other two pathway enzymes. Across a range of enzyme expression levels, we observed a considerable drop in PDV production when either VioB or VioE was targeted to the peroxisome, but not VioA (**Figure 3-19**). The effect of VioB sequestration in the peroxisome peaked at a 4.5-fold drop in activity (at the pTEF1 promoter level), while sequestration of VioE achieved 13-fold reduced activity (at the pRNR2 and pRPL18B promoter levels). This relative difference in sequestration performance may be due to less efficient import of the bulky (111 kDa) VioB or directional differences in IPA imine dimer flux across the peroxisomal membrane. Since the volume of peroxisomes is roughly one hundredth that of the cytosol[119], it would be expected that the gradient driving IPA imine dimer across the peroxisomal membrane is up to 100-fold greater when VioB is in the peroxisome. Though VioB sequestration was not as effective as VioE sequestration, the two enzymes' shared intermediate, IPA imine dimer, appears to be mostly peroxisome-impermeant.

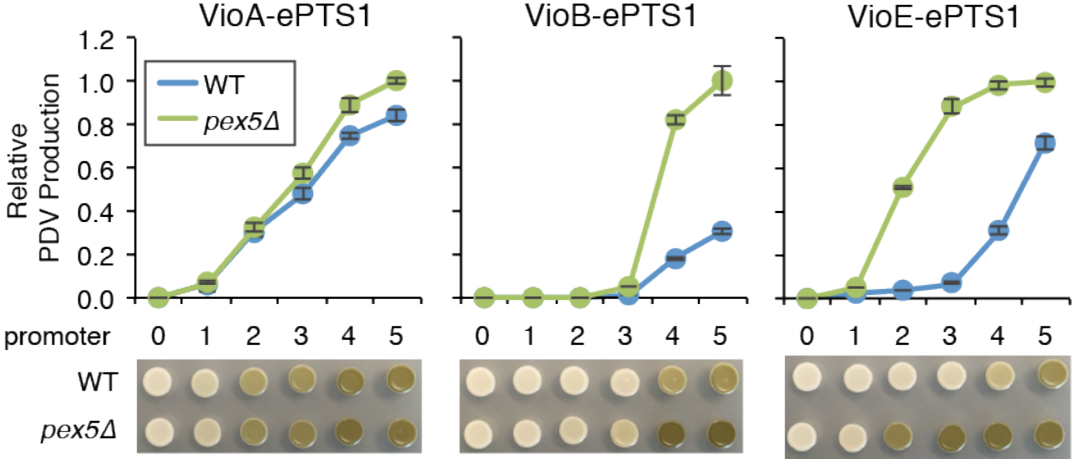


Figure 3-19. Characterization of peroxisomal permeability in the PDV pathway. PDV was extracted from yeast strains expressing an ePTS1-tagged version of VioA, VioB, or VioE together with untagged versions of the remaining two enzymes. Strains were constructed in either a wildtype (blue lines) or *pex5Δ* (green lines) background, with the ePTS1-tagged enzyme being expressed under a variety of promoters, numbered in rank order of expression strength: 0=no gene present, 1=pREV1, 2=pRNR2, 3=pRPL18B, 4=pTEF1, 5=pTDH3. Error bars represent the mean ± s.d. of four biological replicates. The extraction data are accompanied by spots showing the yeast after 48 hours growth (bottom).

To verify the null effect of peroxisomal targeting on VioA, we again turned to the VioE-fusion sequestration assay. As illustrated in **Figure 3-20**, a

VioE-VioA fusion was expressed either in a system designed to test VioE activity by expressing VioA and VioB cytosolically, or in a VioA-monitoring strain where VioB and VioE were expressed cytosolically, with *pex5Δ* controls for each. We found that the VioE-VioA fusion lost VioE activity when targeted to the peroxisome, demonstrating efficient peroxisomal import (**Figure 3-20b,c**). In contrast, the VioA portion of the fusion remained active irrespective of peroxisomal targeting. These results provide *in vivo* evidence that only VioA remains active as part of the PDV pathway while peroxisomally sequestered, further suggesting that both VioA's substrate (tryptophan) and product (IPA imine) can cross the peroxisomal membrane efficiently, but not VioB/VioE's IPA imine dimer.

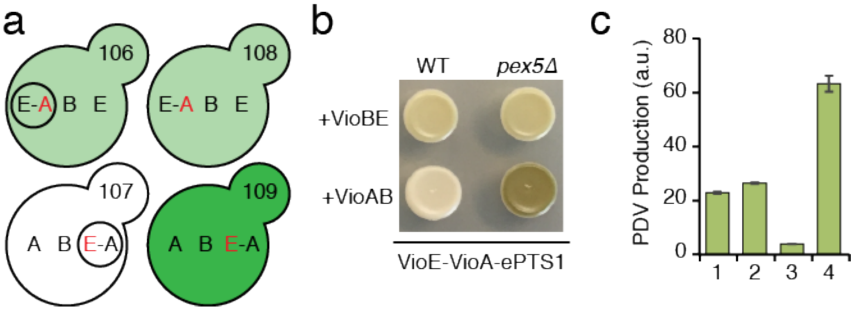


Figure 3-20. Confirmation of VioA sequestration using VioE as a fusion tag for peroxisomal import measurement. (**a**) To confirm that VioA was properly localized to the peroxisome, a fusion protein VioE-VioA-ePTS1 (listed as A-E in the diagram) was expressed at pRPL18B levels along with cytosolically localized enzymes from the PDV pathway. To test the activity of VioA in the A-E fusion, VioB and VioE were coexpressed cytosolically in a wildtype (1, upper left) or *pex5Δ* strain (2, upper right). To test the activity of VioE in the A-E fusion, VioA and VioB were coexpressed cytosolically, again in a wildtype (3, lower left) or *pex5Δ* strain (4, lower right). (**b**) Yeast spots corresponding to the strains described in (**a**) after 48 hours growth. (**c**) Strains 106-109 were grown and extracted as in (**a**) and PDV measured. Error bars represent the mean ± s.d. of four biological replicates.

3.2.5. PDV production can be compartmentalized under specific conditions

Nonspecific permeability to small intermediates poses a problem for the isolation of pathways in the peroxisome. While our enzyme sequestration assays could prove useful for identifying the source of this permeability, our initial efforts to do so have proven unsuccessful. Even so, the last two steps in the PDV pathway, VioB and VioE, share an intermediate, IPA imine dimer, which does not appear to efficiently cross the native peroxisomal membrane. IPA imine dimer can also spontaneously form the dead-end side product chromopyrrolic acid (CPA) (**Figure 3-1**). We hypothesized that, by compartmentalizing VioB and VioE in the peroxisome, we could demonstrate both enzymes remain functional in the peroxisome, and potentially even reduce CPA formation by channeling IPA imine dimer to VioE.

First, we had to account for the import of an additional enzyme by measuring how much displacement occurs when multiple heterologous cargo proteins compete for peroxisomal import. To do this, we expressed a constant amount of peroxisomally-targeted VioE in a strain that overexpressed cytosolic VioA and VioB. We then investigated how much of a competitor protein, RFP-ePTS1, must be expressed to competitively displace VioE from the peroxisomal import system (**Figure 3-21a**). With no competitor expressed, VioE was efficiently sequestered in the peroxisome, and the strain produced 95% less PDV than its *pex5Δ* control (**Figure 3-21b,c**). Expression of the competitor had negligible effects on VioE-targeting and PDV production even up to the strongest single promoter, pTDH3, where a drop from 95% to 85% VioE sequestration was observed. Only when multiple copies of pTDH3 were used to express RFP-ePTS1 was VioE-ePTS1 effectively displaced, suggesting that the peroxisomal import machinery possesses plenty of excess capacity.

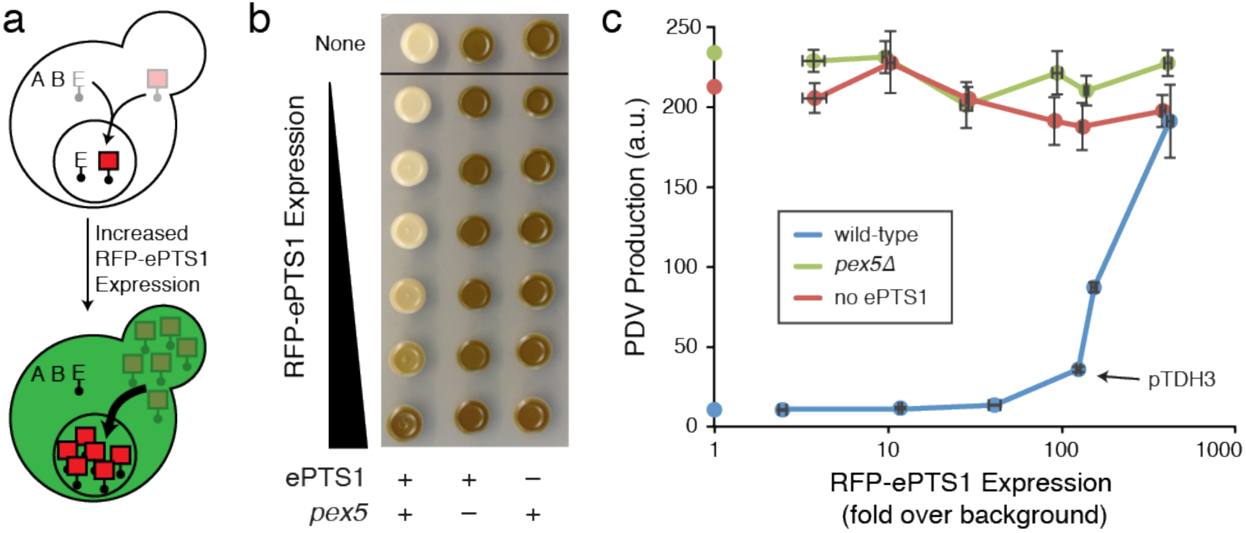


Figure 3-21. Peroxisomal import can be saturated at very high levels of cargo expression. (**a**) Starting with Strains 135, 136, and 137 coexpressing cytosolic VioA, cytosolic VioB, and a moderate fixed amount of peroxisomally-targeted YFP-VioE-ePTS1, increasing amounts of RFP-ePTS1 were coexpressed in an effort to crowd VioE out of the peroxisome. If VioE was successfully prevented from entering the peroxisome by RFP cargo crowding, then increased PDV production should result as cytosolic levels of left-behind VioE rise. (**b**) PDV spot assay of Strains 138-161 showing that PDV production does indeed rise with peroxisomal cargo crowding. Controls with defective peroxisomal import (*pex5Δ*) and untagged VioE confirm the maximal level of PDV production with cytosolic VioE remains constant. (**c**) PDV extracts from the same cells shown in (**b**) to provide a more quantitative estimate of the maximum protein

expression level before crowding becomes an issue. The highlighted expression level (arrow) indicates RFP being driven by a single copy of the strongest promoter in *S. cerevisiae,* pTDH3. Error bars are the mean ± s.d. of four biological replicates.

Having established that the peroxisomal import machinery had sufficient capacity for both proteins, we moved to testing the simultaneous compartmentalization of VioB and VioE. In a strain expressing cytosolic VioA, we expressed VioB and VioE each with an ePTS1 tag for peroxisomal compartmentalization or no tag for cytosolic targeting. We used both wildtype and *pex5Δ* cells to control for any import-independent differences in enzyme activity due to fusion of the ePTS1 tag. We were particularly interested in whether peroxisomal targeting improves PDV production when one or more enzyme partners are scarce to highlight the effects of locally concentrating enzymes and intermediates by compartmentalization. We selected two different expression regimes to test: 1) a VioB-limited regime where substrate channeling is less likely to be observed because there is plenty of VioE available to convert IPA imine dimer (pRPL18B-VioB/pRNR2-VioE) and 2) a VioE-limited regime where substrate channeling effects should be most visible (pTEF1-VioB/pREV1-VioE).

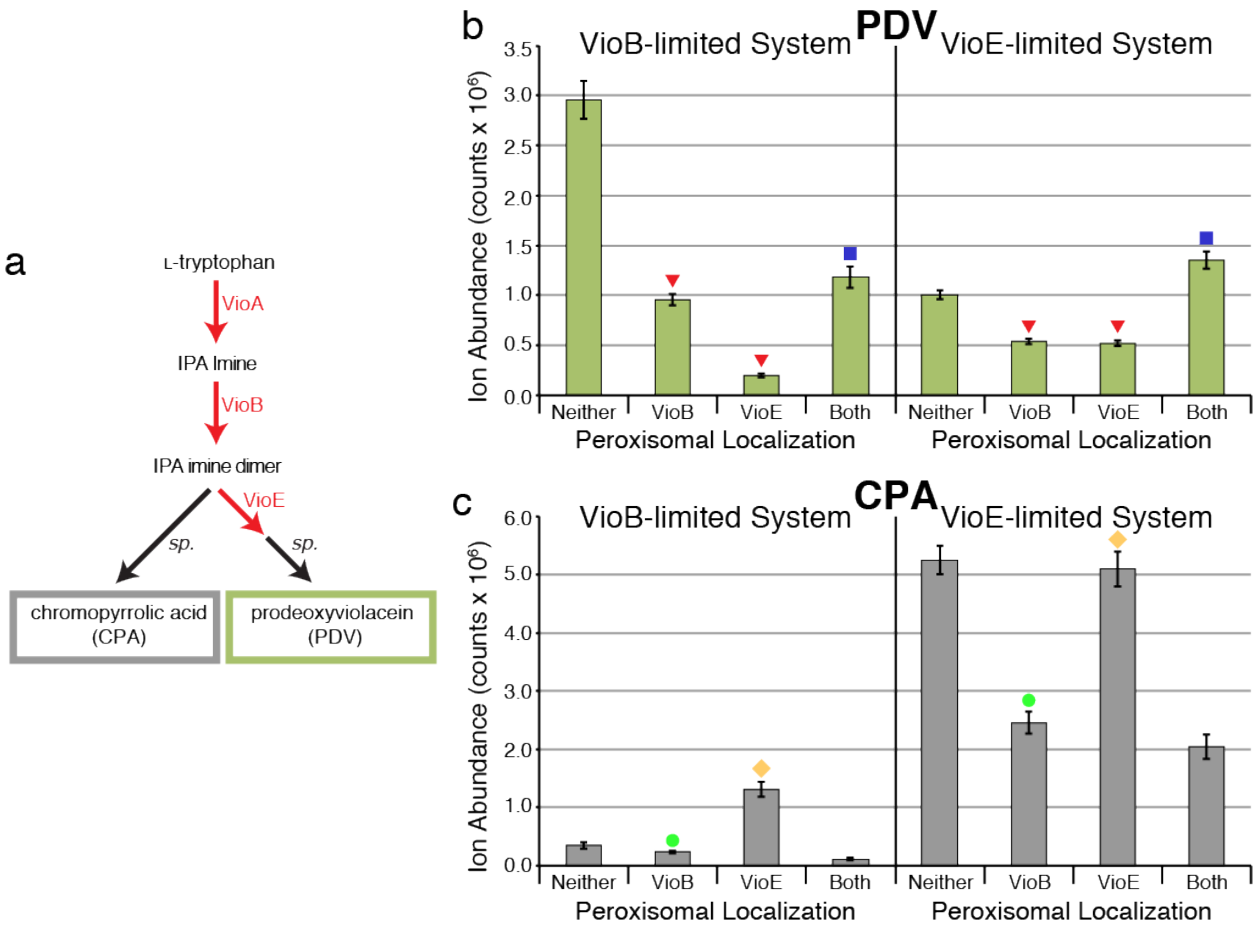


Figure 3-22. Colocalizing PDV pathway enzymes in the peroxisome reduces byproduct CPA in all cases and also increases PDV production in a VioE-limited regime. (**a**) The PDV pathway, now including chromopyrrolic acid (CPA), the colorless spontaneous byproduct that spontaneously forms from IPA imine dimer in the absence of VioE activity. (**b,c**) Bars show amounts of PDV or CPA detected in yeast expressing VioB and VioE in the cytosol or peroxisome, with error bars showing the mean ± s.d. of six biological replicates. Two expression regimes of VioB and VioE are tested: a VioB-limited regime with less VioB and more VioE, and a VioE-limited regime with more VioB and less VioE. (**b**) Upon separation of VioB and VioE into different compartments, PDV production consistently decreases (red ▼) relative to both enzymes together in the cytosol. However, when VioB and VioE are together in the peroxisome (blue ■), the change in PDV production relative to cytosolic localization depends on whether excess VioB is present, with excess VioB allowing a net increase in PDV. **(c)** CPA formation generally decreases upon VioB import into the peroxisome (green ●), but increases or stays the same upon VioE import into the peroxisome (gold ♦).

As enzymes are targeted to the peroxisome, three trends in PDV production appear: when VioB and VioE are in separate compartments, less PDV is produced; VioB loses some of its activity upon import; and when excess VioB is provided, PDV production actually increases when the pathway is moved from the cytosol to the peroxisome. In both regimes, strains with VioB and VioE directed to the same compartment made more PDV than strains in which the enzymes were split between the peroxisome and cytosol (**Figure 3-22b**), suggesting that both enzymes remained active and PDV was produced in the peroxisome. However, the effect of moving both enzymes from the cytosol to the peroxisome was quite different in the two regimes. In the VioB-limited expression regime, PDV production was 60% lower when VioB and VioE were compartmentalized in the peroxisome (**Figure 3-22b**). In contrast, in the VioE-limited regime, we observed 35% greater PDV production when VioB and VioE moved from the cytosol to the peroxisome (**Figure 3-22b**).

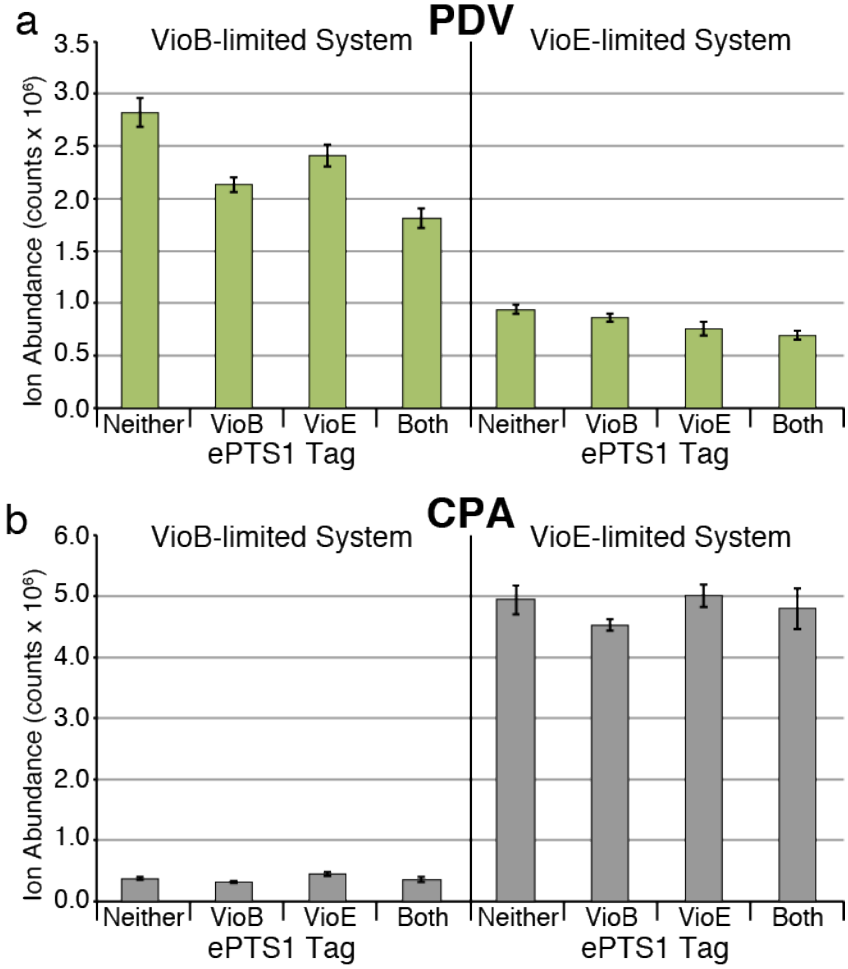


Figure 3-23. ***pex5Δ*** **import controls for Figure 3-22.** Measurement of CPA and PDV in yeast cytosolically expressing VioA while with or without ePTS1 tags on VioB and VioE in all combinations. All strains are *pex5Δ*, so no peroxisomal import is possible. Data is organized identically to **Figure 3-22**, with two different expression regimes shown: a regime where lower expression of VioB and higher expression of VioE cause VioB to be the limiting enzyme, and vice versa causing VioE to be the limiting enzyme. (**a**) Measurement of PDV extracted from cells. The ePTS1 tag seems to lead to decreased

activity from the enzyme to which it is fused. (**b**) Measurement of CPA extracted from cells. Again, the ePTS1 tag seems to lead to a slight decrease in activity for each enzyme. Error bars are the mean ± s.d. of six biological replicates.

To examine the disparity between these regimes, we measured accumulation of the off-pathway product chromopyrrolic acid (CPA), which forms spontaneously from the product of VioB in the absence of VioE (**Figure 3-22a**). CPA should accumulate when VioB's product cannot access VioE, such as when VioB and VioE are split between the peroxisome and cytosol. Indeed, we observed the expected increase in CPA when VioE was sequestered in the peroxisome in the VioB-limited regime (**Figure 3-22c**). The VioE-limited regime showed little change in CPA, likely because of reduced pathway draw from VioE (**Figure 3-22c**). However, peroxisomally sequestered VioB generated 33% or 53% less CPA in the VioB- and VioE-limited regimes, respectively. The consistent decrease in the CPA product from VioB upon peroxisomal localization suggests that VioB has reduced activity when imported into the peroxisome. There are several explanations for this behavior, including an unfavorable lumenal environment due to pH[127,128] or other factors, or that VioB itself is being imported in a nonfunctional state. It is possible that ePTS1-based import outpaces heme loading of VioB, as peroxisomal import can occur prior to cargo folding[129]. Furthermore, it is thought that free heme (617 Da) is unavailable in the peroxisome[130], so apo-VioB would be unable to acquire heme after import.

Even with the apparent decline in the specific activity of VioB in the peroxisome, if excess VioB is supplied as in the VioE-limited regime, PDV production increases by 35% upon import of both enzymes (**Figure 3-22b**). We suspect this increase is not due to protection of VioE from degradation, as a slight decrease in YFP fluorescence was observed when VioE-YFP was successfully imported into the peroxisome (**Figure 3-7b**). Instead, we hypothesize that either the lumenal environment is favorable for VioE activity or substrate channeling is occurring. Substrate channeling could be created by colocalizing the source (VioB) and sink (VioE) of IPA imine dimer, thus reducing the accumulation of IPA imine dimer and reducing the formation of CPA byproduct. Under this specific expression regime, we found that wasted flux can be decreased and desirable production increased by colocalization of VioB and VioE in the peroxisome.

3.3. Discussion

Engineering pathway compartmentalization requires precise control over both protein and metabolite localization, but the tools for measuring and manipulating these traits in peroxisomes are lacking. We report the development of novel *in vivo* assays for analyzing protein and metabolite import into the peroxisome as well as the identification of an efficient, modular targeting tag to direct heterologous cargo to the yeast peroxisome. These tools offered new insights into basic peroxisome biology and allowed us to establish key design principles for repurposing the peroxisome to compartmentalize engineered metabolic pathways.

A fundamental requirement of a synthetic organelle is the ability to import protein effectively and minimize cytosolic activity of cargo proteins prior to import. We find the yeast peroxisome to be highly efficient in this regard, with the native peroxisomal import machinery able to handle import of cargo driven by multiple copies of the high-strength promoter, pTDH3. As a single copy of pTDH3 can drive production of approximately 4% of the cell's total protein[131], import capacity should not be a problem for most protein-based applications. Even so, further enhancements should be achievable by overexpressing peroxisome biogenesis genes, removing endogenous peroxisomal cargo, or using hosts with naturally voluminous peroxisomes, such as *Pichia pastoris*[98]. Speed of protein import is also satisfactory, as we found that an enhanced PTS1 (ePTS1) tag bearing a linker with several basic residues was able to effect rapid peroxisomal import for a wide variety of proteins. Cargo proteins were compartmentalized at rates that outpaced TEV protease in a competition assay, arrested growth in essential protein sequestration assays, and cleared the cytosol of fluorescent protein within an hour of induction. While the ePTS1 tag appears to be modular, we have developed a novel assay that can aid in high-throughput measurement and optimization of an individual protein's import efficiency. By fusing the enzyme VioE to a protein of interest, the degree of sequestration of the fusion in the peroxisome can be determined by loss of VioE's product, the pigment PDV. With the native protein import machinery capacious, rapid, and versatile, and several tools available to help make further improvements, we moved on to the next challenge: permeability.

A second fundamental requirement of a synthetic organelle is control over metabolite transport in order to import substrates, contain intermediates, and exclude cross-reactive metabolites. Our initial efforts to sequester heterologous enzymes suggested that the peroxisomal membrane allows free diffusion of some, but not all small hydrophilic metabolites consistent with the previously hypothesized peroxisomal aqueous pore with a molecular weight cutoff of 300-400 Da[111]. To get a better estimate of pore size, we developed a BGL-based assay to generate the first *in vivo* polymer exclusion measurements for peroxisomal permeability, estimating the pore radius between 0.57-0.65 nm. This value was consistent with *in vitro* polymer exclusion measurements[114] and suggests the 2-4 nm permeability states estimated using reconstituted peroxisomal protein and ion conductance[104] are unlikely to be encountered in metabolic engineering applications. Though our estimate will be predictive for stable molecules, reactivity and molecular lifetime also play a role, as the near-diffusion-limited enzyme catalase shows latency even though its substrate, hydrogen peroxide (34 Da) is thought to easily cross the peroxisomal membrane[112]. Similar effects might explain our ability to compartmentalize the reactive intermediate of the PDV pathway, IPA imine dimer. For pathways with stable intermediates, however, the current threshold of peroxisomal permeability poses a problem. Engineered pathways frequently feature small intermediates that can be compartmentalized only if peroxisomal membrane permeability is reduced. We expect this permeability for hydrophilic molecules is due to one or more membrane proteins rather than the lipid membrane itself. Pxmp2 has been identified as a channel-forming peroxisomal membrane protein[132] in mammalian cells, but no equivalent channel has been confirmed in yeast. It is also possible that the protein importomer, which opens to a reported radius of 4.5 nm when importing large cargo[104], could be the culprit. Testing this hypothesis will likely require inactivation of the importomer and development of an alternative route for directing cargo to the peroxisomal lumen, a considerable challenge. Thus, peroxisomal permeability remains an open challenge that places hard restrictions on which pathways can be effectively sequestered.

With these protein import tools and permeability rules, we were able to produce the first rigorous demonstration of a heterologous pathway compartmentalized in the peroxisome with a trapped intermediate. The model pathway consisted of VioB and VioE, which share a peroxisome-impermeant intermediate, IPA imine dimer, generated from a permeant substrate, IPA imine. Both VioB and VioE were active in the peroxisomal lumen, but VioB appeared to halve in activity upon import while VioE's activity increased. With excess VioB, we found that peroxisomal import led to a 35% increase in production of PDV and a 61% reduction in the off-pathway byproduct CPA. VioB's loss of activity highlights the need for further modification of the peroxisome's lumenal environment to maintain or enhance the activity of individual enzymes, especially cofactor-dependent enzymes such as VioB. On the other hand, VioE's gain in activity highlights the potential for the peroxisome to enhance production, potentially by concentrating scarce enzymes and intermediates to achieve substrate channeling. Together, these results provide a glimpse into the rewards a synthetic organelle can offer and the future cost of constructing that organelle.

With its rapid, versatile import system and disposable native function, the peroxisome holds great promise and is already suitable for protein import-only applications, but several challenges stand in the way of unlocking its full potential. Reductions in small molecule permeability and accommodations to maintain enzyme activity will be important next steps, with an eye towards the final goal of modifying the chemical environment of the peroxisomal lumen (e.g. pH, redox state, cofactor balance) to allow catalysis of reactions that would be infeasible in the cytosol. Achievement of these long-term goals could one day provide unprecedented control over the chemistry that can be performed in the cell.

3.4. Extensions and Future Directions

3.4.1. Engineering glucoside transport into the peroxisome

As **Figure 3-14** showed, X-cellotrioside does not enter the peroxisome, while X-glucoside and X-cellobioside do. However, all of these X-glycosides enter the cell via the cellodextrin transporter CDT1 (**Figure 3-16**). Though CDT1 naturally arrives at the plasma membrane, appending a peroxisomal membrane protein could redirect CDT1 to the peroxisomal membrane. Indeed, Pex15p fragments, specifically the C-terminal fragment (amino acids 315-383), have successfully redirected fusion proteins to the peroxisome[133]. Thus, we fused CDT1 (or CDT2, a protein that can transport X-cellobioside only, but has no dependence on proton gradient) to an RFP (mKate2) for tracking and Pex15-C for targeting to make CDT1-RFP-Pex15C and CDT2-RFP-Pex15C. The fusion proteins appear to transport to the peroxisome with little background elsewhere in the cell (**Figure 3-24a**), so CDT1 and CDT2 were tested for their ability to reduce X-cellobioside latency or transport X-cellotrioside, with a dummy transporter (LAC12 from *K. lactis*) serving as the negative control.

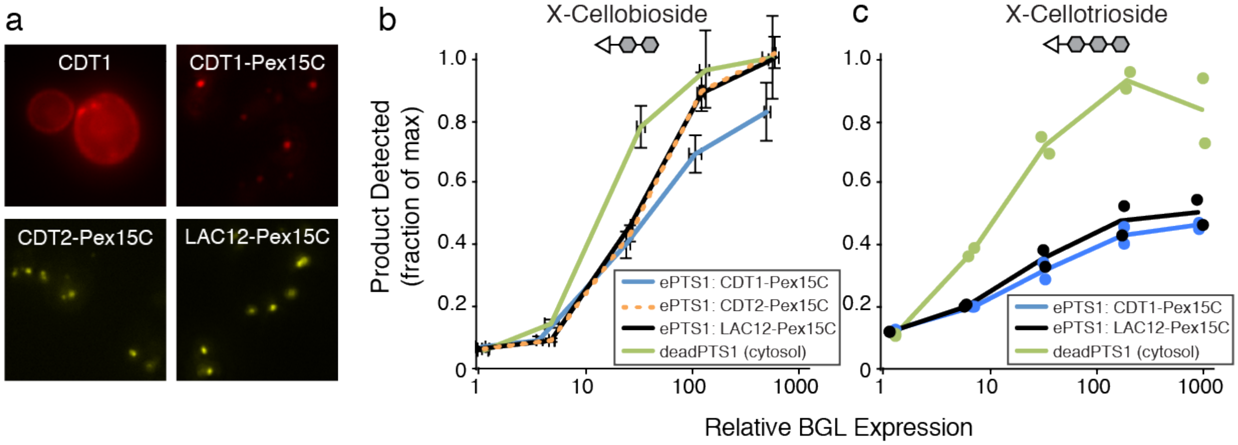


Figure 3-24. Cellodextrin transporters can be targeted to the peroxisome but appear to be inactive. (**a**) Fluorescence microscopy of mKate-tagged (CDT1 and CDT1-Pex15C) or Venus-tagged (CDT2-Pex15C and LAC12-Pex15C) transporters. Punctate fluorescence suggests peroxisomal localization. (**b**) Relative fluorescent product release from X-cellobioside in yeast strains expressing increasing amounts of peroxisomal (blue, orange, and black lines) or cytosolic (green lines) β-glucosidase. Different peroxisomally-tagged transporters are tested: CDT1-Pex15C (blue line), CDT2-Pex15C (orange dotted line), and a negative control, LAC12-Pex15C (black line). Error bars are the mean ± s.d. of six biological replicates. (**c**) X-cellotrioside was added to yeast strains expressing peroxisomal (blue, black) or cytosolic (green) β-glucosidase. Transporters tested are CDT1-Pex15C (blue) and a negative control, LAC12-Pex15C (black). Only two replicates per data point were conducted; these are indicated with dots.

Unfortunately, we did not observe any appreciable difference in X-cellobioside latency (**Figure 3-24b**) or X-cellotrioside transport (**Figure 3-24c**) between CDT1-RFP-Pex15C, CDT2-RFP-Pex15C, and the negative control, LAC12-RFP-Pex15C. There are several possible explanations for this failure. Firstly, CDT1 is a proton symporter and requires a proton gradient for activity; without the gradient, only a third of the activity remains[134], leaving CDT1 with roughly the same activity as CDT2. The pH of the peroxisomal lumen remains a matter of debate[127,128], and an opposing gradient could actually inhibit transport. Secondly, there could simply be insufficient activity on the part of CDT1. Only pRPL18B levels of CDT1 could be expressed to avoid disrupting peroxisomal morphology. Under optimal conditions at the plasma membrane, pRPL18B levels of CDT1 were able to transport approximately 15% of maximum signal (CDT2 was able to transport 5%). Considering the signal difference between cytosolic and peroxisomal BGL is 40% and the error is 5%, CDT1 was unlikely to be able to fully close the gap, and CDT2 may not have been detectable above error.

There exist mutations to improve CDT1 activity, N209S/F262Y[135] and C82S[136], as well as CDT2 activity (Q207H/N311H/I505T[137]), and these mutations would be recommended in future formulations of this experiment. The control would be most ideally an inactive CDT1 mutant such as D307A[138], E476A[138], E194A[139], or R174A[139]. Finally, it is possible that membrane proteins directed to the peroxisome by the PEX15C tag are simply failing to fold properly or integrate into the membrane. This challenge would have to be solved by finding different tags, such as ALDP fragments[140] or fragments of other peroxisomal proteins[141]. In any case, the ability to direct functional protein to the peroxisomal membrane remains an open challenge with significant implications for controlling peroxisomal metabolism.

3.4.2. Maintaining activity of imported VioB

Much to our dismay, we observed that VioB lost roughly half of its activity upon peroxisomal import (**Figure 3-22c**). Though we have several hypotheses, including a different, unfavorable chemical environment in the peroxisomal lumen; ePTS1-based import could outpace folding or heme loading of VioB[129], leaving apo-VioB in the peroxisome without heme available[130]; or the import process itself could cause VioB to misfold or aggregate. The too-fast-import hypothesis was particularly easy to test, as we had several lower-efficiency (hypothetically slower) tags from our library (**Figure 3-6**). We selected LTKVMT (87% efficiency versus 90% for ePTS1), EACRSY (79%), SSVYEN (49%), and AGDLEY (26%).

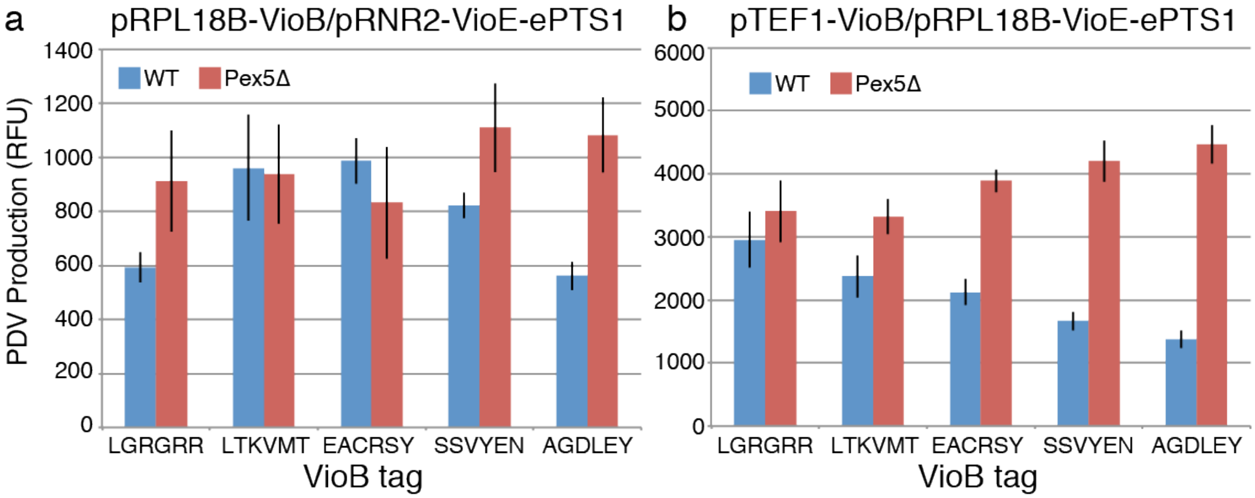


Figure 3-25. Slower import of VioB may rescue activity in certain contexts. VioE was tagged with ePTS1 while VioB was tagged with a peroxisomal targeting sequence containing one of five possible linkers: LGRGRR (ePTS1), LTKVMT, EACRSY, SSVYEN, and AGDLEY, in order of decreasing efficiency. Constructs were tested in both wildtype (blue) and peroxisomal import deficient (red) strains. (**a**) A heavily VioB-limited regime, with pRPL18B levels of VioB and pRNR2 levels of VioE. These are the same conditions under which **Figure 3-22b** was generated for the "VioB-limited regime." (**b**) An intermediate regime where both VioB and VioE may be limiting, with pTEF1 levels of VioB and pRPL18B levels of VioE. Similar to the "VioE-limited regime" in **Figure 3-22b** except VioE expression has been increased from pREV1 to pRPL18B, a roughly tenfold increase. Error bars are the mean ± s.d. of eight biological replicates.

As **Figure 3-25a** shows, "slower" tags such as LTKVMT and EACRSY can resolve whatever folding or heme loading issues may be occurring due to too-fast import in the same heavily VioB-limited regime we used in **Figure 3-22b**, negating the decrease in productivity seen with peroxisomal import. However, the improvement disappears at higher expression levels (**Figure 3-25b**), such as the intermediate regime where both VioB and VioE are expressed roughly fourfold more than in **Figure 3-25a**. Whether this is due to moving into a regime of excess VioB (where a loss of some VioB activity doesn't matter) or subtle slowing of peroxisomal import with high levels of VioB that makes all of the tags appear less efficient (i.e. LGRGRR in this condition is as fast as LTKVMT in F**igure 3-25a**) is unclear.

To narrow down whether heme loading played a role in the slow tags' function, we also tested increases in cytosolic heme. A heme-specific plasma membrane permease, HRG4 from *C. elegans*, is known to be necessary and sufficient for heme transport in yeast[142], so we tested this in conjunction with PDV production. Specifically, we used strains with high levels of HRG4 and cytosolic VioA (pCCW12, pTDH3, respectively) along with either high levels of cytosolic VioB and VioE (pTDH3 for both) or moderate levels of peroxisomally-targeted ePTS1-tagged VioB and VioE (pTEF1, pRPL18B respectively). These strains were tested in SD-LU media (adjusted to pH 8 to maintain heme in solution) with and without 25 μM of added hemin (which converts to heme in solution). We did not observe a significant difference in PDV production with either cytosolic or peroxisomally targeted enzymes (**Figure 3-26**), suggesting that PDV production (and by extension VioB's activity) is not significantly limited by heme availability.

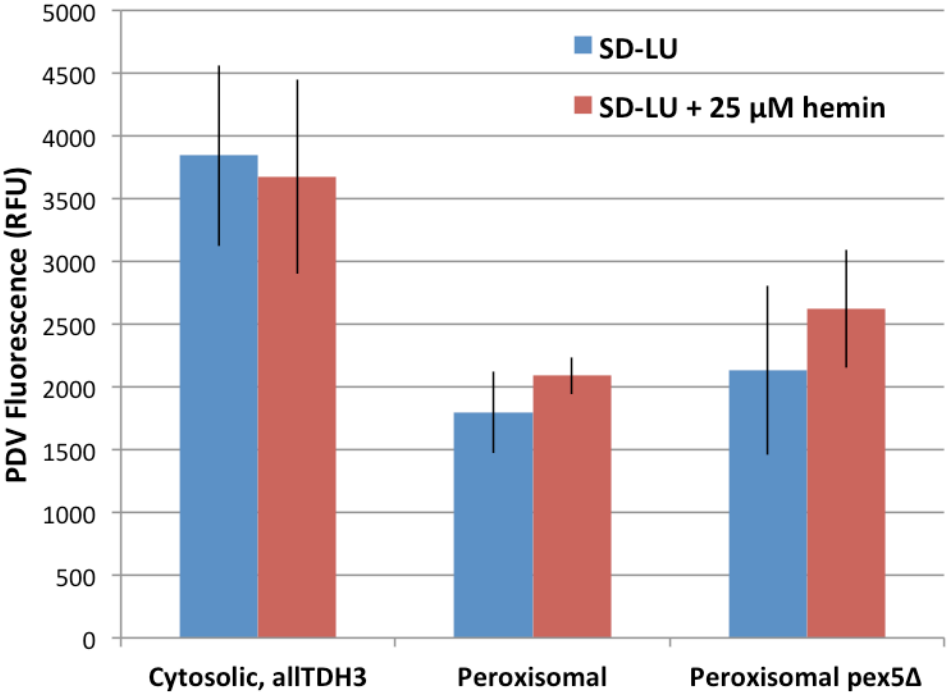


Figure 3-26. Added heme does not significantly affect PDV production. Yeast expressing a heme permease (HRG4) and the PDV pathway were tested with and without 25 μM of added hemin, the cofactor for VioB. In the cytosolic case, VioB and VioE were both expressed at TDH3 levels with no targeting tags. For the peroxisomal case, VioB and VioE were tagged with ePTS1 and expressed at pTEF1 and pRPL18B levels, respectively. The peroxisomal strains were made in a wildtype or *pex5Δ* peroxisomal import knockout. Error bars are the mean ± s.d. of six biological replicates.

Taken together, this evidence suggests that "slow" tags may be an adequate solution to address the loss of activity upon import of sensitive proteins like VioB. However, there is the significant caveat that slow tags don't appear to unlock new heights of peroxisomal PDV production in regimes with higher expression of both VioB and VioE. Indeed, it seems as if the apparent substrate channeling advantage of peroxisomal compartmentalization of VioB and VioE can be observed only if VioE is heavily limited and CPA formation is consuming lots of IPA imine dimer.

3.4.3. Investigating the identity of the aqueous pore

While our manuscript was under consideration, another manuscript was released claiming that Pex11p[143] was a peroxisomal channel protein. Of course, we rushed to test this with both our PDV assay (by sequestering VioA to see if IPA imine and/or tryptophan were now impermeant) and our β-glucosidase assay (to see if X-cellobioside was now impermeant). Unfortunately, neither the PDV assay (data not shown) nor the BGL assay (**Figure 3-27**) showed any effect with the *pex11Δ* knockout.

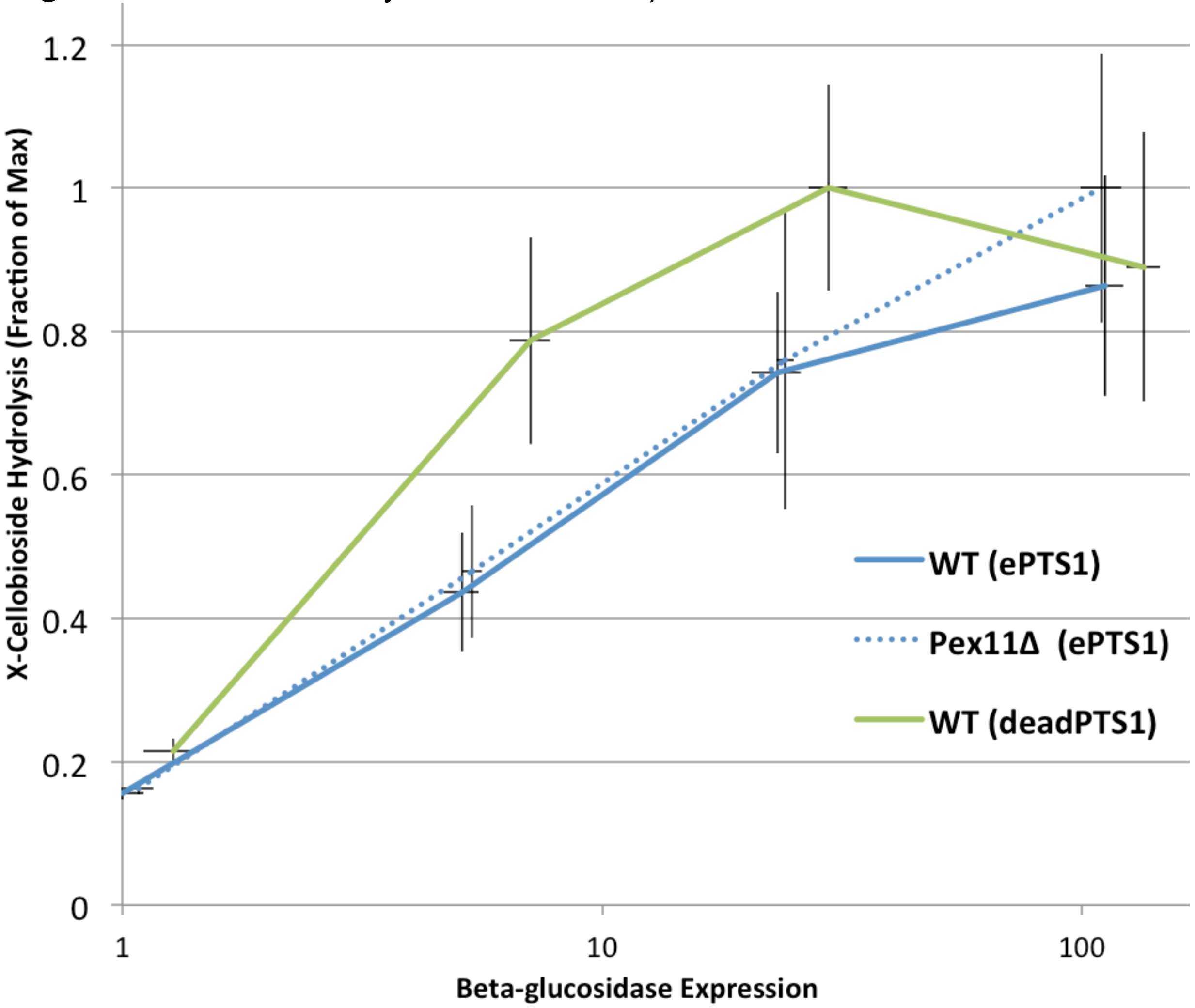


Figure 3-27. *Pex11Δ* does not affect transport of X-cellobioside into the peroxisome. X-cellobioside was fed to cells expressing CDT1 at the plasma membrane and a β-glucosidase (BGL) in the peroxisome (ePTS1) or the cytosol (deadPTS1) in wildtype or *pex11Δ* backgrounds. If *pex11Δ* had successfully abolished X-cellobioside permeability, the *pex11Δ* curve (blue dotted line) would be lower and flatter than the WT curve (blue solid line). Error bars are the mean ± s.d. of fourteen biological replicates.

Our own suspicion is that the peroxisomal protein import pore (including Pex14p, Pex5p, and companion proteins) is actually responsible for the leakiness. This hypothesis is difficult to test, as a knockout of PEX5 or PEX14 will make it difficult to get the assay protein (VioA or BGL) into the peroxisome. One potential option is to stop all peroxisomal cargo synthesis, because, without cargo, Pex5p and Pex14p will not interact and form a pore[104]. We selected cycloheximide, a ribosomal inhibitor, for this purpose. Though we added cycloheximide up to a final concentration 50 μg/mL, we saw no effect on peroxisomal permeability (**Figure 3-28**). This means we will need to devise new and exciting ways to get protein into the peroxisome to test our hypothesis with *pex14Δ* knockouts instead.

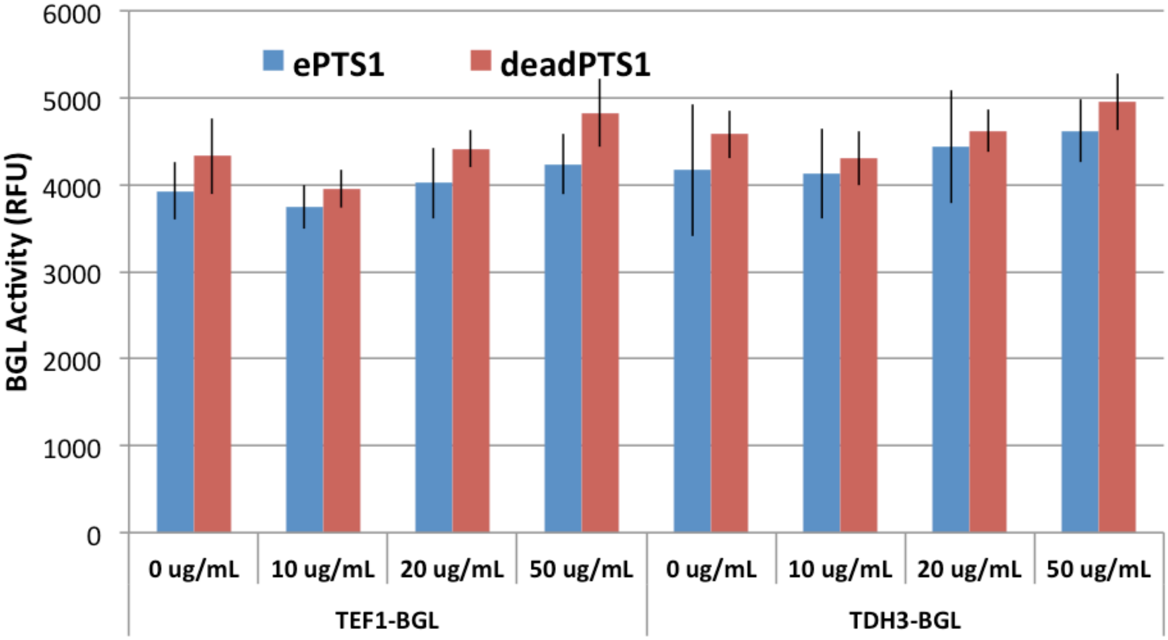


Figure 3-28. Cycloheximide does not appear to affect peroxisomal permeability to X-cellobioside. X-cellobioside was fed to cells expressing CDT1 at the plasma membrane and a β-glucosidase (BGL) at pTEF1 or pTDH3 levels in the peroxisome (ePTS1, blue bars) or the cytosol (deadPTS1, red bars). Cycloheximide concentrations are shown in μg/mL underneath the bars. If cycloheximide had successfully abolished X-cellobioside permeability, the blue bars would be significantly lower than the red bars as cycloheximide concentration increases. Error bars are the mean ± s.d. of three replicates.

3.4.4. Repurposing the peroxisome for protein purification

The unmodified peroxisome already has one clear and unambiguous superpower: it can sequester lots of protein from the cytosol (**Figure 3-21**). With this special ability, peroxisomes can separate and purify a single important protein in high yields from the rest of the cell, at higher yield and without the need for per-protein engineering or risk of misfolding offered by the secretory pathway[144–146]. There are two options for obtaining the purified protein: either collect the purified protein by separating and lysing peroxisomes, or turn peroxisomes themselves into high-yield secretory vesicles and collect the protein from the media.

While easy protocols exist for lysing cells while keeping peroxisomes intact, the separation of those peroxisomes from other components such as endoplasmic reticulum and mitochondria frequently requires ultracentrifugation on a density gradient [147–149]. There have been several exciting achievements in simpler organelle pulldowns using protein affinity. The most notable of these is a group that extracted peroxisomes from the post-nuclear fraction of rat cell lysate using beads coated with an anti-PMP70 antibody[150]. Other groups have used affinity purification to clean up a density centrifugation peroxisomal preparation; one expressed a Pex30-HA fusion and used anti-HA-coated beads[151], another expressed GFP-Pex13 and used anti-GFP-coated beads[152]. Mitochondria have been purified using only magnetic beads coated in an anti-CYB5B antibody, which was actually more effective in obtaining pure, intact mitochondria than centrifugation[153]. Another group automated their mitochondrial process, using a similar purification with anti-TOM22[154]. However, there may be room for improvement – for example, the anti-PMP70 purification left behind significant numbers of peroxisomes[150]. Our ideal design uses a peroxisomal protein, such as Ant1p or Pex15p, fused to a reversibly-binding affinity tag (i.e. Strep-tag, which can be picked up with Strep-Tactin beads and released using desthiobiotin[155]). Furthermore, engineering TEV protease sites into Inp1p and Pex30p, Inp2p, and Pex11p, the main tethers between the peroxisome, the ER, cytoskeleton, and mitochondria[156], would allow addition of TEV protease to free the peroxisomes from these components for higher protein purity. Unfortunately, gentle reader, I never got around to making these constructs, so this idea is left as an exercise to the reader. Even if perfected, protein purification via lysis of cells and capture of peroxisomes is still a complicated affair, and secretion is preferable because it leaves cells alive.

Turning peroxisomes into secretory vesicles, dubbed "peroxicretion," has already been reported in the literature[157]. The core idea of this work was combining the cytosolic fragment of a v-SNARE protein, which fuses secretory vesicles to the plasma membrane, to a peroxisomal membrane protein. However, these initial experiments, in *Aspergillus niger*, did not show substantial peroxicretion and it was ultimately ambiguous as to whether the secretory peroxisomes worked as intended. We attempted to recreate this work in *S. cerevisiae* using a fusion of the cytosolic parts of yeast v-SNAREs Snc1p (amino acids 1-93) or Snc2p (amino acids 1-92) with either Pex15p (data not shown) or Ant1p (**Figure 3-29**). We observed colocalization of SNC1$_{1\text{-}93}$-ANT1-CFP and SNC2$_{1\text{-}92}$-ANT1-CFP with YFP-PTS1, indicating successful direction of the v-SNARE fusion to peroxisomes. We also observed strange empty CFP-only loops, which may suggest plasma membrane fusion of our fused SNC1$_{1\text{-}93}$-ANT1-CFP and SNC2$_{1\text{-}92}$-ANT1-CFP proteins (**Figure 3-29**). Similar-looking loops were also present with ANT1-CFP, though those loops always contained YFP, suggesting they were greatly expanded peroxisomes. However, we did not detect any release of fluorescent cargo into the media by microtiter plate fluorescence (data not shown). Based on these highly equivocal results, we decided to completely rework our strategy.

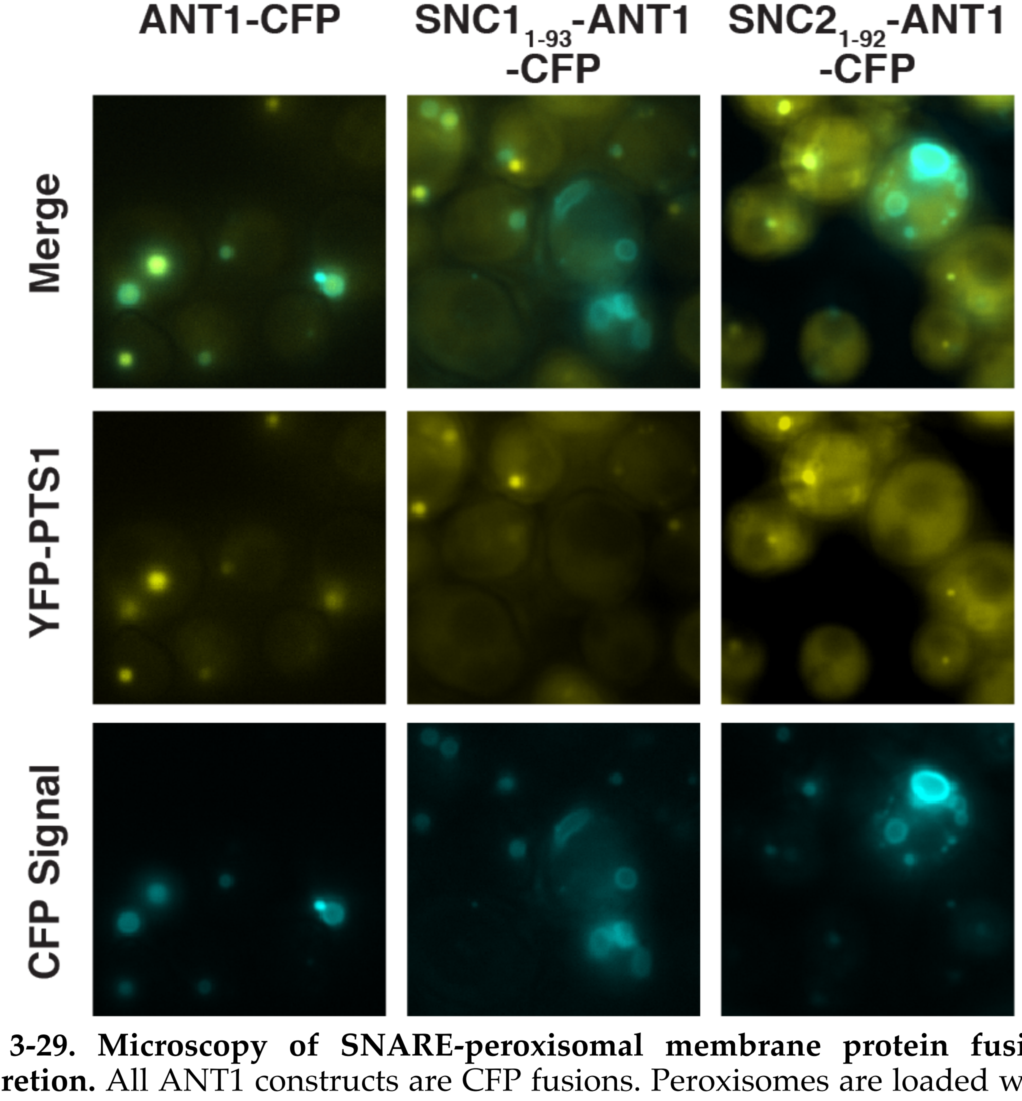


Figure 3-29. Microscopy of SNARE-peroxisomal membrane protein fusions for peroxicretion. All ANT1 constructs are CFP fusions. Peroxisomes are loaded with YFP-PTS1 as cargo. The peroxisomes seen here are unusually distended, and appear ringed by ANT1-CFP.

There were several key weaknesses in our original strategy. Foremost, we lacked a sensitive, specific readout of peroxicretion. One of the best solutions is an enzyme-linked assay, especially hydrolyases with colorimetric/fluorimetric assays. Perhaps the best option is NanoLuc, a small, monomeric, secreted luciferase[158]. Multimeric enzymes that could work are β-glucuronidase (gus[159] or GUSPLUS[160]) and invertase Suc2p[161,162], though the glycosylation acquired during secretion appears to be important for Suc2p's stability[163,164]. Other options include Western blot detection of FLAG, HA, or hexahistidine peptides, or use of superfolder GFP, which has been used in the periplasmic space[165,166]. The next component is the SNARE-peroxisomal membrane protein fusion. Since the transmembrane domains of neuronal SNARES are dispensable[167,168] (and behave similarly to yeast SNARES[169]), our initial designs using $SNC1_{1\text{-}93}$ and $SNC2_{1\text{-}92}$ were reasonable. It may be worthwhile to try a library of truncations of the cytosolic part of Ant1p or any other fusion partner, as the rigidity of the SNARE may be important. A good assay and a well-designed SNARE-peroxisomal protein fusion are key, but there are also additional improvements to the system that may prove important.

The first of these additional strategies is hijacking of vesicular tethering. The native yeast secretory vesicle does not operate on v-SNAREs alone – the v-SNAREs are assisted by the exocyst, a complex of proteins that brings the vesicle close enough to the membrane for SNARE interactions to occur[170]. Ypt31p or Ypt32p are membrane-associated proteins that recruit Myo2p and start the cascade that assembles the exocyst by recruiting Sec2p, which activates Sec4p, which in turn recruits the rest of the machinery[171]. Minimally, peroxisomally-tethered Sec4p with Sec2p or Ypt31p/Ypt32p should be sufficient to enhance peroxicretion. As usual, this would be achieved by a peroxisomal membrane protein fusion. The other consideration is preexisting peroxisomal tethering. The peroxisomes naturally have four main tether proteins, Inp1p, Pex30p, Inp2p, and Pex11p, which tie peroxisomes to the ER, cytoskeleton, and mitochondria[156]. Inp1p and Inp2p together are important for peroxisomal inheritance, with Inp1p responsible for keeping a subpopulation of peroxisomes in the mother cell, while Inp2p interacts with Myo2p to bring peroxisomes to the daughter cell[172]. Since the peroxicreting yeast are expected to be in stationary phase, Inp2p can be knocked out as there should be no fresh daughter cells and the older cells can regenerate their supply of peroxisomes *de novo*[173]. Similarly, a Pex30p knockout accelerates peroxisome proliferation in *S. cerevisiae* and is therefore desirable[174]. Pex11p knockouts have not affected peroxisomal import in our hands (**Figure 3-**

27), so Pex11p can be dispensed with as well. The remaining tether, Inp1p, can be knocked out or repurposed to our advantage – a group has fused Tom70p, a mitochondrial protein, to Inp1p, and redirected peroxisomes to the mitochondria[175]. Thus, Inp1p could serve as a scaffold for organizing the exocyst proteins mentioned earlier. Taken together, these four key points – assay development, SNARE engineering, vesicular transport hijacking, and peroxisomal untethering – form the roadmap (summarized in **Figure 3-30**) for developing peroxicretion as a high-yield system.

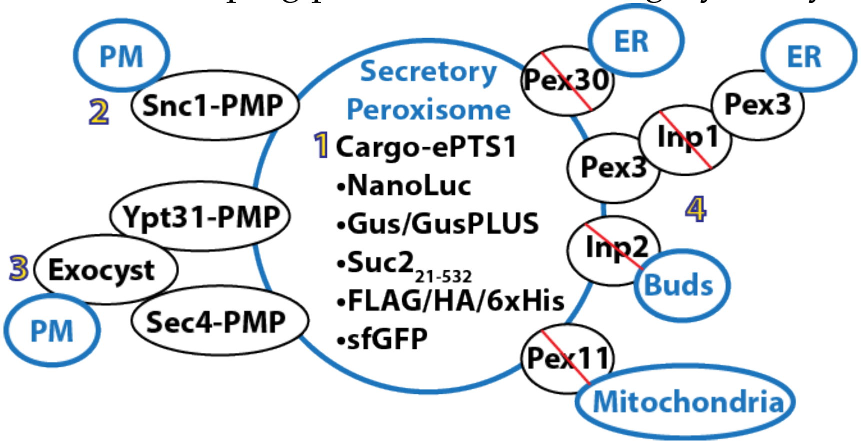


Figure 3-30. A roadmap for peroxicretion. Proteins are marked in black, while destinations are blue. Key requirements are numbered. (**1**) **Assay development** – a robust, sensitive, and specific assay for release of protein must be developed. Options include luciferase readout (NanoLuc), glycoside hydrolysis (Gus, GusPLUS, Suc2$_{21\text{-}532}$), Western blot (FLAG, HA, 6xHis), or fluorescence (sfGFP). (**2**) **SNARE engineering** – a library of different ways to get v-SNAREs to the peroxisomal membrane should be tested, including different peroxisomal membrane proteins (PMPs) such as Ant1p or PEX15C and different strategies (leucine zippers or other peptides to scaffold with PMPs). Snc1 fragments should be functional when fused to PMPs. (**3**) **Vesicular transport hijacking** – Ypt31p and Sec4p should be fused to PMPs and expressed to recruit the exocyst complex, which should help guide the peroxisome to the plasma membrane and encourage v-SNARE interaction. (**4**) **Peroxisomal untethering** –tethers that may prevent the peroxisomes from moving to the plasma membrane should be knocked out. These are Inp1p (which may be repurposed), Inp2p, Pex11p, and Pex30p.

3.4.5. Synergy with benzylisoquinoline alkaloids.

Please refer to **Chapter 2** for how the peroxisome could be used to isolate a cytochrome P450 (CYP) and companion reductase (CPR) from host cell CPR interference.

3.4.6. Synergy with biological protecting groups

Although the peroxisome permits fairly large molecules to cross in and out, there are some key metabolites that don't cross without the help of transporters. Among them are acyl-CoA, ATP (and likely PAPS), and possibly UDP-sugars. These large metabolites are donors for biological protecting groups of all shapes and sizes, and all three classes (acyl-CoA, ATP/PAPS, and UDP-sugars) are naturally used in yeast cell metabolism. As a consequence, these donor molecules are good candidates for sequestration to establish separate pools for engineered pathways, as yeast can naturally use these donor molecules. This consumption can cause toxicity if the donor molecules are carrying modified groups not normally seen in the cell (e.g. UDP-xylose when normally only UDP-glucose is observed), or depletion of a pool due to unregulated host cell usage (i.e. the cell consumes excess UDP-glucose because it is available). Furthermore, these biological protecting groups (especially hydrophilic options like glucuronic acid and sulfate) can be used to trap molecules in the peroxisome by increasing their hydrodynamic radius. In summary, peroxisomes can be used to store pools of donor molecules for adding biological protecting groups to metabolites, and biological protecting groups can be used to trap metabolites in peroxisomes. See **Chapter 4** for details on how biological protecting groups can change the properties (including size and stability) of valuable molecules.

3.5. Materials and Methods

3.5.1. Strains and growth media

The base *S. cerevisiae* strain for all experiments in this article was BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*). Base strain BY4741 and BY4741 *pex5Δ* (Strain 1) were both ordered from Open Biosystems – GE Dharmacon (Lafayette, CO, USA). Wildtype yeast cultures were grown in YPD (10 g l^{-1} Bacto Yeast Extract; 20 g l^{-1} Bacto Peptone; 20 g l^{-1} Dextrose). Selection of auxotrophic markers (URA3 or LEU2) was performed in synthetic complete media (6.7 g/l Difco Yeast Nitrogen Base without amino acids (Spectrum Chemical); 2 g l^{-1} Drop-out Mix Synthetic Minus appropriate amino acids, without Yeast Nitrogen Base (US Biological); 20 g l^{-1} Dextrose). All strains constructed for this work are listed in **Table 3-1**.

Golden gate assembly reactions were transformed in chemically competent *E. coli* prepared from strain TG1 (Lucigen). Linker libraries were transformed in TransforMax EPI300 (Epicentre) electrocompetent *E. coli*. Transformed cells were selected on Lysogeny Broth (LB) containing the antibiotics ampicillin or kanamycin. Sequences of plasmids produced for this work are supplied on request.

Table 3-1. List of yeast strains used in this work. Strains were made by chromosomal integration of plasmids directly into the parent strain, except when a repair partner is noted as in Strain 54 and Strains 56-71. These strains were cotransformed with the plasmid and repair DNA listed, plated, and cured of the Cas9 plasmid once CRISPR genomic modification was confirmed. Also note that all references to the red fluorescent protein mKate2 are abbreviated as RFP, Venus is referred to as YFP, and mTurquoise2 is listed as CFP.

#	Strain Name	Strain Parent	Plasmid Name	Description	Yeast Marker
1	YJD051	BY4741	N/A	Openbiosystems *pex5Δ* :: KanMX (KO of YDR244W) Clone ID 3603 Catalog # YSC6273-201934299	Kan
2	yWCD321	BY4741	pWCD1443	pTEF1-VioA,pTDH3-VioB	LEU2
3	yWCD325	YJD051	pWCD1443	pTEF1-VioA,pTDH3-VioB	LEU2, Kan
4	yWCD788	yWCD321	pWCD1134	URA3 Empty Vector	URA3, LEU2
5	yWCD791	yWCD321	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2
6	yWCD797	yWCD325	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2, Kan
7	yZNR266	BY4741	pWCD2520	pRNR2-Pex11-CFP	LEU2
8	yZNR267	YJD051	pWCD2520	pRNR2-Pex11-CFP	LEU2, Kan
9	yWCD886	yZNR266	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2
10	yWCD887	yZNR267	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2, Kan
11	Library	yWCD321	pWCD2420L	pRPL18B-VioE-6xDNK-PTS1	URA3, LEU2
12	yWCD788	yWCD321	pWCD1134	URA3 Empty vector	URA3, LEU2
13	yWCD789	yWCD321	pWCD2422	pREV1-VioE-YFP-PTS1	URA3, LEU2
14	yWCD790	yWCD321	pWCD2423	pRNR2-VioE-YFP-PTS1	URA3, LEU2
15	yWCD791	yWCD321	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2
16	yWCD792	yWCD321	pWCD2425	pTEF1-VioE-YFP-PTS1	URA3, LEU2
17	yWCD793	yWCD321	pWCD2426	pTDH3-VioE-YFP-PTS1	URA3, LEU2
18	yWCD794	yWCD325	pWCD1134	URA3 Empty vector	URA3, LEU2, Kan
19	yWCD795	yWCD325	pWCD2422	pREV1-VioE-YFP-PTS1	URA3, LEU2, Kan
20	yWCD796	yWCD325	pWCD2423	pRNR2-VioE-YFP-PTS1	URA3, LEU2, Kan
21	yWCD797	yWCD325	pWCD2424	pRPL18B-VioE-YFP-PTS1	URA3, LEU2, Kan

22	yWCD798	yWCD325	pWCD2425	pTEF1-VioE-YFP-PTS1	URA3, LEU2, Kan
23	yWCD799	yWCD325	pWCD2426	pTDH3-VioE-YFP-PTS1	URA3, LEU2, Kan
24	yWCD800	yWCD321	pWCD2427	pREV1-VioE-YFP-ePTS1	URA3, LEU2
25	yWCD801	yWCD321	pWCD2428	pRNR2-VioE-YFP-ePTS1	URA3, LEU2
26	yWCD802	yWCD321	pWCD2429	pRPL18B-VioE-YFP-ePTS1	URA3, LEU2
27	yWCD803	yWCD321	pWCD2430	pTEF1-VioE-YFP-ePTS1	URA3, LEU2
28	yWCD804	yWCD321	pWCD2431	pTDH3-VioE-YFP-ePTS1	URA3, LEU2
29	yWCD805	yWCD325	pWCD2427	pREV1-VioE-YFP-ePTS1	URA3, LEU2
30	yWCD806	yWCD325	pWCD2428	pRNR2-VioE-YFP-ePTS1	URA3, LEU2, Kan
31	yWCD807	yWCD325	pWCD2429	pRPL18B-VioE-YFP-ePTS1	URA3, LEU2, Kan
32	yWCD808	yWCD325	pWCD2430	pTEF1-VioE-YFP-ePTS1	URA3, LEU2, Kan
33	yWCD809	yWCD325	pWCD2431	pTDH3-VioE-YFP-ePTS1	URA3, LEU2, Kan
34	yZNR231	BY4741	pZNR0727	HIS3 Empty vector	HIS3
35	yWCD543	BY4741	pWCD1873	pREV1-TEVprotease-CFP	HIS3
36	yWCD544	BY4741	pWCD1874	pRNR2-TEVprotease-CFP	HIS3
37	yWCD545	BY4741	pWCD1875	pRPL18B-TEVprotease-CFP	HIS3
38	yWCD546	BY4741	pWCD1876	pTEF1-TEVprotease-CFP	HIS3
39	yWCD547	BY4741	pWCD1877	pTDH3-TEVprotease-CFP	HIS3
40	yWCD548	yZNR231	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
41	yWCD549	yWCD543	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
42	yWCD550	yWCD544	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
43	yWCD551	yWCD545	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
44	yWCD552	yWCD546	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
45	yWCD553	yWCD547	pJAG011	pRPL18B-RFP-TEVsite-YFP-PTS1	HIS3, URA3
46	yWCD554	yZNR231	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3
47	yWCD555	yWCD543	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3
48	yWCD556	yWCD544	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3

49	yWCD557	yWCD545	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3
50	yWCD558	yWCD546	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3
51	yWCD559	yWCD547	pWCD1879	pRPL18B-RFP-TEVsite-YFP-ePTS1	HIS3, URA3
52	yWCD560	yZNR231	pWCD1134	URA3 Empty vector	HIS3, URA3
53	yWCD565	yWCD547	pWCD1134	URA3 Empty vector	HIS3, URA3
54	yZNR142	BY4741	pZNR1271 +R1	pPex5::pGal1	None
55	yZNR232	yZNR142	pZNR0915	pRPL18B-RFP-ePTS1	None
56	yZNR157	yZNR142	pZNR1261 +R2	CDC14-ePTS1	None
57	yZNR158	yZNR142	pZNR1261 +R3	CDC14-PTS1	None
58	yZNR147	yZNR142	pZNR1262 +R4	CDC28-ePTS1	None
59	yZNR148	yZNR142	pZNR1262 +R5	CDC28-PTS1	None
60	yZNR159	yZNR142	pZNR1263 +R6	TYS1-ePTS1	None
61	yZNR160	yZNR142	pZNR1263 +R7	TYS1-PTS1	None
62	yZNR153	yZNR142	pZNR1264 +R8	SPC42-ePTS1	None
63	yZNR154	yZNR142	pZNR1264 +R9	SPC42-PTS1	None
64	yZNR174	yZNR157	pZNR1326 +R10	*pex14Δ*	None
65	yZNR175	yZNR158	pZNR1326 +R10	*pex14Δ*	None
66	yZNR165	yZNR147	pZNR1326 +R10	*pex14Δ*	None
67	yZNR166	yZNR148	pZNR1326 +R10	*pex14Δ*	None
68	yZNR176	yZNR159	pZNR1326 +R10	*pex14Δ*	None
69	yZNR177	yZNR160	pZNR1326 +R10	*pex14Δ*	None
70	yZNR171	yZNR153	pZNR1326 +R10	*pex14Δ*	None
71	yZNR172	yZNR154	pZNR1326 +R10	*pex14Δ*	None
72	yWCD319	BY4741	pWCD1441	pTDH3-VioB pHHF2-VioE	LEU2
73	yWCD320	BY4741	pWCD1442	pTEF1-VioA pHHF2-VioE	LEU2
74	yWCD323	YJD051	pWCD1441	pTDH3-VioB pHHF2-VioE	LEU2, Kan
75	yWCD324	YJD051	pWCD1442	pTEF1-VioA pHHF2-VioE	LEU2, Kan
76	yWCD888	yWCD319	pWCD2526	pREV1-VioA-ePTS1	URA3, LEU2
77	yWCD889	yWCD319	pWCD2527	pRNR2-VioA-ePTS1	URA3,

					LEU2
78	yWCD890	yWCD319	pWCD2528	pRPL18B-VioA-ePTS1	URA3, LEU2
79	yWCD891	yWCD319	pWCD2529	pTEF1-VioA-ePTS1	URA3, LEU2
80	yWCD892	yWCD319	pWCD2530	pTDH3-VioA-ePTS1	URA3, LEU2
81	yWCD893	yWCD320	pWCD2531	pREV1-VioB-ePTS1	URA3, LEU2
82	yWCD894	yWCD320	pWCD2532	pRNR2-VioB-ePTS1	URA3, LEU2
83	yWCD895	yWCD320	pWCD2533	pRPL18B-VioB-ePTS1	URA3, LEU2
84	yWCD896	yWCD320	pWCD2534	pTEF1-VioB-ePTS1	URA3, LEU2
85	yWCD897	yWCD320	pWCD2535	pTDH3-VioB-ePTS1	URA3, LEU2
86	yWCD898	yWCD321	pWCD2536	pREV1-VioE-ePTS1	URA3, LEU2
87	yWCD899	yWCD321	pWCD2537	pRNR2-VioE-ePTS1	URA3, LEU2
88	yWCD900	yWCD321	pWCD2538	pRPL18B-VioE-ePTS1	URA3, LEU2
89	yWCD901	yWCD321	pWCD2539	pTEF1-VioE-ePTS1	URA3, LEU2
90	yWCD902	yWCD321	pWCD2540	pTDH3-VioE-ePTS1	URA3, LEU2
91	yWCD903	yWCD323	pWCD2526	pREV1-VioA-ePTS1	URA3, LEU2, Kan
92	yWCD904	yWCD323	pWCD2527	pRNR2-VioA-ePTS1	URA3, LEU2, Kan
93	yWCD905	yWCD323	pWCD2528	pRPL18B-VioA-ePTS1	URA3, LEU2, Kan
94	yWCD906	yWCD323	pWCD2529	pTEF1-VioA-ePTS1	URA3, LEU2, Kan
95	yWCD907	yWCD323	pWCD2530	pTDH3-VioA-ePTS1	URA3, LEU2, Kan
96	yWCD908	yWCD324	pWCD2531	pREV1-VioB-ePTS1	URA3, LEU2, Kan
97	yWCD909	yWCD324	pWCD2532	pRNR2-VioB-ePTS1	URA3, LEU2, Kan
98	yWCD910	yWCD324	pWCD2533	pRPL18B-VioB-ePTS1	URA3, LEU2, Kan
99	yWCD911	yWCD324	pWCD2534	pTEF1-VioB-ePTS1	URA3, LEU2, Kan
100	yWCD912	yWCD324	pWCD2535	pTDH3-VioB-ePTS1	URA3, LEU2,

					Kan
101	yWCD913	yWCD325	pWCD2536	pREV1-VioE-ePTS1	URA3, LEU2, Kan
102	yWCD914	yWCD325	pWCD2537	pRNR2-VioE-ePTS1	URA3, LEU2, Kan
103	yWCD915	yWCD325	pWCD2538	pRPL18B-VioE-ePTS1	URA3, LEU2, Kan
104	yWCD916	yWCD325	pWCD2539	pTEF1-VioE-ePTS1	URA3, LEU2, Kan
105	yWCD917	yWCD325	pWCD2540	pTDH3-VioE-ePTS1	URA3, LEU2, Kan
106	yWCD918	yWCD319	pWCD2541	pRPL18B-VioE-VioA-ePTS1	URA3, LEU2
107	yWCD919	yWCD321	pWCD2541	pRPL18B-VioE-VioA-ePTS1	URA3, LEU2
108	yWCD920	yWCD323	pWCD2541	pRPL18B-VioE-VioA-ePTS1	URA3, LEU2, Kan
109	yWCD921	yWCD325	pWCD2541	pRPL18B-VioE-VioA-ePTS1	URA3, LEU2, Kan
110	yZNR073	BY4741	pZNR0728	pCCW12-Nc_CDT1	LEU2
111	yZNR075	BY4741	pZNR0521	LEU2 Empty vector	LEU2
112	yZNR103	yZNR073	pZNR0873	pREV1-VioE-YFP-BGL-ePTS1	URA3, LEU2
113	yZNR104	yZNR073	pZNR0872	pRNR2-VioE-YFP-BGL-ePTS1	URA3, LEU2
114	yZNR105	yZNR073	pZNR0871	pRPL18B-VioE-YFP-BGL-ePTS1	URA3, LEU2
115	yZNR106	yZNR073	pZNR0864	pTEF1-VioE-YFP-BGL-ePTS1	URA3, LEU2
116	yZNR107	yZNR073	pZNR0870	pTDH3-VioE-YFP-BGL-ePTS1	URA3, LEU2
117	yZNR108	yZNR073	pZNR0875	pTDH3-VioE-YFP-BGL-deadPTS1	URA3, LEU2
118	yZNR109	yZNR073	pZNR0868	pTEF1-VioE-YFP-BGL-deadPTS1	URA3, LEU2
119	yZNR110	yZNR073	pZNR0876	pRPL18B-VioE-YFP-BGL-deadPTS1	URA3, LEU2
120	yZNR111	yZNR073	pZNR0877	pRNR2-VioE-YFP-BGL-deadPTS1	URA3, LEU2
121	yZNR112	yZNR073	pZNR0878	pREV1-VioE-YFP-BGL-deadPTS1	URA3, LEU2
122	yZNR113	yZNR075	pZNR0875	pTDH3-VioE-YFP-BGL-deadPTS1	URA3, LEU2
123	yZNR114	yZNR073	pWCD1134	URA3 Empty vector	URA3, LEU2
124	yZNR115	yWCD321	pWCD1134	URA3 Empty vector	URA3, LEU2
125	yZNR116	yWCD321	pZNR0873	pREV1-VioE-YFP-BGL-ePTS1	URA3, LEU2

126	yZNR117	yWCD321	pZNR0872	pRNR2-VioE-YFP-BGL-ePTS1	URA3, LEU2
127	yZNR118	yWCD321	pZNR0871	pRPL18B-VioE-YFP-BGL-ePTS1	URA3, LEU2
128	yZNR119	yWCD321	pZNR0864	pTEF1-VioE-YFP-BGL-ePTS1	URA3, LEU2
129	yZNR120	yWCD321	pZNR0870	pTDH3-VioE-YFP-BGL-ePTS1	URA3, LEU2
130	yZNR121	yWCD321	pZNR0875	pTDH3-VioE-YFP-BGL-deadPTS1	URA3, LEU2
131	yZNR122	yWCD321	pZNR0868	pTEF1-VioE-YFP-BGL-deadPTS1	URA3, LEU2
132	yZNR123	yWCD321	pZNR0876	pRPL18B-VioE-YFP-BGL-deadPTS1	URA3, LEU2
133	yZNR124	yWCD321	pZNR0877	pRNR2-VioE-YFP-BGL-deadPTS1	URA3, LEU2
134	yZNR125	yWCD321	pZNR0878	pREV1-VioE-YFP-BGL-deadPTS1	URA3, LEU2
135	yWCD591	BY4741	pWCD1932	pTDH3-VioA,pTDH3-VioB,pRPL18B-YFP-VioE-ePTS1	URA3
136	yWCD592	YJD051	pWCD1932	pTDH3-VioA,pTDH3-VioB,pRPL18B-YFP-VioE-ePTS1	URA3, Kan
137	yWCD593	BY4741	pWCD1933	pTDH3-VioA,pTDH3-VioB,pRPL18B-YFP-VioE-deadPTS1	URA3
138	yWCD594	yWCD591	pWCD1934	HIS3 Empty vector	URA3, HIS3
139	yWCD595	yWCD591	pWCD1935	pREV1-RFP-ePTS1	URA3, HIS3
140	yWCD596	yWCD591	pWCD1936	pRNR2-RFP-ePTS1	URA3, HIS3
141	yWCD597	yWCD591	pWCD1937	pRPL18B-RFP-ePTS1	URA3, HIS3
142	yWCD598	yWCD591	pWCD1938	pTEF1-RFP-ePTS1	URA3, HIS3
143	yWCD599	yWCD591	pWCD1924	pTDH3-RFP-ePTS1	URA3, HIS3
144	yWCD600	yWCD591	pWCD1925	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1	URA3, HIS3
145	yWCD601	yWCD591	pWCD1926	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1,pTDH3-RFP-ePTS1	URA3, HIS3
146	yWCD602	yWCD592	pWCD1934	HIS3 Empty vector	URA3, HIS3, Kan
147	yWCD603	yWCD592	pWCD1935	pREV1-RFP-ePTS1	URA3, HIS3, Kan
148	yWCD604	yWCD592	pWCD1936	pRNR2-RFP-ePTS1	URA3, HIS3, Kan
149	yWCD605	yWCD592	pWCD1937	pRPL18B-RFP-ePTS1	URA3, HIS3, Kan
150	yWCD606	yWCD592	pWCD1938	pTEF1-RFP-ePTS1	URA3, HIS3, Kan
151	yWCD607	yWCD592	pWCD1924	pTDH3-RFP-ePTS1	URA3, HIS3,

					Kan
152	yWCD608	yWCD592	pWCD1925	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1	URA3, HIS3, Kan
153	yWCD609	yWCD592	pWCD1926	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1,pTDH3-RFP-ePTS1	URA3, HIS3, Kan
154	yWCD610	yWCD593	pWCD1934	HIS3 Empty vector	URA3, HIS3
155	yWCD611	yWCD593	pWCD1935	pREV1-RFP-ePTS1	URA3, HIS3
156	yWCD612	yWCD593	pWCD1936	pRNR2-RFP-ePTS1	URA3, HIS3
157	yWCD613	yWCD593	pWCD1937	pRPL18B-RFP-ePTS1	URA3, HIS3
158	yWCD614	yWCD593	pWCD1938	pTEF1-RFP-ePTS1	URA3, HIS3
159	yWCD615	yWCD593	pWCD1924	pTDH3-RFP-ePTS1	URA3, HIS3
160	yWCD616	yWCD593	pWCD1925	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1	URA3, HIS3
161	yWCD617	yWCD593	pWCD1926	pTDH3-RFP-ePTS1,pCCW12-RFP_alt-ePTS1,pTDH3-RFP-ePTS1	URA3, HIS3
162	yWCD516	BY4741	pWCD1824	pTDH3-VioA	LEU2
163	yWCD517	YJD051	pWCD1824	pTDH3-VioA	LEU2, Kan
164	yZNR233	yWCD516	pZNR1425	pRPL18B-VioB-YFP-ePTS1,pRNR2-VioE-RFP-ePTS1	URA3, LEU2
165	yZNR234	yWCD516	pZNR1427	pRPL18B-VioB-YFP-ePTS1,pRNR2-VioE-RFP	URA3, LEU2
166	yZNR235	yWCD516	pZNR1429	pRPL18B-VioB-YFP,pRNR2-VioE-RFP-ePTS1	URA3, LEU2
167	yZNR236	yWCD516	pZNR1431	pRPL18B-VioB-YFP,pRNR2-VioE-RFP	URA3, LEU2
168	yZNR237	yWCD517	pZNR1425	pRPL18B-VioB-YFP-ePTS1,pRNR2-VioE-RFP-ePTS1	URA3, LEU2, Kan
169	yZNR238	yWCD517	pZNR1427	pRPL18B-VioB-YFP-ePTS1,pRNR2-VioE-RFP	URA3, LEU2, Kan
170	yZNR239	yWCD517	pZNR1429	pRPL18B-VioB-YFP,pRNR2-VioE-RFP-ePTS1	URA3, LEU2, Kan
171	yZNR240	yWCD517	pZNR1431	pRPL18B-VioB-YFP,pRNR2-VioE-RFP	URA3, LEU2, Kan
172	yZNR241	yWCD516	pZNR1433	pTEF1-VioB-YFP-ePTS1,pREV1-VioE-RFP-ePTS1	URA3, LEU2
173	yZNR242	yWCD516	pZNR1435	pTEF1-VioB-YFP-ePTS1,pREV1-VioE-RFP	URA3, LEU2
174	yZNR243	yWCD516	pZNR1437	pTEF1-VioB-YFP,pREV1-VioE-RFP-ePTS1	URA3, LEU2
175	yZNR244	yWCD516	pZNR1439	pTEF1-VioB-YFP,pREV1-VioE-RFP	URA3, LEU2
176	yZNR245	yWCD517	pZNR1433	pTEF1-VioB-YFP-ePTS1,pREV1-VioE-RFP-ePTS1	URA3, LEU2,

					Kan
177	yZNR246	yWCD517	pZNR1435	pTEF1-VioB-YFP-ePTS1,pREV1-VioE-RFP	URA3, LEU2, Kan
178	yZNR247	yWCD517	pZNR1437	pTEF1-VioB-YFP,pREV1-VioE-RFP-ePTS1	URA3, LEU2, Kan
179	yZNR248	yWCD517	pZNR1439	pTEF1-VioB-YFP,pREV1-VioE-RFP	URA3, LEU2, Kan

3.5.2. Yeast strain construction

Yeast expression vectors were built using Golden Gate Assembly as described in the YTK system[118]. All plasmids were designed for direct integration into the yeast genome via homologous recombination at the URA3 or LEU2 or HIS3 locus. Five hundred nanograms of plasmid were linearized by digestion with NotI and transformed directly into yeast using a standard lithium acetate transformation[176]. Cells were plated onto dropout plates corresponding to the auxotrophic marker present in the integration cassette. Replicate colonies were picked directly from this transformation plate for further analysis. The genomic modifications in Strains 54 and 56-71 were generated using CRISPR/Cas9, in a fashion outlined in the yeast toolkit[118]. Cells were cotransformed with a Cas9+sgRNA expression vector and a repair DNA. Transformants were grown on uracil-selective media to ensure maintenance of the plasmid until the desired genomic modification could be confirmed by sequencing, at which point the cells were transferred to 5-fluoroorotic acid (1 g l^{-1}) plates to encourage clearance of the plasmid. Once the plasmid clearance was verified by PCR for loss of the URA3 marker, the strains were ready for use. Sequences of the modified loci and the linear repair DNA produced for this work are available on request.

3.5.3. Linker library construction

The 6-amino acid linker library was constructed by digesting the pRPL18B-VioE expression vector pWCD1661 with BamHI and XhoI and inserting the annealed oligos EA03 (5′-GATCTGGTAGCDNKDNKDNKDNKDNK-DNKTCCAAATTGTAAC-3′) and EA04 (5′-TCGAGTTACAATTTGGA-MNHMNHMNHMNHMNHMNHGCTACCA-3′) which together contained the (DNK)$_6$ region followed by the canonical PTS1 and a stop codon. Once ligated, this new plasmid library (pWCD2420L) was transformed into EPI300 *E. coli* as described above. Colonies from those transformations were pooled, miniprepped and the resulting DNA was transformed into Strain 2 and plated on synthetic complete 2% dextrose minus uracil to generate the library (Strain 11) used in **Figure 3-7a** and **Figure 3-4**.

3.5.4. Fluorescence microscopy

For confirmation of protein localization, strains were grown in synthetic complete media (minus prototrophic nutrients) with 2% glucose overnight, back diluted 50-fold into fresh media, and grown for an additional 6 hours. Cultures were concentrated by centrifugation, washed in pH 7.4 phosphate buffered saline (PBS), and spotted onto plain glass slides to be examined with a confocal fluorescence microscope.

Strains 9, 10, and 40-51 were examined with confocal microscopy using a CSU-X1 spinning disk confocal microscope (Yokogawa) with Leica optics including a 100x objective and 442, 488, and 561 nm excitation lasers for cyan fluorescent protein (CFP, specifically mTurquoise2), YFP (Venus), and RFP (mKate2) respectively. Images were taken using a QuantEM 512SC EMCCD camera (Photometrics). All images were analyzed using Fiji (http://fiji.sc).

3.5.5. Fluorescent protein measurements

For quantification of relative amounts of fluorescent proteins, cells were pelleted and resuspended in an equal volume of phosphate buffered saline (PBS) pH 7.4 to remove background fluorescence. 100 μl of cells was transferred to a microtiter plate and fluorescence was measured on a Tecan Safire2 or M1000 infinite plate reader. RFP (mKate2) was detected at excitation 580/5 nm and emission 617/5 nm, YFP (Venus) was detected at excitation 516/5 nm and emission 530/5 nm, and CFP (mTurquoise2) was detected at excitation 435/5 nm and emission 478/5 nm.

3.5.6. Plate and spot assays for PDV production

Libraries were generated and plated as described above (**Figure 3-4**), while spots for PDV production were generated by plating 10 μl of saturated culture on synthetic complete 2% dextrose plates missing uracil. Unless otherwise noted, plates were grown for 48 hours at 30°C before pictures were taken using a consumer digital camera.

3.5.7. Production of prodeoxyviolacein

Yeast expressing the PDV pathway (Strains 12-33, 76-109, 124-134, 138-161, and 164-179) were grown overnight at 30˚C with 750 rpm shaking in synthetic dextrose medium without uracil (SD-Ura). Saturated cultures were then diluted 50-fold into 500 μl of fresh SD-Ura and grown for 48 hours, at which point 100 μl of culture was taken for fluorescent protein measurement. The remainder continued incubating for another 12 hours (for a total of 60 hours) before PDV extraction was done.

3.5.8. Extraction and quantification of prodeoxyviolacein production

300 μl of cells were pelleted and the spent media was discarded. Cells were resuspended in 150 μl of glacial acetic acid and transferred to thin-walled PCR tubes. The tubes were then incubated at 95°C for 15 minutes, mixed by inversion, and incubated for another 15 minutes. Cell debris was removed first by centrifugation for 5 minutes at 4700 rpm (4816 x*g* RCF) and subsequently by filtration of the supernatant with an Acroprep Advance 0.2 μm filter plate (Pall Corporation, Port Washington, NY, USA). 100 μl of filtered extract was then transferred to either glass vials or a microtiter plate for quantification via HPLC/MS or bulk fluorescence.

Extracts were separated on an Agilent 1260 Infinity Quaternary LC System using a Zorbax SB C18 (30 × 2.1 mm, 3.5 μm particle size) reversed phase column at 30°C using a 0.5 ml min^{-1} flow rate and an established protocol[177]. Briefly, compounds were separated using a 12-minute complex gradient consisting of: start at 5% B; hold at 5% B for 1.5 min; 16.9% per minute to 98% B; hold at 98% B for 2 min; 3.1%/s to 5% B; hold at 5% B for 2.5 min. Buffer A was water with 0.1% v/v formic acid and buffer B was acetonitrile with 0.1% formic acid. For Strains 164-179, PDV was detected with an Agilent 6520 Q-TOF mass spectrometer in positive mode at a retention time of 5.55 minutes and m/z $[M+H]^+$ of 312.12 and chromopyrrolic acid (CPA) was detected at a retention time of 5.45 minutes and m/z of 386.11. For Strains 12-33 and 76-109, PDV product was detected with an Agilent diode array detector (DAD) at 610 nm.

In the course of our experiments, we found that PDV production was associated with the development of a strongly fluorescent product detectable at excitation 535/5 nm and emission 585/5 nm. Though we were not able to identify the product, we were able to determine that the bulk fluorescence of 100 μl extracts measured on a Tecan M1000 microplate reader linearly corresponds to the amount of PDV detected in those extracts when separated and quantified by HPLC/MS (R^2=0.997) (**Figure 3-5**). Thus, for high-throughput applications such as the linker library (Strain

11) as well as some of our experimental validation (Strains 124-134), we chose to rely on bulk fluorescence for estimating PDV production.

3.5.9. Library quantification and fitting

PDV production of the linker library was quantified fluorescently as described above, and normalized by subtracting autofluorescence from VioE-deficient cells and dividing by productivity in VioE-expressing cells with defective peroxisomal import (*pex5Δ*). Net charge was calculated and plotted against relative PDV production using Plotly (http://plot.ly) as seen in **Figure 3-7a**. A logistic model was then devised to model charge versus relative PDV production, $F_{PDV} = \frac{1}{(1+e^{-k*(x-x_0)})}$ where F_{PDV} is the relative amount of PDV production (ranging from 0 to 1), x is the net charge of the linker at physiological pH (either unweighted or weighted scores), and k and x_0 are fitting parameters. The logistic curve was fit to the full data set using SigmaPlot 13 (Systat Software) (**Figure 3-6**). Parameters for best fit with unweighted net charge were k = -1.1394 and x_0 = -0.8469 with an R^2 = 0.5392 and standard error of regression SER = 0.1928 while the parameters with net charge weighted twice as heavily for the three residues nearest the PTS1 were k = -0.7574 and x_0 = -1.3515 with an R^2 = 0.6067 and SER = 0.1781.

3.5.10. TEV cleavage assay and western blotting

Strains 40-53 were grown to saturation in synthetic complete media (minus uracil) with 2% glucose overnight and back diluted 50-fold into fresh media. After 6 hours of growth with shaking at 30°C, confocal fluorescence microscopy was performed as described above. Relative expression levels for TEV-CFP were determined using bulk fluorescence as described above. The relative fluorescence was obtained via pixel saturation measurement using FIJI (http://fiji.sc), and relative amounts of cleaved and uncleaved protein were calculated as follows: Based on 0.051 μm^2 average cross-sectional area per peroxisome and 0.98 peroxisomes per μm^3 of cell volume from Tam *et al.*[119], we calculate peroxisomes account for 0.85% of the total cell volume. Further algebraic conversion establishes the fraction of protein cleaved before import is equal to $F_{cleaved} = \frac{1}{1+0.0085*B}$ where B is the relative brightness of the peroxisome over cytosol.

For western blotting, 2.5 ODs of cells were pelleted, washed in water, resuspended in 200 mM NaOH and incubated at room temperature for 5 minutes. Cells were again pelleted, resuspended in 50 µl of PAGE sample buffer and boiled at 95°C for 5 minutes. Samples were diluted 10-fold in PAGE sample buffer and 8 µl was loaded onto a NuPAGE Novex 4-12% Bis-Tris gel (Life Technologies) and run for 2.5 hours at 80 volts. Proteins were transferred onto PVDF transfer membrane in NuPAGE transfer buffer (Life Technologies) and blocked overnight in TBST (0.05% Tween-20) with 5% milk. The membrane was washed twice with TBST for 5 minutes and incubated for 1 hour with a rabbit polyclonal anti-RFP antibody Evrogen #AB234 (Evrogen, Moscow, Russia), at a dilution of 1:5,000. After six 5-minute washes in TBST, the membrane was incubated for 1 hour with an HRP conjugated anti-rabbit antibody Abcam #ab16284 (Abcam, Cambridge, MA, USA) at a dilution of 1:5,000. After six 5-minute washes in TBST, the secondary HRP antibody was detected by chemiluminescence using a ChemiDoc XRS imager (Bio-Rad).

3.5.11. Inducible Pex5p expression and RFP sequestration time course

The 533 bases of the region upstream of GAL1 were inserted before the start codon of PEX5 to serve as a new, inducible promoter using a CRISPR/Cas9 system as outlined above. This new *pPex5::pGal1* strain, Strain 54, was further transformed to constitutively express RFP-ePTS1, producing Strain 55. Cells were grown in synthetic complete with 2% raffinose, and in the conditions shown the cells were washed with PBS and resuspended in synthetic complete media with 2% raffinose and 1% galactose and immediately placed on a standard microscope slide with cover slip. No additional action was used to immobilize the cells. Fluorescence was periodically measured on a Zeiss Axioexplorer D1 with a 100x DIC objective using the Texas Red 45 filter set (excitation 560/40 nm, emission 630/75 nm).

3.5.12. Essential protein sequestration

Strain 54 was further modified to add an ePTS1 (GSLGRGRRSKL) or PTS1 (SKL) to the end of Cdc14p, Cdc28p, Tys1p, or Spc42p while leaving the native promoter and terminator intact, resulting in Strains 56-63. These strains were then further modified by a CRISPR-mediated complete deletion of the *PEX14* ORF, resulting in Strains 64-71 with a *pPEX5::pGAL1 pex14Δ* genotype. Strains 56-71 were then propagated on YPD, and overnight cultures were washed in synthetic complete with 2% raffinose and then diluted into the same at a 1:100 dilution. After dilution, cells were allowed to grow at 30˚C, 750 rpm for 2 hours before 10 *µ*l of cells were plated onto 2% agar plates containing synthetic complete with 2% raffinose and either 1% galactose or 0% galactose. Plates were incubated at 30°C for 48 hours, after which a picture was taken.

3.5.13. β-glucosidase mediated permeability measurements

Strains 112-123 expressing the cellodextrin transporter CDT1 and the fusion protein VioE-Venus-BGL-ePTS1 or -deadPTS1 (created by mutating the final residue of the ePTS1 tag to threonine) were grown overnight in YPD. Then, these cultures were diluted 1:100 into synthetic dextrose medium missing uracil and leucine, and incubated shaking at 750 rpm for 14 hours at 30°C. All following work was done at room temperature (22°C). The cells were washed with water, transferred to an anaerobic chamber, and concentrated sixfold in buffer consisting of 2X PBS augmented with 40 mM citrate and brought to pH 3.0 with HCl and NaOH. 50 μl per well of these concentrated cells were loaded into a 384 well plate, to which 50 μl of 800 μM X-substrate in water was added. This gave a final assay concentration of 400 μM X-substrate, 3-fold concentrated cells, 20 mM citrate, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 3.0. The plate was then mixed by pipetting and tightly sealed to maintain the anaerobic conditions, and centrifuged twice for 30 seconds at 500 rcf, rotating the plate to ensure even pelleting. The plate was then transferred to a Tecan M1000 Infinite plate reader, and, 15 minutes after the substrates were added, fluorescence was read via the bottom, allowing a greater signal from pelleted cells.

Venus fluorescence was read at excitation 515/5 nm and emission 530/5 nm, while the 5-Bromo-4-chloro-3-indoxyl (BCI) fluorescence was read at excitation 410/5 nm and emission 490/5 nm. While the hydrolysis of X-Gal and other BCI dyes is normally observed via cyan product 5,5′-dibromo-4,4′-dichloro-indigo formation and absorbance at 610 nm, the BCI monomer exists transiently and can be observed using fluorescence (**Figure 3-15**). Anaerobic conditions were used to prolong the lifetime of the fluorescent product, as reactions that generate the nonfluorescent cyan product are oxygen-dependent. Also note that all hydrolysis experiments were done under acidic conditions (pH ~3), as extracellular activity of BGL from lysate is suppressed and CDT1, being a proton symporter, has enhanced transport activity[134]. Strains expressing VioA and VioB were also transformed with VioE-Venus-BGL-ePTS1 and PDV was extracted and quantified fluorescently as described above (**Figure 3-17**).

X-Glucoside (5-Bromo-4-chloro-3-indoxyl-β-D-glucopyranoside) and X-Cellobioside were obtained from Iris Biotech GmbH (Marktredwitz, Germany) and X-Cellotrioside was obtained from Carbosynth US, LLC (San Diego, CA, USA).

3.5.14. Peptide affinity measurements

The C-terminal fragment (residues 295-612)[178] of *S. cerevisiae* PEX5 was fused to an N-terminal hexahistidine tag and expressed in *E. coli* BL21 (DE3) using a pETDUET vector (Pex5C-pETDUET plasmid pBG400 was a gift of Brooke M. Gardner, UC Berkeley). Cells were induced at mid-logarithmic growth with 1 mM IPTG and incubated for 16 hours at 30°C with 200 rpm shaking. The His6-Pex5c protein was then using a HisTrap FF 5 mL nickel affinity column (GE Biosciences) and subsequently a HiTrap Q HP 5 mL anion exchange column (GE Biosciences). Protein concentration was then measured by Bradford assay, A280 nm on a Nanodrop ND1000, and SDS-PAGE densitometry against a bovine serum albumin samples of known concentration. The protein stock solution was 130 μM His6-Pex5c (5 mg ml^{-1}) in a buffer of 200 mM NaCl, 25 mM HEPES, pH 7.6. 5-carboxyfluorescein-LGRGRRSKL (5FAM-LGRGRRSKL), LGRGRRSKL, and DELYKGSKL peptides were purchased from Elim Biopharmaceuticals Inc. (Hayward, CA, USA) and resuspended in water at 20 mM, 100 mM, and 100 mM respectively. All peptide binding experiments were done with an assay buffer containing 20 nM 5FAM-LGRGRRSKL in 50 mM HEPES pH 7.4, 100 mM NaCl, 0.05% Tween 20. Experiments were equilibrated for 16 hours at 4°C before reading. All raw fluorescence and fluorescence polarization measurements were collected using a Perkin Elmer Victor V3.5 fluorometric plate reader in 384-well small volume non-binding polystyrene plates (Greiner). Fluorescence was measured using 480 nm excitation with 30 nm bandwidth and 535 nm emission with 40 nm bandwidth; raw fluorescence was captured over 1 second, while fluorescence polarization was captured over 0.5 second in the L-format.

Direct binding titrations tested 6.5 μM to 0.0015 nM in twofold increments of His6-Pex5C against 20 nM of 5FAM-LGRGRRSKL. Fluorescence anisotropy was converted to fraction bound using the equation $f_b = \frac{r - r_{free}}{Q(r_{bound} - r) + (r - r_{free})}$ from Maynard et al.[107] where f_b is the fraction of labeled peptide bound to protein, Q is the quantum yield ratio of bound peptide to unbound (accounting for the loss of fluorescence upon binding), r is the fluorescence anisotropy at a given protein concentration, r_{bound} is the anisotropy at maximum binding, and r_{free} is the anisotropy with no protein bound. Key parameters found were $Q = \frac{fluorescence_{bound}}{fluorescence_{free}} = 0.686276,$ and the K_D of Pex5p C-terminal fragment binding to 5FAM-LGRGRRSKL was 355±6.38 nM.

Competition experiments started with a mixture of 375 nM Pex5c and 20 nM 5FAM-LGRGRRSKL and adding 4 mM to 3.8 nM unlabeled DELYKGSKL or 0.5 mM to 0.47 nM unlabeled LGRGRRSKL to compete off 5FAM-LGRGRRSKL from Pex5c. Fluorescence polarization data was analyzed with floating values for the anisotropy of bound and unbound 5FAM-LGRGRRSKL; these values were r_{bound} = 0.2841±0.0045 and r_{free} = 0.0627±0.0025 for DELYKGSKL, and r_{bound} = 0.2993±0.0041 and r_{free} = 0.0642±0.0019 for LGRGRRSKL. Fluorescence anisotropy data were fit to curves using SigmaPlot 13 (Systat Software) and Equation 17 as described in Roehrl et al[179].

Chapter 4. Employing a glucosyl biochemical protecting group for sustainable indigo dyeing

4.1. Introduction

Sustainable, clean sources of specialty chemicals are a matter of increasing importance, and alternatives to classical syntheses are being sought[1]. These clean alternatives, termed "green chemistry," do away with toxic solvents and reagents, replacing them with biocatalysis and microbial production[2]. Microbial production of fine and commodity chemicals can claim several successes, including monomer and polymer products such as poly-L-lactic acid[180], 1,3-propanediol[181], 1,4-butanediol[182], as well as the antimalarial drug artemisinin[19]. Making “drop-in” replacements of key chemicals with biosourced versions while keeping the overall process unchanged is excellent for spurring acceptance of new, biosourced chemicals, but sometimes the process itself needs to be reworked to make full use of the power of microbial production.

One key market where this principle holds true is indigo dyeing. Each year, approximately 40,000 tons of indigo are produced worldwide to support the dyeing of 3 billion pairs of blue jeans[5]. The indigo used in this dye process is produced from aniline using cyanide or chloroacetic acid, requiring extensive cleanup before wastewater can be discharged[5]. Although Amgen[183] and Genencor[184] engineered *E. coli* to produce high titers of indigo, up to 18 g/L[184], microbially-sourced indigo has not achieved commercial success. The key problem is that biosourced indigo, being a drop-in replacement for petroleum-produced indigo, was susceptible to being strategically priced out of the market by existing players[180,185]. However, with increasing regulatory concern for water quality in China[186], where much of the world’s textiles are produced, the public’s attention is not just on how indigo is produced, but how it is used.

A closer look at denim dyeing reveals three key chemicals have remained constants in this process for over a century: indigo dye, sodium hydroxide (also called caustic, used to control pH), and sodium hydrosulfite (also called dithionite, a reducing agent)[187]. Hydrosulfite is of particular concern, as it is a fire hazard and its use results in sulfates, sulfites, and sulfide in the wastewater[187]; a typical textile plant in India might have wastewater 25 times above acceptable sulfide levels[188]. Sodium hydrosulfite is used to reduce indigo dye to its activated, water-soluble form, leucoindigo. This solution of leucoindigo can then be applied to cotton yarn, where it penetrates cotton fibers, binds cellulose and reoxidizes to form insoluble blue crystals of indigo trapped in the cotton. The use of leucoindigo at high pH is an indispensable part of the process, but the use of hydrosulfite to generate leucoindigo is not. Commercial indigo suppliers already sell catalytically pre-reduced leucoindigo, which, due to leucoindigo's spontaneous oxidation in air, must be kept under inert atmosphere until use and used with special equipment. Unfortunately, these drawbacks have prevented pre-reduced leucoindigo from capturing the indigo market[187].

A sustainably sourced process that also removes hazardous hydrosulfite but doesn't require an inert atmosphere would be the ideal solution. We turned to nature for inspiration and developed a biomimetic textile dyeing process based around indican, the stable, pre-reduced indigo precursor that indigo-producing plants such as *Polygonum tinctorium* use as a storage molecule[189]. Using a combined transcriptomics – proteomics approach, we identified the coding sequence of the glucosyltransferase responsible for producing indican. We then engineered *E. coli* to produce indican with titers reaching 4.2 g / L in a single batch fermentation. Finally, we demonstrate a method for concentrating and using this indican to dye cotton. This breakthrough establishes the first sustainable and hydrosulfite-free method to dye cotton using only microbial fermentation products.

4.2. Results

4.2.1. A biomimetic strategy for reductant-free indigo dyeing

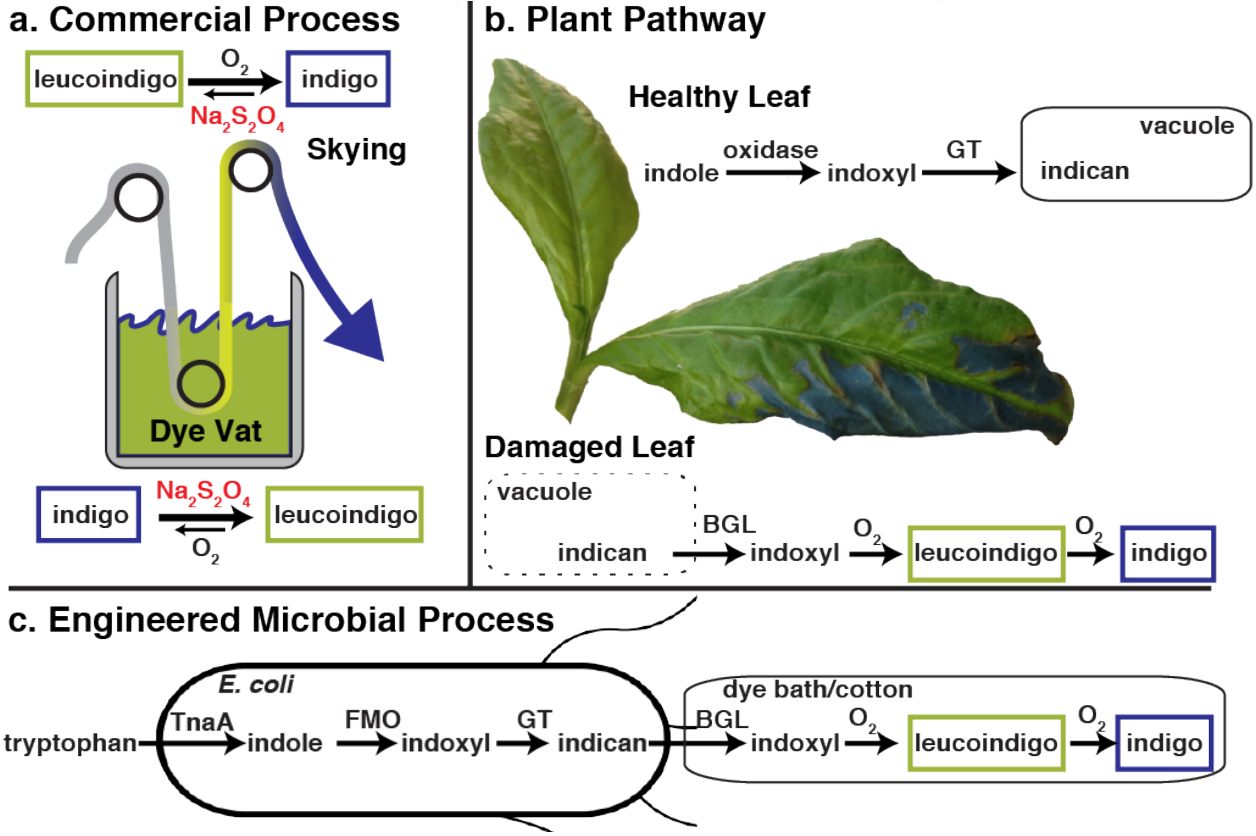


Figure 4-1. A microbial process mimicking plant indigo production allows reductant-free dyeing. (**a**) In current industrial processes, dye mills reduce indigo to soluble leucoindigo using sodium hydrosulfite ($Na_2S_2O_4$), a hazardous chemical. Cotton yarn is dipped into a vat of leucoindigo and then exposed to air to allow the indigo to oxidize, crystallizing in the yarn (skying). (**b**) Indigo-producing plants such as *Polygonum tinctorium* (pictured) do not form indigo when they are alive. Instead, they use a glucosyltransferase (GT) to glucosylate the indigo precursor indoxyl, forming indican, which is then stored in the vacuole. Upon damage or predation, the vacuole is lysed and indican is exposed to β-glucosidases (BGL), releasing indoxyl and forming indigo. (**c**) In our engineered microbial process, tryptophanase (TnaA) converts tryptophan into indole, which is then oxidized by flavin monooxygenase (FMO) into indoxyl. The *Polygonum* UDP-glucosyltransferase (GT) is used to protect indoxyl in the form of indican. Indican is harvested from the culture media, concentrated, and applied to cotton along with a β-glucosidase (BGL) to release free indoxyl. The free indoxyl then oxidizes to leucoindigo and indigo, effectively dyeing the cotton.

Denim dyeing is achieved today the same way it has been achieved for the past hundred years: a dye mill receives synthetic indigo from a chemical supplier, the indigo is mixed in a vat with the reducing agent sodium hydrosulfite to generate soluble leucoindigo, and cotton yarn is passed through the vat to pick up leucoindigo, which is then exposed to air to allow oxidation and production of indigo crystals trapped inside the cotton fibers (**Figure 4-1a**). Typically, sodium hydrosulfite must be added continuously to excess, as air can react with both hydrosulfite directly as well as leucoindigo inside the vat[187]. Although chemical manufacturers can ship pre-reduced leucoindigo under inert atmosphere, the additional expense and dye mill modifications necessary to achieve the greatest decrease in hydrosulfite use have prevented pre-reduced leucoindigo from taking over the market[187]. To find a more workable solution, we turned to nature. Indigo-producing plants such as *Polygonum tinctorium, Indigofera tinctoria*, and *Indigofera suffruticosa* don't actually produce indigo while they are alive and healthy[190] (**Figure 4-1b**). Instead, they produce indican, a glucosylated, air-stable form of the reactive indigo precursor molecule, indoxyl. When the plant is damaged, indican is released from the vacuole[189] and comes in contact with β-glucosidases (BGLs) released from the chloroplast[190]. These BGLs hydrolyze indican, releasing indoxyl molecules, which then dimerize and oxidize into indigo. This process proceeds through leucoindigo as an intermediate, suggesting that a combination of indican and BGL can be used to generate leucoindigo on-demand without a reducing agent. However, since *P. tinctorium* typically contains 14 grams of indican per kilogram of fresh leaves, plant production can reach up to 600 kilograms of indican per hectare per year[191]. Meeting the world's demand for indigo with plant-derived indican would require 70,000 hectares of land, a β-glucosidase-free plant line (so that indican is not hydrolyzed upon harvest), and the capacity to process 8 million tons of plant matter annually (roughly the amount of sorghum produced in the USA in 2010[192]). Thus, a microbial solution is desirable to cut costs and increase yield.

Our solution is to produce the indican necessary for this dyeing process in the bacterium *E. coli* (**Figure 4-1c**). High titers of indigo from either tryptophan (23 g/L after 42 hours) or glucose (18 g / L after 72 hours) have already been demonstrated[184], so we decided to pick the simplest pathway so we could focus on indican production. Thus, we selected the tryptophan-based pathway using endogenous tryptophanase TnaA to minimize the amount of required strain engineering, keeping the glucose-based pathway using deregulated tryptophan biosynthesis as a later option for utilizing inexpensive glucose feedstocks. We then had to select an indole oxidase. New, single-gene indole oxidases such as flavin monooxygenase[193–195] (FMO) and the cytochrome P450 CYP102A1[196] (BM3) and mutants[197–200] have been developed since the landmark Amgen[183] and

Genencor[184] studies of indigo production Another consideration was which BGL to use – several promising options have been reported in the literature[122,190,201–204]. We settled on FMO for our indole oxidase, and *B. circulans* BglA[201,202] for our β-glucosidase. At this point, the only enzymatic functionality remaining to be identified was that of the indican synthase – a UDP-glucosyltransferase that acts on indoxyl to produce indoxyl glucoside, or indican. Although indican synthases have been purified and tested in the literature[189,205], no sequence has yet been reported for these enzymes. Thus, it was necessary to discover the sequence ourselves.

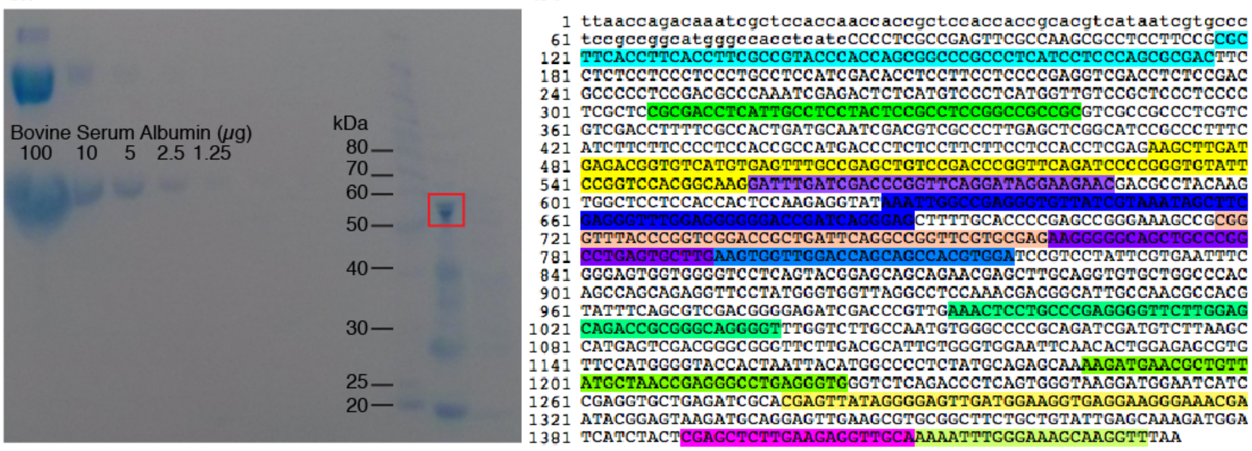
4.2.2. Identification of indican synthases PtUGT1 and PtUGT2

To identify the missing indican synthase, we first prepared a cDNA library from mRNA extracted from the leaves of the indigo-producing plant *Polygonum tinctorium*. The leaves, in particular, are known to house the indican synthase and store produced indican[189]. We sequenced this cDNA library using an Illumina HiSeq 2500 with a 150 bp paired-end read kit, receiving 294,263,920 reads, of which 280,333,100 were high-quality reads. We then digitally normalized these reads and assembled the normalized reads into potential mRNA transcripts with either Trinity[206,207] or Oases[208] and pooled the results from each *de novo* program. This left us with 5,231,568 total predicted transcripts, which we filtered down to 854 predicted UDP-glycosyltransferase (UGT) transcripts by selecting only transcripts which matched (via BLASTX[209]) known UDP-glycosyltransferases. Selecting only unique matches left 15 distinct candidate transcripts (**Table 4-1**). Although it would be possible to clone each individual transcript and test for indican synthase activity, the process of discovering and cloning the full-length transcript from predicted transcript fragments is challenging and time-consuming.

Table 4-1. Key *P. tinctorium* transcriptome assembly statistics.

Sequencing platform	Illumina HiSeq 2500, Rapid Run
Read length	150-bp paired-end
Number of reads (raw)	294,263,920
Yield (Mbases)	42,927
Percent of bases >= Q30	90.52%
Clean reads after trimming	280,333,100
Reads after digital normalization	28,227,790
Number of contigs	5,231,568
N50 of combined assembly	766
Mean contig length	507
Glucosyltransferase contigs	854

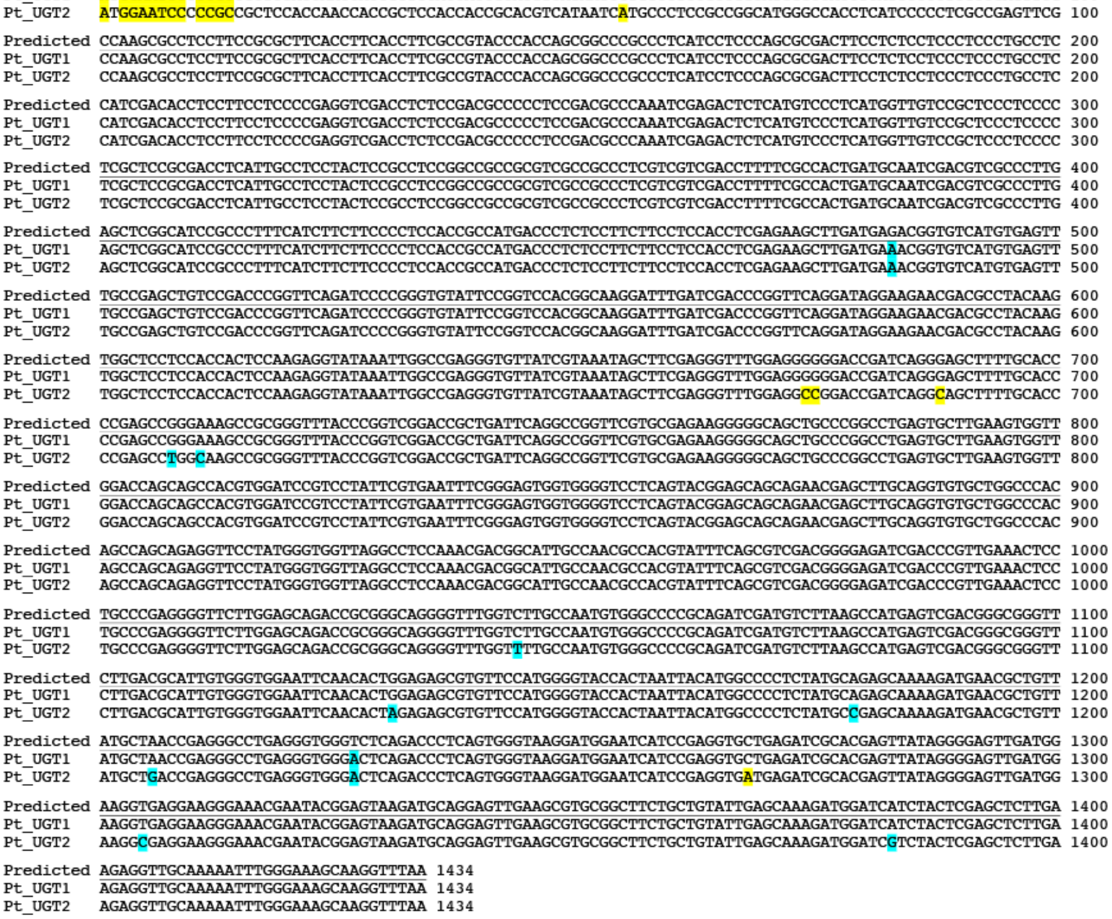
Proteomic data can be used to accelerate the process by matching protein fragments with predicted peptides from the transcriptome, so we purified the *Polygonum tinctorium* indican synthase according to Minami et al[189]. As expected, we observed a ~55 kDa protein band (**Figure 4-2a**) which we trypsin-digested and analyzed with an LC/MS-MS. The fragments observed matched best with one particular predicted transcript in particular, offering 38% coverage of the coding sequence with high-abundance proteins distributed throughout the CDS (**Figure 4-2b**).



```
      *   10    *   20    *   30    *   40    *   50    *   60
   1 ttaaccagacaaatcgctccaccaaccaccgctccaccaccgcacgtcataatcgtgccc
  61 tccgccggcatgggccacctcatcCCCCTCGCCGAGTTCGCCAAGCGCCTCCTTCCGCGC
 121 TTCACCTTCACCTTCGCCGTACCCACCAGCGGCCCGCCCTCATCCTCCCAGCGCGACTTC
 181 CTCTCCTCCCTCCCTGCCTCCATCGACACCTCCTTCCTCCCCGAGGTCGACCTCTCCGAC
 241 GCCCCCTCCGACGCCCAAATCGAGACTCTCATGTCCCTCATGGTTGTCCGCTCCCTCCCC
 301 TCGCTCCGCGACCTCATTGCCTCCTACTCCGCCTCCGGCCGCCGCGTCGCCGCCCTCGTC
 361 GTCGACCTTTTCGCCACTGATGCAATCGACGTCGCCCTTGAGCTCGGCATCCGCCCTTTC
 421 ATCTTCTTCCCCTCCACCGCCATGACCCTCTCCTTCTTCCTCCACCTCGAGAAGCTTGAT
 481 GAGACGGTGTCATGTGAGTTTGCCGAGCTGTCCGACCCGGTTCAGATCCCCGGTGTATT
 541 CCGGTCCACGGCAAGGATTTGATCGACCCGGTTCAGGATAGGAAGAACGACGCCTACAAG
 601 TGGCTCCTCCACCACTCCAAGAGGTATAAATTGGCCGAGGGTGTTATCGTAAATAGCTTC
 661 GAGGGTTTGGAGGGGGGACCGATCAGGGAGCTTTTGCACCCCGAGCCGGGAAAGCCGCGG
 721 GTTTACCCGGTCGGACCGCTGATTCAGGCCGGTTCGTGCGAGAAGGGGGCAGCTGCCCGG
 781 CCTGAGTGCTTGAAGTGGTTGGACCAGCAGCCACGTGGATCCGTCCTATTCGTGAATTTC
 841 GGGAGTGGTGGGGTCCTCAGTACGGAGCAGCAGAACGAGCTTGCAGGTGTGCTGGCCCAC
 901 AGCCAGCAGAGGTTCCTATGGGTGGTTAGGCCTCCAAACGACGGCATTGCCAACGCCACG
 961 TATTTCAGCGTCGACGGGGAGATCGACCCGTTGAAACTCCTGCCCGAGGGGTTCTTGGAG
1021 CAGACCGCGGGCAGGGGTTTGGTCTTGCCAATGTGGGCCCCGCAGATCGATGTCTTAAGC
1081 CATGAGTCGACGGGCGGGTTCTTGACGCATTGTGGGTGGAATTCAACACTGGAGAGCGTG
1141 TTCCATGGGGTACCACTAATTACATGGCCCCTCTATGCAGAGCAAAAGATGAACGCTGTT
1201 ATGCTAACCGAGGGCCTGAGGGTGGGTCTCAGACCCTCAGTGGGTAAGGATGGAATCATC
1261 CGAGGTGCTGAGATCGCACGAGTTATAGGGGAGTTGATGGAAGGTGAGGAAGGGAAACGA
1321 ATACGGAGTAAGATGCAGGAGTTGAAGCGTGCGGCTTCTGCTGTATTGAGCAAAGATGGA
1381 TCATCTACTCGAGCTCTTGAAGAGGTTGCAAAAATTTGGAAAGCAAGGTTTAA
```

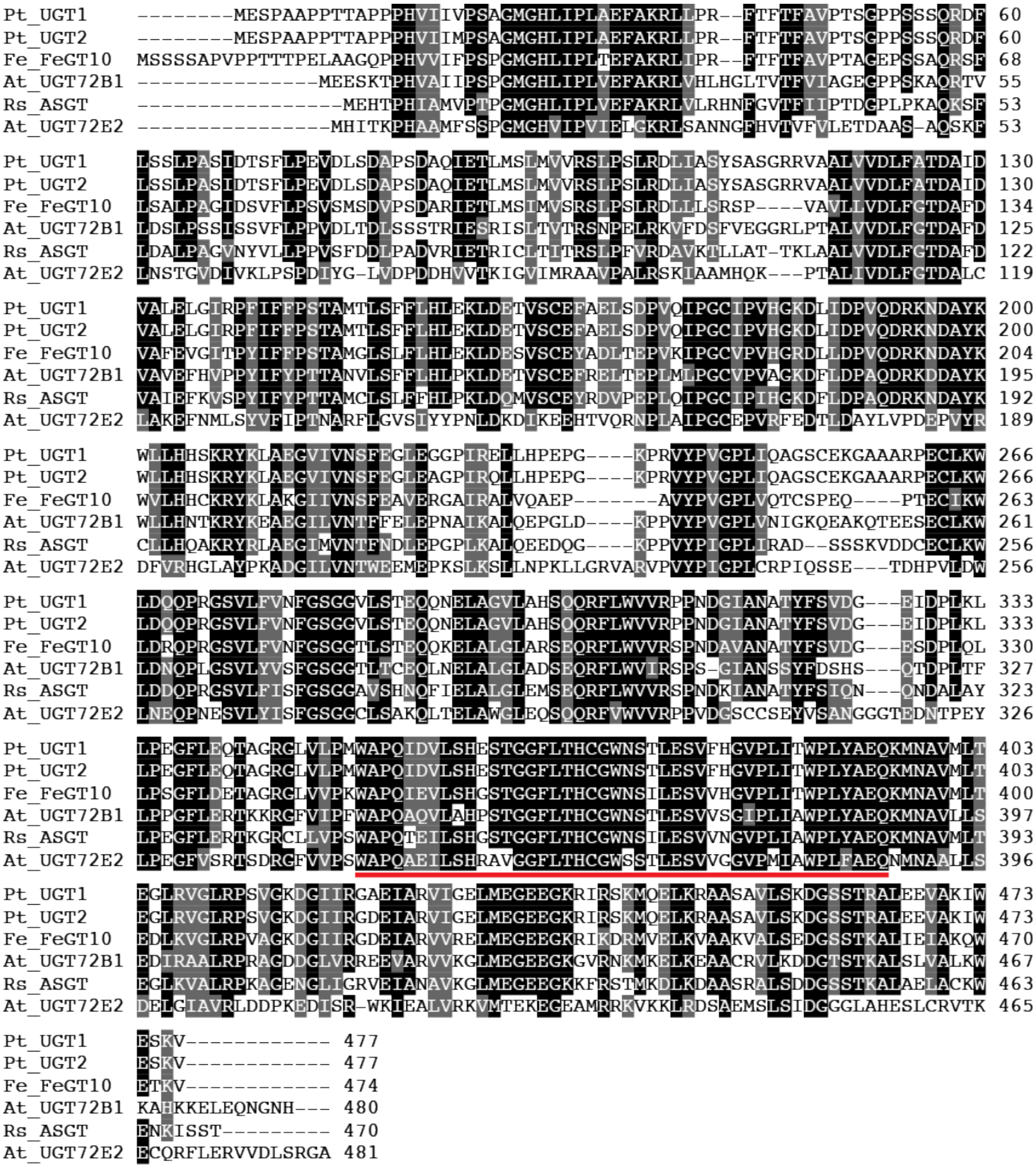
Figure 4-2. Isolation and mass spectrometry analysis of the indican synthase from *Polygonum tinctorium*. (**a**) PVDF membrane blot of an SDS-PAGE gel run with indican synthase fractionated from *P. tinctorium* leaf lysate. The red box highlights the expected indican synthase band, which we cut out and sent for trypsin digest and mass spectrometry analysis. (**b**) Coverage of the transcript with the highest abundance of peptides observed. Regions of the transcript represented by a corresponding peptide observed in the mass spectrometry are highlighted in color, with different colors corresponding to different peptides.

Using the GeneRacer Kit, we successfully cloned and sequenced the full-length cDNA of this transcript. Surprisingly, we obtained two slightly different sequences; one matched the predicted transcript almost perfectly, while the other had several silent and nonsilent mutations (a 1% mismatch rate). We named the glucosyltransferases coded by these cDNAs Pt_UGT1, and Pt_UGT2, respectively (**Figure 4-3**). A BLASTP query of the Uniprot database revealed that Pt_UGT1 is most similar to *Fagopyrum esculentum* FeGT10 (83.2% similarity, e-value = 0, Uniprot: A0A0A1H7P3), with *Rauvolfia serpentina* arbutin synthase[210] (75.3% similarity, e-value = 0, Uniprot: Q9AR73) and *Arabidopsis thaliana* UGT72B1[211,212] (73.5% similarity, e-value = 0, Uniprot: Q9M156) serving as the well-documented representatives of this class of UDP-glucosyltransferase (**Figure 4-4**). Interestingly, these relatives can glucosylate hydroquinone, another reactive molecule, to form arbutin, a stable storage molecule for this natural pesticide[210]. Based on the high degree of similarity, we suspect Pt_UGT1 and Pt_UGT2 may have maintained the ability to glucosylate hydroquinone.



```
Predicted TTAACCAGACAAATCGCTCCACCAACCACCGCTCCACCACCGCACGTCATAATCGTGCCCTCCGCCGGCATGGGCCACCTCATCCCCCTCGCCGAGTTCG 100
Pt_UGT1   ATGGAATCCCCCGCCGCTCCACCAACCACCGCTCCACCACCGCACGTCATAATCGTGCCCTCCGCCGGCATGGGCCACCTCATCCCCCTCGCCGAGTTCG 100
Pt_UGT2   ATGGAATCCCCCGCCGCTCCACCAACCACCGCTCCACCACCGCACGTCATAATCATGCCCTCCGCCGGCATGGGCCACCTCATCCCCCTCGCCGAGTTCG 100

Predicted CCAAGCGCCTCCTTCCGCGCTTCACCTTCACCTTCGCCGTACCCACCAGCGGCCCGCCCTCATCCTCCCAGCGCGACTTCCTCTCCTCCCTCCCTGCCTC 200
Pt_UGT1   CCAAGCGCCTCCTTCCGCGCTTCACCTTCACCTTCGCCGTACCCACCAGCGGCCCGCCCTCATCCTCCCAGCGCGACTTCCTCTCCTCCCTCCCTGCCTC 200
Pt_UGT2   CCAAGCGCCTCCTTCCGCGCTTCACCTTCACCTTCGCCGTACCCACCAGCGGCCCGCCCTCATCCTCCCAGCGCGACTTCCTCTCCTCCCTCCCTGCCTC 200

Predicted CATCGACACCTCCTTCCTCCCCGAGGTCGACCTCTCCGACGCCCCCTCCGACGCCCAAATCGAGACTCTCATGTCCCTCATGGTTGTCCGCTCCCTCCCC 300
Pt_UGT1   CATCGACACCTCCTTCCTCCCCGAGGTCGACCTCTCCGACGCCCCCTCCGACGCCCAAATCGAGACTCTCATGTCCCTCATGGTTGTCCGCTCCCTCCCC 300
Pt_UGT2   CATCGACACCTCCTTCCTCCCCGAGGTCGACCTCTCCGACGCCCCCTCCGACGCCCAAATCGAGACTCTCATGTCCCTCATGGTTGTCCGCTCCCTCCCC 300

Predicted TCGCTCCGCGACCTCATTGCCTCCTACTCCGCCTCCGGCCGCCGCGTCGCCGCCCTCGTCGTCGACCTTTTCGCCACTGATGCAATCGACGTCGCCCTTG 400
Pt_UGT1   TCGCTCCGCGACCTCATTGCCTCCTACTCCGCCTCCGGCCGCCGCGTCGCCGCCCTCGTCGTCGACCTTTTCGCCACTGATGCAATCGACGTCGCCCTTG 400
Pt_UGT2   TCGCTCCGCGACCTCATTGCCTCCTACTCCGCCTCCGGCCGCCGCGTCGCCGCCCTCGTCGTCGACCTTTTCGCCACTGATGCAATCGACGTCGCCCTTG 400

Predicted AGCTCGGCATCCGCCCTTTCATCTTCTTCCCCTCCACCGCCATGACCCTCTCCTTCTTCCTCCACCTCGAGAAGCTTGATGAGACGGTGTCATGTGAGTT 500
Pt_UGT1   AGCTCGGCATCCGCCCTTTCATCTTCTTCCCCTCCACCGCCATGACCCTCTCCTTCTTCCTCCACCTCGAGAAGCTTGATGAAACGGTGTCATGTGAGTT 500
Pt_UGT2   AGCTCGGCATCCGCCCTTTCATCTTCTTCCCCTCCACCGCCATGACCCTCTCCTTCTTCCTCCACCTCGAGAAGCTTGATGAAACGGTGTCATGTGAGTT 500

Predicted TGCCGAGCTGTCCGACCCGGTTCAGATCCCCGGGTGTATTCCGGTCCACGGCAAGGATTTGATCGACCCGGTTCAGGATAGGAAGAACGACGCCTACAAG 600
Pt_UGT1   TGCCGAGCTGTCCGACCCGGTTCAGATCCCCGGGTGTATTCCGGTCCACGGCAAGGATTTGATCGACCCGGTTCAGGATAGGAAGAACGACGCCTACAAG 600
Pt_UGT2   TGCCGAGCTGTCCGACCCGGTTCAGATCCCCGGGTGTATTCCGGTCCACGGCAAGGATTTGATCGACCCGGTTCAGGATAGGAAGAACGACGCCTACAAG 600

Predicted TGGCTCCTCCACCACTCCAAGAGGTATAAATTGGCCGAGGGTGTTATCGTAAATAGCTTCGAGGGTTTGGAGGGGGGACCGATCAGGGAGCTTTTGCACC 700
Pt_UGT1   TGGCTCCTCCACCACTCCAAGAGGTATAAATTGGCCGAGGGTGTTATCGTAAATAGCTTCGAGGGTTTGGAGGGGGGACCGATCAGGGAGCTTTTGCACC 700
Pt_UGT2   TGGCTCCTCCACCACTCCAAGAGGTATAAATTGGCCGAGGGTGTTATCGTAAATAGCTTCGAGGGTTTGGAGGCCGGACCGATCAGGCAGCTTTTGCACC 700

Predicted CCGAGCCGGGAAAGCCGCGGGTTTACCCGGTCGGACCGCTGATTCAGGCCGGTTCGTGCGAGAAGGGGGCAGCTGCCCGGCCTGAGTGCTTGAAGTGGTT 800
Pt_UGT1   CCGAGCCGGGAAAGCCGCGGGTTTACCCGGTCGGACCGCTGATTCAGGCCGGTTCGTGCGAGAAGGGGGCAGCTGCCCGGCCTGAGTGCTTGAAGTGGTT 800
Pt_UGT2   CCGAGCCTGGCAAGCCGCGGGTTTACCCGGTCGGACCGCTGATTCAGGCCGGTTCGTGCGAGAAGGGGGCAGCTGCCCGGCCTGAGTGCTTGAAGTGGTT 800

Predicted GGACCAGCAGCCACGTGGATCCGTCCTATTCGTGAATTTCGGGAGTGGTGGGGTCCTCAGTACGGAGCAGCAGAACGAGCTTGCAGGTGTGCTGGCCCAC 900
Pt_UGT1   GGACCAGCAGCCACGTGGATCCGTCCTATTCGTGAATTTCGGGAGTGGTGGGGTCCTCAGTACGGAGCAGCAGAACGAGCTTGCAGGTGTGCTGGCCCAC 900
Pt_UGT2   GGACCAGCAGCCACGTGGATCCGTCCTATTCGTGAATTTCGGGAGTGGTGGGGTCCTCAGTACGGAGCAGCAGAACGAGCTTGCAGGTGTGCTGGCCCAC 900

Predicted AGCCAGCAGAGGTTCCTATGGGTGGTTAGGCCTCCAAACGACGGCATTGCCAACGCCACGTATTTCAGCGTCGACGGGGAGATCGACCCGTTGAAACTCC 1000
Pt_UGT1   AGCCAGCAGAGGTTCCTATGGGTGGTTAGGCCTCCAAACGACGGCATTGCCAACGCCACGTATTTCAGCGTCGACGGGGAGATCGACCCGTTGAAACTCC 1000
Pt_UGT2   AGCCAGCAGAGGTTCCTATGGGTGGTTAGGCCTCCAAACGACGGCATTGCCAACGCCACGTATTTCAGCGTCGACGGGGAGATCGACCCGTTGAAACTCC 1000

Predicted TGCCCGAGGGGTTCTTGGAGCAGACCGCGGGCAGGGGTTTGGTCTTGCCAATGTGGGCCCCGCAGATCGATGTCTTAAGCCATGAGTCGACGGGCGGGTT 1100
Pt_UGT1   TGCCCGAGGGGTTCTTGGAGCAGACCGCGGGCAGGGGTTTGGTCTTGCCAATGTGGGCCCCGCAGATCGATGTCTTAAGCCATGAGTCGACGGGCGGGTT 1100
Pt_UGT2   TGCCCGAGGGGTTCTTGGAGCAGACCGCGGGCAGGGGTTTGGTTTTGCCAATGTGGGCCCCGCAGATCGATGTCTTAAGCCATGAGTCGACGGGCGGGTT 1100

Predicted CTTGACGCATTGTGGGTGGAATTCAACACTGGAGAGCGTGTTCCATGGGGTACCACTAATTACATGGCCCCTCTATGCAGAGCAAAAGATGAACGCTGTT 1200
Pt_UGT1   CTTGACGCATTGTGGGTGGAATTCAACACTGGAGAGCGTGTTCCATGGGGTACCACTAATTACATGGCCCCTCTATGCAGAGCAAAAGATGAACGCTGTT 1200
Pt_UGT2   CTTGACGCATTGTGGGTGGAATTCAACACTAGAGAGCGTGTTCCATGGGGTACCACTAATTACATGGCCCCTCTATGCCGAGCAAAAGATGAACGCTGTT 1200

Predicted ATGCTAACCGAGGGCCTGAGGGTGGGTCTCAGACCCTCAGTGGGTAAGGATGGAATCATCCGAGGTGCTGAGATCGCACGAGTTATAGGGGAGTTGATGG 1300
Pt_UGT1   ATGCTAACCGAGGGCCTGAGGGTGGGACTCAGACCCTCAGTGGGTAAGGATGGAATCATCCGAGGTGCTGAGATCGCACGAGTTATAGGGGAGTTGATGG 1300
Pt_UGT2   ATGCTGACCGAGGGCCTGAGGGTGGGACTCAGACCCTCAGTGGGTAAGGATGGAATCATCCGAGGTGATGAGATCGCACGAGTTATAGGGGAGTTGATGG 1300

Predicted AAGGTGAGGAAGGGAAACGAATACGGAGTAAGATGCAGGAGTTGAAGCGTGCGGCTTCTGCTGTATTGAGCAAAGATGGATCATCTACTCGAGCTCTTGA 1400
Pt_UGT1   AAGGTGAGGAAGGGAAACGAATACGGAGTAAGATGCAGGAGTTGAAGCGTGCGGCTTCTGCTGTATTGAGCAAAGATGGATCATCTACTCGAGCTCTTGA 1400
Pt_UGT2   AAGGCGAGGAAGGGAAACGAATACGGAGTAAGATGCAGGAGTTGAAGCGTGCGGCTTCTGCTGTATTGAGCAAAGATGGATCGTCTACTCGAGCTCTTGA 1400

Predicted AGAGGTTGCAAAAATTTGGGAAAGCAAGGTTTAA 1434
Pt_UGT1   AGAGGTTGCAAAAATTTGGGAAAGCAAGGTTTAA 1434
Pt_UGT2   AGAGGTTGCAAAAATTTGGGAAAGCAAGGTTTAA 1434
```

Figure 4-3. Multiple sequence alignment of Pt_UGT1 and Pt_UGT2 versus NGS-predicted transcript CDS. Mutations are annotated relative to the predicted transcript, with silent mutations in cyan and missense mutations in yellow. With the exception of the low-coverage N-terminal region, Pt_UGT1 matched the NGS results with only two silent mutations. Pt_UGT2 had four missense mutations relative to Pt_UGT1: V19M, G225A, E230Q, and A423D.



```
Pt_UGT1    --------MESPAAPPTTAPPPHVIIVPSAGMGHLIPLAEFAKRLLPR--FTFTFAVPTSGPPSSSQRDF 60
Pt_UGT2    --------MESPAAPPTTAPPPHVIIMPSAGMGHLIPLAEFAKRLLPR--FTFTFAVPTSGPPSSSQRDF 60
Fe_FeGT10  MSSSSAPVPPTTTPELAAGQPPHVVIFPSPGMGHLIPLTEFAKRLIPR--FTFTFAVPTAGEPSSAQRSF 68
At_UGT72B1 ---------------MEESKTPHVAIIPSPGMGHLIPLVEFAKRLVHLHGLTVTFVIAGEGPPSKAQRTV 55
Rs_ASGT    -----------------MEHTPHIAMVPTPGMGHLIPLVEFAKRLVLRHNFGVTFIIPTDGPLPKAQKSF 53
At_UGT72E2 ----------------MHITKPHAAMFSSPGMGHVIPVIELGKRLSANNGFHVTVFVLETDAAS-AQSKF 53

Pt_UGT1    LSSLPASIDTSFLPEVDLSDAPSDAQIETLMSLMVVRSLPSLRDLIASYSASGRRVAALVVDLFATDAID 130
Pt_UGT2    LSSLPASIDTSFLPEVDLSDAPSDAQIETLMSLMVVRSLPSLRDLIASYSASGRRVAALVVDLFATDAID 130
Fe_FeGT10  LSALPAGIDSVFLPSVSMSDVPSDARIETLMSIMVSRSLPSLRDLLLSRSP----VAVLLVDLFGTDAFD 134
At_UGT72B1 LDSLPSSISSVFLPPVDLTDLSSSTRIESRISLTVTRSNPELRKVFDSFVEGGRLPTALVVDLFGTDAFD 125
Rs_ASGT    LDALPAGVNYVLLPPVSFDDLPADVRIETRICLTITRSLPFVRDAVKTLLAT-TKLAALVVDLFGTDAFD 122
At_UGT72E2 LNSTGVDIVKLPSPDIYG-LVDPDDHVVTKIGVIMRAAVPALRSKIAAMHQK---PTALIVDLFGTDALC 119

Pt_UGT1    VALELGIRPFIFFPSTAMTLSFFLHLEKLDETVSCEFAELSDPVQIPGCIPVHGKDLIDPVQDRKNDAYK 200
Pt_UGT2    VALELGIRPFIFFPSTAMTLSFFLHLEKLDETVSCEFAELSDPVQIPGCIPVHGKDLIDPVQDRKNDAYK 200
Fe_FeGT10  VAFEVGITPYIFFPSTAMGLSLFLHLEKLDESVSCEYADLTEPVKIPGCVPVHGRDLLDPVQDRKNDAYK 204
At_UGT72B1 VAVEFHVPPYIFYPTTANVLSFFLHLPKLDETVSCEFRELTEPLMLPGCVPVAGKDFLDPAQDRKDDAYK 195
Rs_ASGT    VAIEFKVSPYIFYPTTAMCLSLFFHLPKLDQMVSCEYRDVPEPLQIPGCIPIHGKDFLDPAQDRKNDAYK 192
At_UGT72E2 LAKEFNMLSYVFIPTNARFLGVSIYYPNLDKDIKEEHTVQRNPLAIPGCEPVRFEDTLDAYLVPDEPVYR 189

Pt_UGT1    WLLHHSKRYKLAEGVIVNSFEGLEGGPIRELLHPEPG----KPRVYPVGPLIQAGSCEKGAAARPECLKW 266
Pt_UGT2    WLLHHSKRYKLAEGVIVNSFEGLEAGPIRQLLHPEPG----KPRVYPVGPLIQAGSCEKGAAARPECLKW 266
Fe_FeGT10  WVLHHCKRYKLAKGIIVNSFEAVERGAIRALVQAEP-------AVYPVGPLVQTCSPEQ----PTECIKW 263
At_UGT72B1 WLLHNTKRYKEAEGILVNTFFELEPNAIKALQEPGLD----KPPVYPVGPLVNIGKQEAKQTEESECLKW 261
Rs_ASGT    CLLHQAKRYRLAEGIMVNTFNDLEPGPLKALQEEDQG----KPPVYPIGPLIRAD--SSSKVDDCECLKW 256
At_UGT72E2 DFVRHGLAYPKADGILVNTWEEMEPKSLKSLLNPKLLGRVARVPVYPIGPLCRPIQSSE---TDHPVLDW 256

Pt_UGT1    LDQQPRGSVLFVNFGSGGVLSTEQQNELAGVLAHSQQRFLWVVRPPNDGIANATYFSVDG---EIDPLKL 333
Pt_UGT2    LDQQPRGSVLFVNFGSGGVLSTEQQNELAGVLAHSQQRFLWVVRPPNDGIANATYFSVDG---EIDPLKL 333
Fe_FeGT10  LDRQPRGSVLFVNFGSGGTLSTEQQKELALGLARSEQRFLWVVRSPNDAVANATYFSVDG---ESDPLQL 330
At_UGT72B1 LDNQPLGSVLYVSFGSGGTLTCEQLNELALGLADSEQRFLWVIRSPS-GIANSSYFDSHS---QTDPLTF 327
Rs_ASGT    LDDQPRGSVLFISFGSGGAVSHNQFIELALGLEMSEQRFLWVVRSPNDKIANATYFSIQN---QNDALAY 323
At_UGT72E2 LNEQPNESVLYISFGSGGCLSAKQLTELAWGLEQSQQRFVWVVRPPVDGSCCSEYVSANGGGTEDNTPEY 326

Pt_UGT1    LPEGFLEQTAGRGLVLPMWAPQIDVLSHESTGGFLTHCGWNSTLESVFHGVPLITWPLYAEQKMNAVMLT 403
Pt_UGT2    LPEGFLEQTAGRGLVLPMWAPQIDVLSHESTGGFLTHCGWNSTLESVFHGVPLITWPLYAEQKMNAVMLT 403
Fe_FeGT10  LPSGFLDETAGRGLVVPKWAPQIEVLSHGSTGGFLTHCGWNSILESVVHGVPLITWPLYAEQKMNAVMLT 400
At_UGT72B1 LPPGFLERTKKRGFVIPFWAPQAQVLAHPSTGGFLTHCGWNSTLESVVSGIPLIAWPLYAEQKMNAVLLS 397
Rs_ASGT    LPEGFLERTKGRCLLVPSWAPQTEILSHGSTGGFLTHCGWNSILESVVNGVPLIAWPLYAEQKMNAVMLT 393
At_UGT72E2 LPEGFVSRTSDRGFVVPSWAPQAEILSHRAVGGFLTHCGWSSTLESVVGGVPMIAWPLFAEQNMNAALLS 396

Pt_UGT1    EGLRVGLRPSVGKDGIIRGAEIARVIGELMEGEEGKRIRSKMQELKRAASAVLSKDGSSTRALEEVAKIW 473
Pt_UGT2    EGLRVGLRPSVGKDGIIRGDEIARVIGELMEGEEGKRIRSKMQELKRAASAVLSKDGSSTRALEEVAKIW 473
Fe_FeGT10  EDLKVGLRPVAGKDGIIRGDEIARVVRELMEGEEGKRIKDRMVELKVAAKVALSEDGSSTKALIEIAKQW 470
At_UGT72B1 EDIRAALRPRAGDDGLVRREEVARVVKGLMEGEEGKGVRNKMKELKEAACRVLKDDGTSTKALSLVALKW 467
Rs_ASGT    EGLKVALRPKAGENGLIGRVEIANAVKGLMEGEEGKKFRSTMKDLKDAASRALSDDGSSTKALAELACKW 463
At_UGT72E2 DELGIAVRLDDPKEDISR-WKIEALVRKVMTEKEGEAMRRKVKKLRDSAEMSLSIDGGGLAHESLCRVTK 465

Pt_UGT1    ESKV------------ 477
Pt_UGT2    ESKV------------ 477
Fe_FeGT10  ETKV------------ 474
At_UGT72B1 KAHKKELEQNGNH--- 480
Rs_ASGT    ENKISST--------- 470
At_UGT72E2 ECQRFLERVVDLSRGA 481
```

Figure 4-4. Multiple sequence alignment of Pt_UGT1 and homologs. Pt_UGT1 and Pt_UGT2 were aligned to their most similar homolog, FeGT10 from *Fagopyrum esculentum*, as well as two less similar but better-documented homologs arbutin synthase from *Rauvolfia serpentina* (Rs_ASGT) and UGT72B1 from *Arabidopsis thaliana*. The next-most similar non-UGT72B GT from *Arabidopsis*, UGT72E2, did not glucosylate indoxyl. Conserved regions are highlighted, with identical residues in black and similar residues in gray. The highly conserved plant secondary product glycosyltransferase (PSPG) box common to plant secondary metabolite GTs is underlined in red.

4.2.3. Microbial production of indican

To verify that Pt_UGT1 is indeed an indican synthase, we expressed Pt_UGT1 in indigo-producing *E. coli* expressing FMO. We observed a disappearance of indigo product (**Figure 4-5a**) and detected the production of colorless indican via LC/MS (**Figure 4-5b,c**). Though this was quite an achievement, we also observed that unmodified *E. coli* produced indigo from exogenously added indican (**Figure 4-6a**), suggesting that *E. coli* possess some mechanism for hydrolyzing indican. The *E. coli* phospho-β-glucosidase BglA is known to confer the ability to hydrolyze arbutin[213], a glucoside with strong structural similarity to indican, so we tested *E. coli* strains with *bglA* knocked out (**Figure 4-6**). As expected, these strains lost the ability to convert indican into indigo (**Figure 4-6a**), resulting in virtually no indican loss over the course of 48 hours, as opposed to 46% loss with the wildtype strain (**Figure 4-6b**). This strain modification also boosted our indican titers 19% from 439 mg / L to 521 mg /L (**Figure 4-6c**).

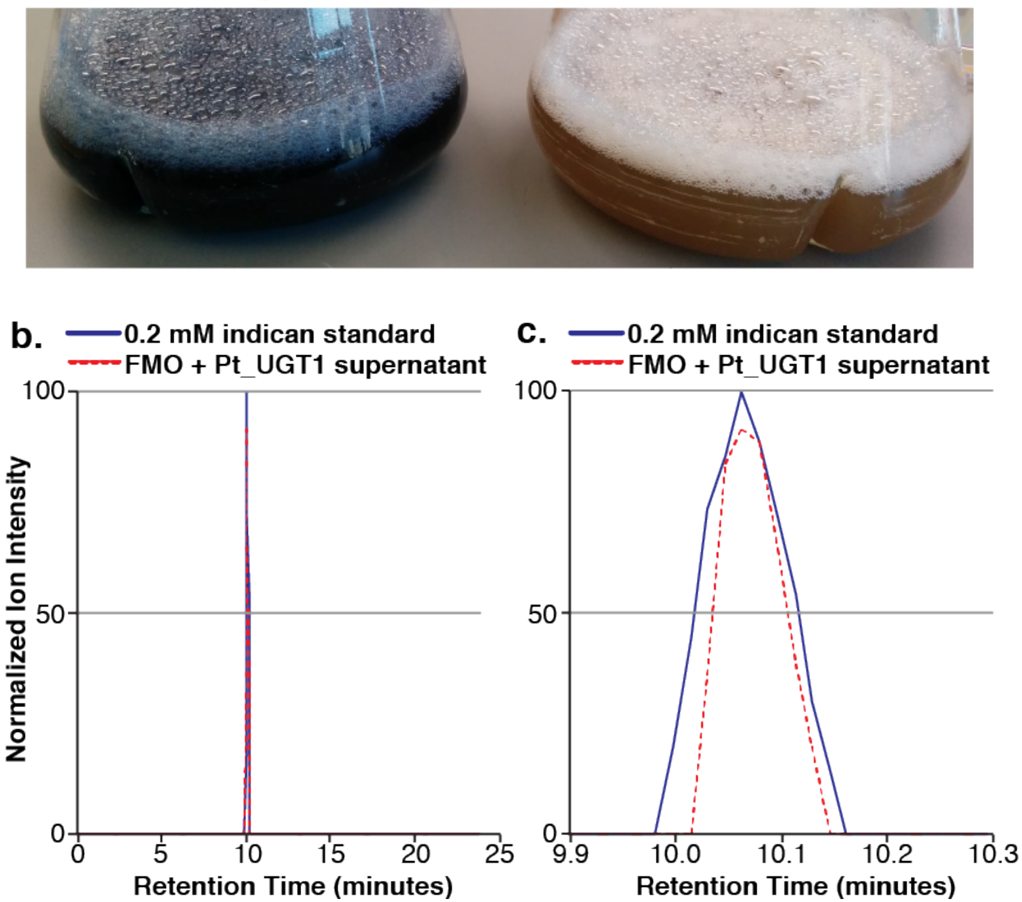


Figure 4-5. Expression of Pt_UGT1 allows microbial production of indican from indoxyl. (**a**) *E. coli* expressing either FMO only or FMO and Pt_UGT1 were grown in LB supplemented with tryptophan. The blue color of the left flask indicates indigo production, while indican is colorless. (**b**) Supernatant from a culture of FMO + Pt_UGT1 strain in MOPS rich media was analyzed via LC/MS. Traces show $[M+H]^+$ ions matching m/z = 296.0929-296.1329 (indican). A genuine indican standard is also shown.

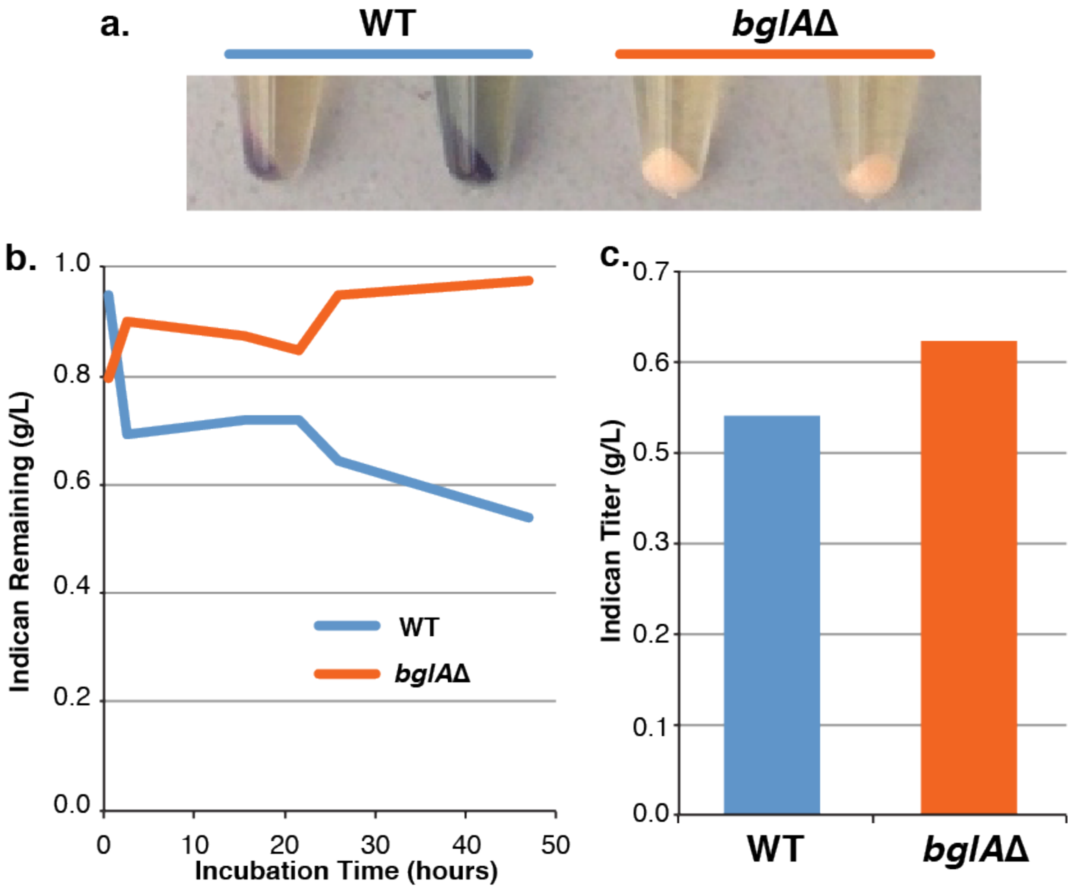


Figure 4-6. Knocking out *bglA* abolishes indican consumption in *E. coli*. (**a**) Pellets of *E. coli* grown for 47 hours in LB supplemented with 1 g / L indican. The blue color indicates hydrolysis of indican and conversion to indigo. The *bglAΔ* strain is knocked out for the phospho-β-glucosidase BglA, while WT is wildtype. (**b**) LC/MS analysis of the supernatants from those same cultures. Again, wildtype (blue) is compared to *bglAΔ* (red). (**c**) Indican production titers from FMO and Pt_UGT1-expressing cells in either a wildtype (blue) or *bglAΔ* background. Cells were grown in LB with 5 mM tryptophan for 36 hours.

With strain optimizations in place, we copied our indican production system (FMO and Pt_UGT1) into the bacterial chromosome and scaled up our production of indican. We found that by increasing the amount of tryptophan fed into the system, we could generate up to 14.3 mM (or 4.2 g/L) indican after 24 hours (**Figure 4-7**). Considering that 25 mM of tryptophan was fed in, this represents a 57% conversion of tryptophan into indican. Unfortunately, indican production did not continue to increase past 25 mM tryptophan. To discover the bottleneck, we analyzed each of the metabolites and side products on the pathway to indican: tryptophan, indole, indigo, and indican (**Figure 4-8**). We omitted indoxyl because indoxyl is unstable and spontaneously forms indigo. Although indigo production appears to increase with high levels of tryptophan, the 38 mM of leftover tryptophan suggests that the main culprit is failure to convert extra tryptophan into indole. Thus, future efforts to enhance indican production will have to improve TnaA activity or use the endogenous tryptophan biosynthesis pathway for producing indole. Conveniently, this system is amenable to moderately high-throughput colorimetric screening on plates for either gain-of-signal phenotypes (increases in indigo production) or loss-of-signal phenotypes (replacement of indigo production with indican production). An example of such a screen is highlighted in **Figure 4-9**.

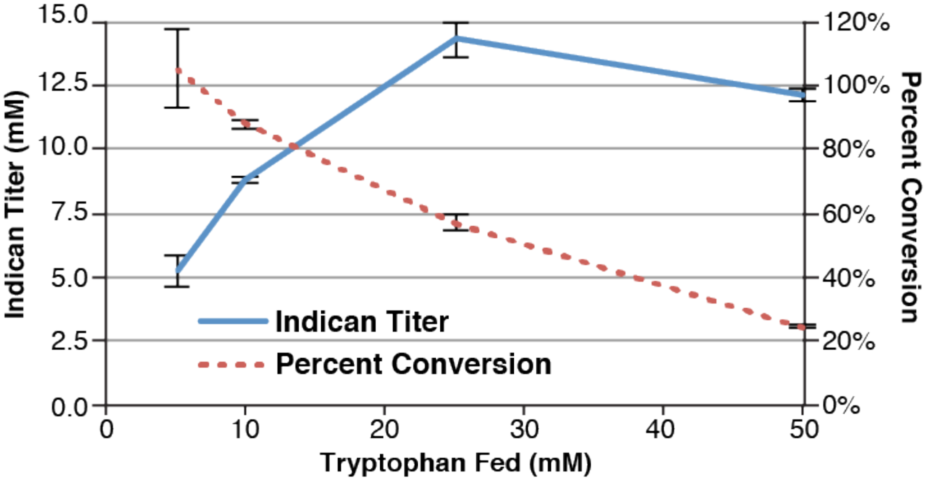


Figure 4-7. Feeding additional tryptophan increases indican titers with diminishing returns. *E. coli* expressing FMO and Pt_UGT1 were grown for 24 hours in EZ Rich MOPS media with the indicated amount of tryptophan. Indican in the supernatant was measured via LC/MS. The percent conversion is calculated as 100% = 1 mol of indican from 1 mol of fed tryptophan. Error bars are the mean ± s.d. of six biological replicates.

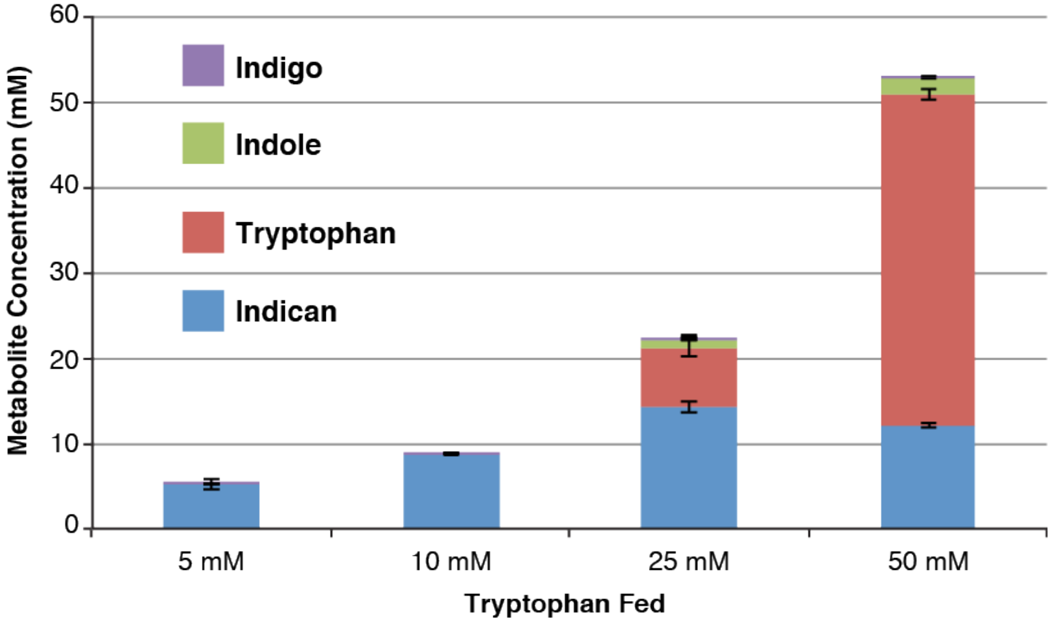


Figure 4-8. Leftover tryptophan accounts for much of the missing indican at high tryptophan concentrations. *E. coli* expressing FMO and Pt_UGT1 were grown for 24 hours in EZ Rich MOPS media with the indicated amount of tryptophan. Indican, tryptophan, and indole in the supernatant were measured via LC/MS, while indigo was extracted from the pellets with DMSO and measured via absorbance. Error bars are the mean ± s.d. of six biological replicates.

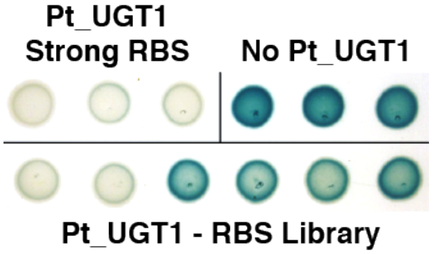


Figure 4-9. Indigo and indigo derivatives naturally lend themselves to colorimetric screens. In this case, 5-bromo-4-chloro-indoxyl acetate (BCIA or X-acetate) was added to LB plates on which *E. coli* were spotted. Natural esterases in the *E. coli* hydrolyse the BCIA, producing BCI, which, like indoxyl, dimerizes to form a blue product. The *E. coli* expressing Pt_UGT1 at high levels are able to effectively suppress the formation of blue product by making colorless X-glucoside, while *E. coli* with lower levels of Pt_UGT1 activity have an intermediate phenotype. The white balance of this image has been adjusted to make the effect easier to see.

4.2.4. Demonstration of indican-based dyeing

With bioproduced indican in hand, we turned to demonstrating indigo dyeing. The *B. circulans* BGL was expressed in *E. coli* and purified (**Figure 4-10**). Supernatant from indican-producing *E. coli* was boiled to concentrate the indican from 0.6 g /L to 3.4 g /L (**Figure 4-11a**). We are still trying to identify the orange, water-soluble contaminant, which appears to be formed from a reaction between the culture media and high concentrations of indican. Cotton swatches were then soaked in boiled indican-laden media for 10 minutes (**Figure 4-11b**), after which droplets of BGL were applied to the still-wet swatches. The BGL was allowed to act on the swatches at ambient temperature for 6 hours, developing the indigo color (**Figure 4-11c**). Finally, the swatches were washed repeatedly with water (**Figure 4-11d**), removing the orange contaminant and leaving behind the indigo. This wash-resistant color fastness indicates that the indigo is truly embedded in the cotton and not merely crystallized on the surface. Thus, this is the first demonstration of successful indigo dyeing using microbial indican and not hazardous chemicals like sodium hydrosulfite.

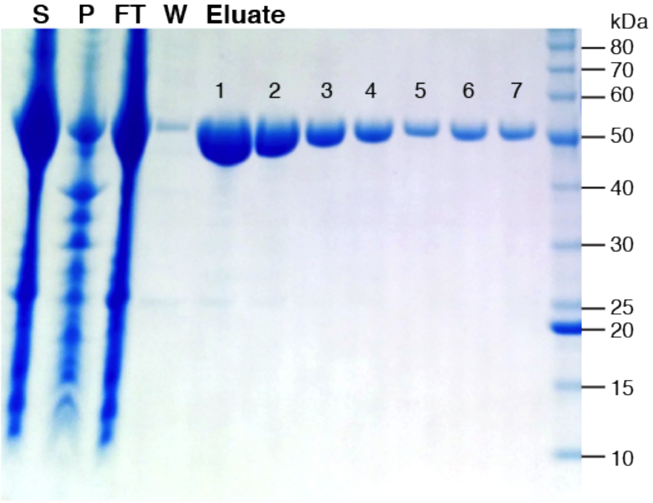


Figure 4-10. Purification of *B. circulans* β-glucosidase from an *E. coli* host. This SDS-PAGE gel shows lysate and Ni-NTA purified protein from *E. coli* BL21 DE3 cells expressing *B. circulans* BglA. Lane "S" is the soluble fraction of the lysate, "P" is the pellet, "FT" is the flow-through from the Ni-NTA column, "W" is the wash flow-through, and 1-7 indicate subsequent imidazole elution fractions. The 54 kDa BglA band appears near the 50 kDA marker. The white balance of this image has been adjusted to make the picture clearer.

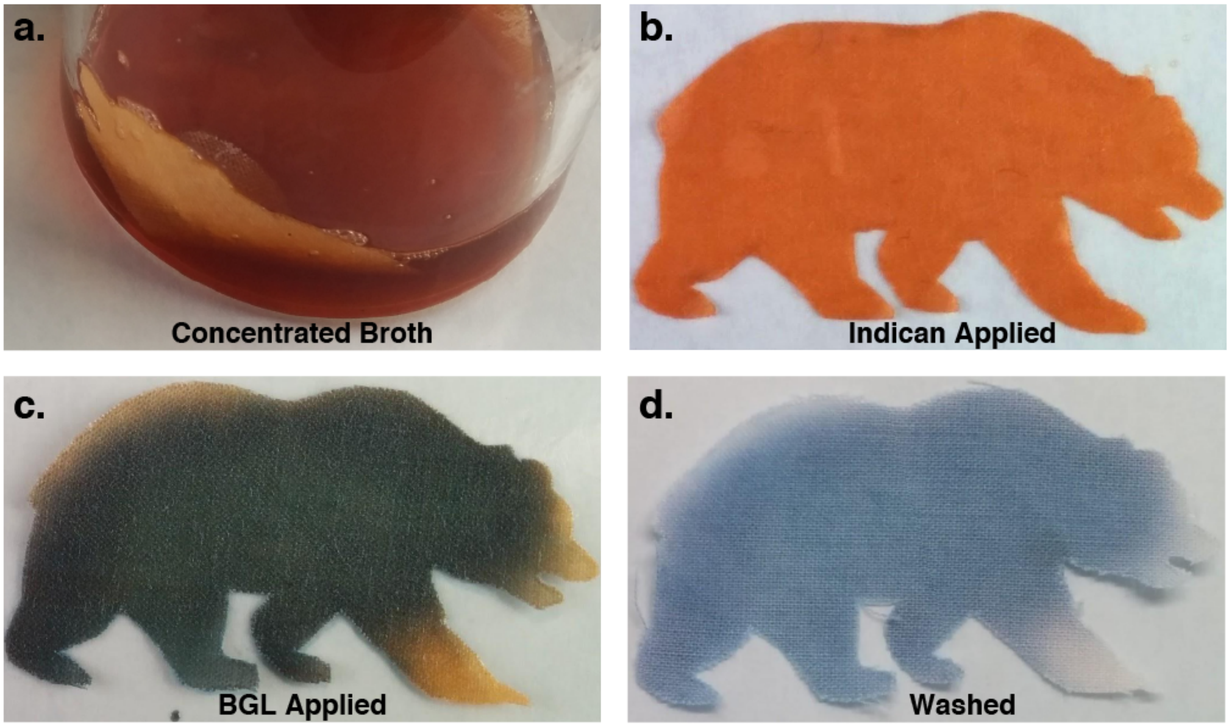


Figure 4-11. Reductant-free indigo dyeing with indican and β-glucosidase. (**a**) Culture broth from indican-producing *E. coli* was concentrated by boiling, leaving a final concentration of 3.4 g / L indican. A cotton swatch is shown soaking in the concentrated broth. An unknown orange contaminant appears to form when high titers of indican are mixed with EZ Rich MOPS media. (**b**) After 10 minutes of soaking, the swatch was removed from the broth and spread on wax paper. (**c**) *B. circulans* BglA was applied to the swatch and allowed to react for six hours, converting the colorless indican to blue indigo. The swatch shown has completed the six-hour incubation. (**d**) At the end of six hours, the swatch is repeatedly washed in water, removing the orange contaminant. The indigo color stays with the fabric, indicating wash-fastness.

4.3. Discussion

While there are numerous strategies available for adapting microbial hosts and their enzymes to optimize production of valuable chemicals[214], sometimes the compounds themselves are the problem. Important fermentation products and intermediates can exhibit instability[215], insolubility[216], or host toxicity[217]. These challenges can be addressed *in vivo* with glucosylation, plants' counterpart to the chemist's protecting group. In plants, glucosyl moieties are frequently attached or removed to modulate small molecule solubility, stability, bioactivity, block off-target reactivity, and provide a convenient handle for molecular recognition by transporters[215]. We demonstrate the successful application of this strategy to the textile industry by engineering microbes to produce the stable indigo precursor, indican, and using that indican to dye cotton without hazardous reducing agents. This breakthrough in environmentally friendly dyeing relies entirely on mixing the fermentation products of two bacterial strains, an indican producer and a β-glucosidase producer.

Additionally, the identification of an indican synthase and discovery that it is a close relative of arbutin synthase fills in a significant gap in our understanding of indigo-producing plants. Though it is not surprising in retrospect that a glucosyltransferase that recognizes a reactive aromatic (hydroquinone) might also be adapted to recognize another reactive aromatic (indoxyl), it is interesting that it was not one of the other candidates put forward based on substrate similarity, such as indole-3-acetic-acid (IAA) glucosyltransferase and 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) glucosyltransferase[218].

While we achieved remarkable indican titers (4.2 g/L after 24 hours) without extensive strain engineering, additional engineering of aromatic acid production will be necessary to produce indican entirely from low-cost starting material like glycerol or glucose and ammonia[184]. Considering the failure of TnaA to completely process tryptophan concentrations of 25 mM and above, using the aromatic acid production pathway (which bypasses TnaA) may boost titers as well. Surprisingly, no engineering of UDP-glucose production was necessary, suggesting that *E. coli* has enormous natural capacity for glucosylation; this natural capacity has been further expanded by eightfold with strain engineering[219]. Strain modifications also have been deployed to add the capability to glycosylate with other sugars, such as xylose and glucuronic acid[220], and glucosyltransferases have previously been engineered to accept new sugars[221]. Together, these enhancements – boosting titers, using glycerol instead of tryptophan to generate indole, and using xylose instead of glucose to supply the protecting group – can further drop production costs to help meet the demands of the indigo market.

China specifically named textiles as one of ten water-polluting industries requiring special interventions to clean up, with action ongoing this year[186]. We believe that stable, protected dye precursors like indican are the breakthrough in green chemistry this field needs right now.

4.4. Extensions and Future Directions

4.4.1. A broader look at biological protecting groups

The fact that we found our indoxyl glucosyltransferase was a close relative of glucosyltransferases that detoxify hydroquinone and even herbicides[211] suggests a very broad role for biological protecting groups. This detoxification can even work across multiple kingdoms, as a *Vitis vinifera* (grape) UDP-glucosyltransferase can detoxify herbicides and pesticides *in plantae* but can also function when transformed into the yeast *Pichia pastoris* to protect against those same chemicals and others[222]. The line between secondary product metabolism and xenobiotic resistance becomes quite blurry when the same enzyme pulls double duty on different molecules, or when the natural product itself is a biologically protected toxic molecule. Whatever the classification of their products and substrates, group transfer enzymes offer a rich and diverse set of options for synthetic biologists. The general classes of these modifications are highlighted in **Table 4-2**, and specific examples of each are described in detail in **Table 4-3**.

Table 4-2. Biological protecting groups in nature.

Modification	Sources	Donor	Features
Glycosylation	Plants, Animals, Bacteria	UDP-Sugar NDP-Sugar	Increased water solubility Stable Cost to charge: 2 ATP equivalents Sugar group options: glucose, galactose, rhamnose, rutinose, glucuronate, xylose…
Phosphorylation	Insects, Bacteria	ATP	Increased water solubility Stable Cost to charge: 1 ATP equivalent
Sulfonation	Animals	PAPS	Increased water solubility Stable Cost to charge: 3 ATP equivalents
Acylation	Animals, Bacteria	Acyl-CoA	Decreased water solubility Enhanced membrane permeance Acid/base labile Cost to charge: 2 ATP equivalents Acyl group options: acetate, malonate, butyrate, coumarate…

Table 4-3. Examples of natural products containing biological protecting groups.

Group	Product	Organism	Purpose	Ref.
Glucose	Indican	*P. tinctorium* *I. tinctoria* *I. suffruticosa* (plants)	Stable, soluble storage form of the unstable indigo precursor, indoxyl	189
Glucose	Betanin	*B. vulgaris* (plant)	Stable form of the pigment betanidin	223
Glucose	Geraniol glucoside	*V. vinifera* (plant)	Soluble storage form of volatile aroma molecule	224

Glucose	Arbutin	*R. serpentina* (plant) *M. sexta* (insect)	Storage of toxic hydroquinone to deter predation (plant) Detoxification of toxic hydroquinone to confound plant defenses (insect)	210,225
Glucose	Deoxynivalenol	*H. vulgare* (plant)	Xenobiotic metabolism, inactivation of pathogen toxin	226
Glucose	Oleandomycin glucoside	*S. antibioticus* (bacterium)	Inactivation of antibiotic in the producing host organism	227,228
Glucose	Tyrosine-O-Glucoside	*M. sexta* (insect)	Soluble storage of tyrosine for development, separation of tyrosine from toxic phenols	229
Glucuronic Acid	Morphine-6-O-Glucuronide	*H. sapiens* (mammal)	Active metabolite of morphine, high analgesic activity	85
Glucuronic Acid	Morphine-3-O-Glucuronide	*H. sapiens* (mammal)	Inactive metabolite of morphine, no analgesic effect	85
Glucuronic Acid	Indoxyl glucuronide	*H. sapiens* (mammal)	Xenobiotic metabolism of indigo precursor, enhances excretion	230
Sulfate	Indoxyl sulfate	*H. sapiens* (mammal)	Xenobiotic metabolism of indigo precursor, enhances excretion	231
Sulfate	Tyrindoxyl sulfate	*D. orbita* (mollusk)	Stable, soluble storage form of tyrian purple precursor	232
Phosphate	Tyrosine-O-phosphate	*D. melanogaster* (insect)	Soluble storage of tyrosine for development	233
Acetate	Chloramphenicol acetate	*E. coli* (bacterium)	Antibiotic inactivation / resistance	234
Acetate	N-acetyl benzidine	*H. sapiens* (mammal)	Xenobiotic metabolism of carcinogens	235
Acetate	13-O-acetyl-1,13-dihydroxy-N-methylcanadine	*P. somniferum* (plant)	Protecting group intermediate in noscapine biosynthesis, used to control reactivity of the 13-hydroxy group	236
Malonate	2-naphthol malonylglucoside	*A. thaliana* (plant)	Tagging xenobiotics with negative charge for sequestration in the vacuole	237

With such a fantastic selection of biological protecting groups, group transfer enzymes, and potential substrates, how would one get started in this field? For synthetic biologists, making a toolkit and a chassis are the first steps. The indigo pathway provides a convenient test case around which these activities can be organized: For broader impact, indoxyl is an unstable molecule that needs to be protected to be useful in reworking the dye process to remove reducing agents. Producing large quantities of protected indoxyl along with a method for deprotecting that indoxyl is valuable as a result. For the practical concerns, indigo offers a colorimetric readout of what fraction of indoxyl is successfully being protected, and this assay is amenable to a moderately high-throughput plate-based format. Also, a wide variety of protected indoxyls have already been observed in nature, including indoxyl sulfate[231], indoxyl glucuronide[230], and indoxyl glucoside (indican), suggesting that the transferases that make these molecules must be available from the source organism. Finally, since 5-bromo-4-chloro-indoxyl (BCI) has become the *de facto* standard for hydrolase assays since its discovery[238,239], six of the seven protected indoxyls shown in **Figure 4-12** are available for purchase. Each of these protecting groups has its own chemical characteristics (**Table 4-4**), advantages, and drawbacks and thus has its place in a complete toolbox.

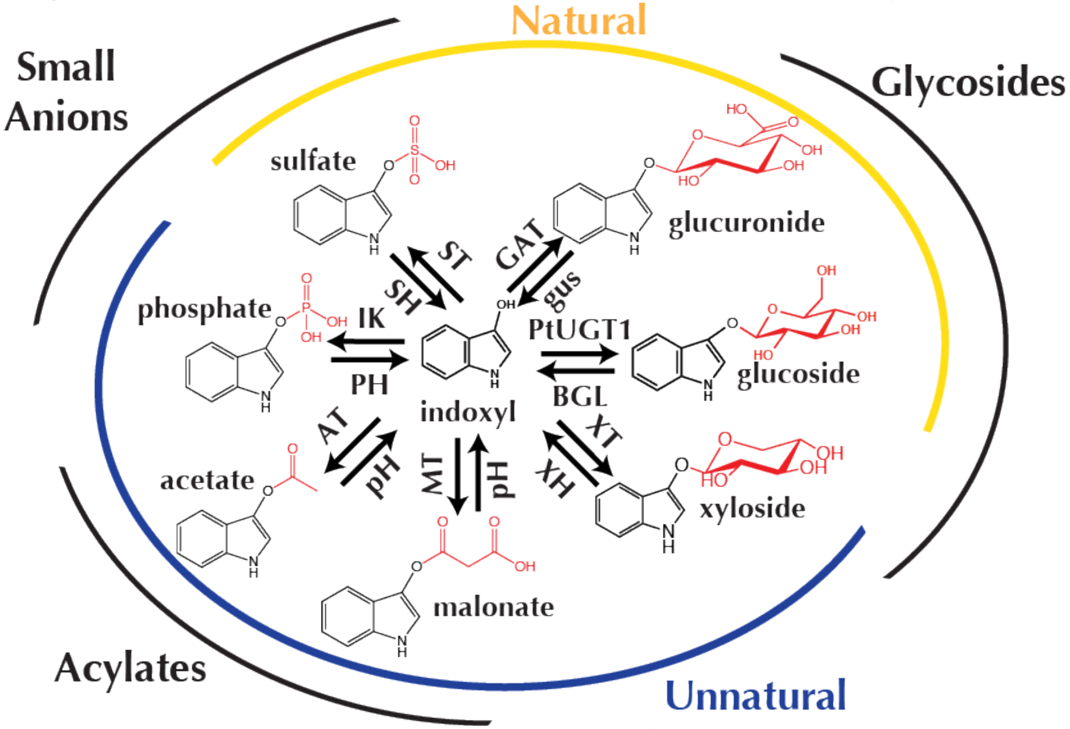


Figure 4-12. A rich collection of protecting groups for indoxyl. Protected indoxyls seen in nature (sulfate, glucuronide, and glucoside) are highlighted with the yellow arc. The blue arc highlights unnatural protected indoxyls. Of all of the indoxyls shown, only indoxyl malonate is not commercially available. Transferases: Pt_UGT1, GAT (glucuronic acid transferase), XT (xylosyltransferase), MT (malonyltransferase), AT (acetyltransferase), IK (indoxyl kinase), ST (sulfotransferase). Hydrolysis: acetate and malonate can likely be removed with base (pH) or an esterase; other protecting groups require the appropriate hydrolase such as gus (β-glucuronidase), BGL (β-glucosidase), XH (xylosidase), PH (phosphatase), SH (sulfohydrolase).

Table 4-4. Characteristics of different protected indoxyls. LogD (taken from http://chemicalize.org) is similar to partition coefficient LogP, but estimated at pH 7.4 to be more physiologically relevant. A high LogD suggests low water solubility. BOD is theoretical biochemical oxygen demand per mol of protecting group released.

Indoxyl-	**LogD**	**BOD (vs hydrosulfite 1.5 mol O_2 / mol)**
Glucoside	-0.48	6 mol O_2/ mol
Glucuronide	-3.31	5 mol O_2/ mol
Xyloside	0.16	5 mol O_2/ mol
Sulfate	-1.02	N/D
Phosphate	-2.11	N/D
Acetate	1.68	2 mol O_2/ mol
Malonate	-1.41	2 mol O_2/ mol

4.4.2. Stability of different protecting groups in hosts

One of the key questions when picking a protecting group is: will the protecting group remain stable in the chassis of interest? And if not, is there a way to solve the problem? To investigate these questions, we took several of the protected indoxyls from **Figure 4-12** and exposed them to either yeast or *E. coli* lysate (**Figure 4-13**). We found that indoxyl sulfate is generally stable in either *E. coli* or yeast, and indoxyl galactoside is conditionally stable in *E. coli* (considering that IPTG induction is used to stimulate consumption of X-gal and probably I-gal) and possibly always stable in yeast. Indoxyl glucoside may also be stable in yeast lysate. However, indoxyl phosphate, acetate, and glucuronide appear to be unstable in both yeast and *E. coli*. Our next question: is there a way to solve the stability problem?

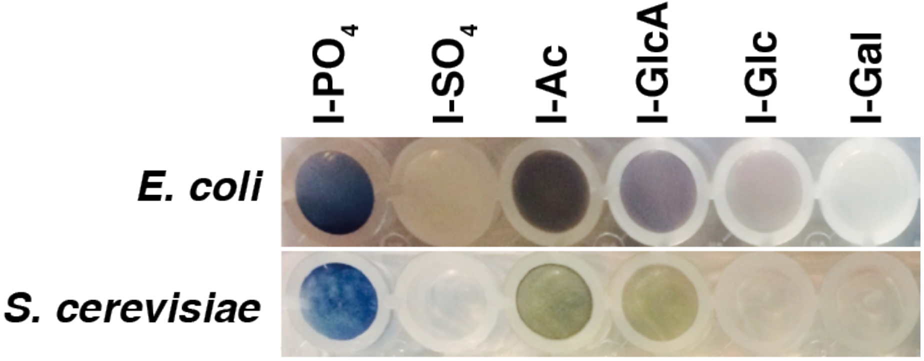


Figure 4-13. Stability of protected indoxyls in *E. coli* and yeast. Lysate from sonication (*E. coli*) or zymolyase treatment and sonication (*S. cerevisiae*) was treated with one of six different protected indoxyls (all colorless to start) for three days at 30°C. Any development of color suggests hydrolysis of the protected indoxyl. I-PO4 is indoxyl phosphate, I-SO4 is indoxyl sulfate, I-Ac is indoxyl acetate, I-GlcA is indoxyl glucuronide, I-Glc is indican, and I-Gal is indoxyl galactoside.

With the development of improved CRISPR tools[118], engineering knockouts in yeast has become almost trivial for most genes. We knocked out five esterases individually and tested their effect on hydrolysis of indoxyl acetate and fluorescein diacetate (**Figure 4-14**). We found that YJL068CΔ and IAH1Δ were the most important, and that YJL068C preferred fluorescein diacetate as a substrate while IAH1 favored indoxyl acetate. Though there were significant improvements in stability, more combinatorial knockouts must be made to drop the esterase activity even further before yeast is a suitable chassis for acetylating products.

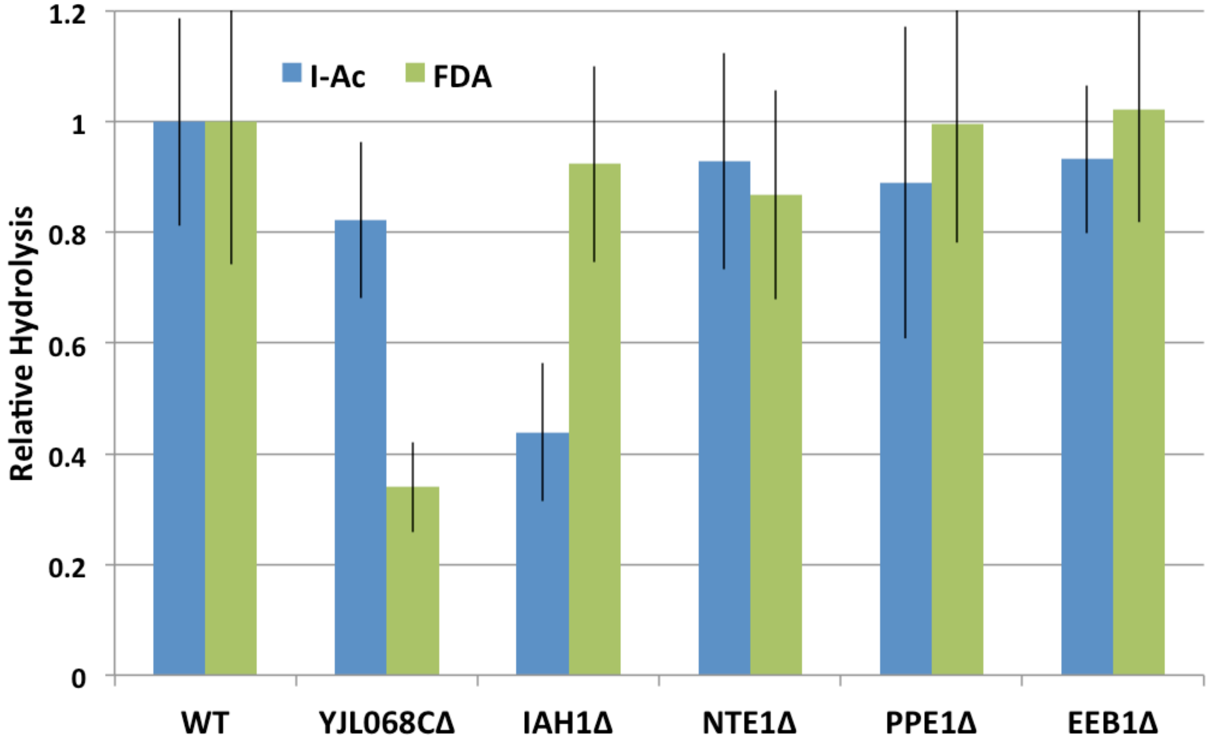


Figure 4-14. YJL068C and IAH1 appear to be the most important esterases for acetylated product stability in yeast. Wildtype (WT) yeast or those with whole-ORF deletions of the indicated esterase were grown to saturation and resuspended in PBS with either indoxyl acetate (blue) or fluorescein diacetate (green). After incubation at room temperature for several minutes, fluorescence was measured and hydrolysis products were detected. Error bars are the mean ± s.d. of six biological replicates.

4.4.3. Indoxyl sulfate as another option for dyeing

One protecting group that showed no hydrolysis in either yeast or *E. coli* was sulfate. This appeared to be very promising, as no strain engineering would be necessary for us to start producing indoxyl sulfate. Also, the sulfotransferases responsible for indoxyl sulfonation are known[231], and indoxyl sulfate has a special role in the human body as a waste product produced from indole received from the gut. It is responsible for the "purple urine bag syndrome" seen in infections of indwelling catheters when bacteria hydrolyze the excreted indoxyl sulfate to produce indigo[240]. We tested several sulfatases and found that one, AtsA from *Pseudomonas aeruginosa*[240], performs exceptionally well at hydrolyzing indoxyl sulfate (**Figure 4-15**). With the sulfatase confirmed, we moved to confirming the sulfotransferases, testing wild-type, feedback-resistant (D249G)[241], and enhanced-activity (Q56E, L67V, A101S, T117S, F222L, V243I, D249G)[241] mutants of human SULT1A1 and SULT1A3. We observed excellent conversion of indoxyl to indoxyl sulfatewhen PAPS was added to the reaction (**Figure 4-16**). However, we found that indoxyl sulfate conversion was strongly dependent on exogenously added PAPS, so engineering to increase the availability and recycling of this cofactor will be important for the high-titer production of indoxyl sulfate and high value products like heparan sulfate[242] and glucosinolates[243].

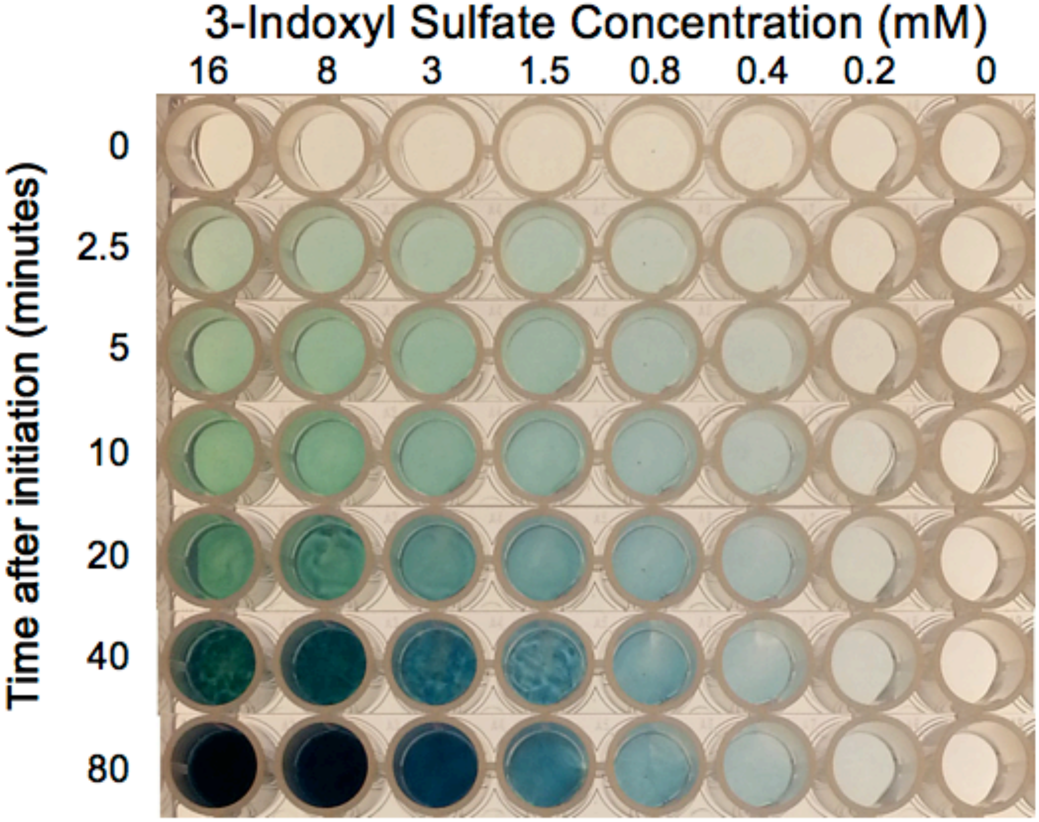


Figure 4-15. Sulfatase AtsA from *Pseudomonas aeruginosa* rapidly hydrolyzes indoxyl sulfate. Indoxyl sulfate was added to purified AtsA at time 0. Moving down, each row is an image of the same wells at a later time point. Blue color started forming almost instantly upon mixing indoxyl sulfate with AtsA.

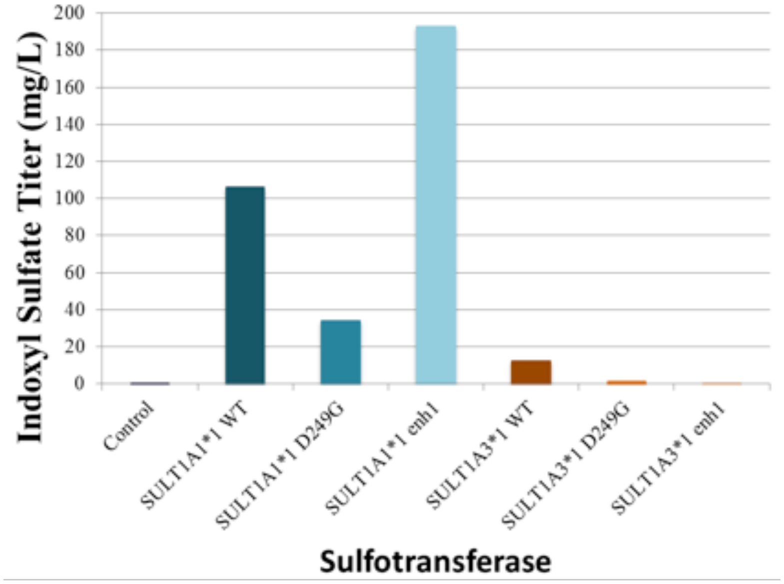


Figure 4-16. SULT1A1 enhanced can sulfonate sizeable amounts of indoxyl when PAPS is supplied. Cells expressing a human SULT were fed 2 mM PAPS and 2 mM indoxyl acetate under anaerobic conditions. The indoxyl acetate was converted to indoxyl and then protected by the SULT. SULT1A1 and SULT1A3 are different isoforms, while D249G is a feedback-resistant mutant and enh1 is a multiple mutant (Q56E/L67V/A101S/T117S/F222L/V243I/D249G) designed to generally enhance activity.

4.4.4. Extension to new colors

A final quirk we happened upon while investigating the indigo pathway was that, despite testing several substrates including halogenated indoles, none of the literature[244–248] we read examining indole oxidases tested 6-bromoindole, the precursor to Tyrian purple (6,6′-dibromoindigo). We added 5-, 6-, and 7- chloro- and bromo- indole to bacteria (with FMO) and observed substituted indigos including Tyrian purple (**Figure 4-17**).

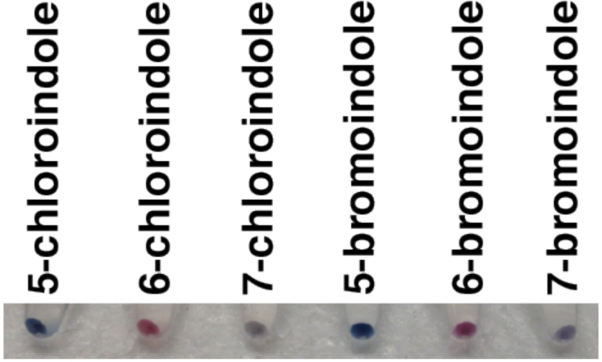


Figure 4-17. Production of Tyrian purple and other substituted indigos in *E. coli*. Cells expressing FMO were grown to saturation, washed, and placed in PBS with 1 mM of the indicated indole and incubated for several hours. 6-bromoindole should give rise to Tyrian purple (a magenta pellet).

4.4.5. Use of biological protecting groups in organic synthesis

Perhaps the most exciting use of biological protecting groups is to enable new and creative organic syntheses. Since our protecting group transferases are largely thought to be promiscuous, we suspected they could be used to modify new and unexpected substrates without any additional mutagenesis. In this case, the new substrate (**Figure 4-18**) was a *meso* compound that needed to be desymmetrized by modifying just one of the two hydroxyl groups. This would allow further syntheses, and an enantiospecific solution is preferred. We tested several glucosyltransferases against this substrate and, much to our excitement, Pt_UGT2 and OleD[227] both worked, with Pt_UGT2 outperforming alternative methods (**Table 4-5**).

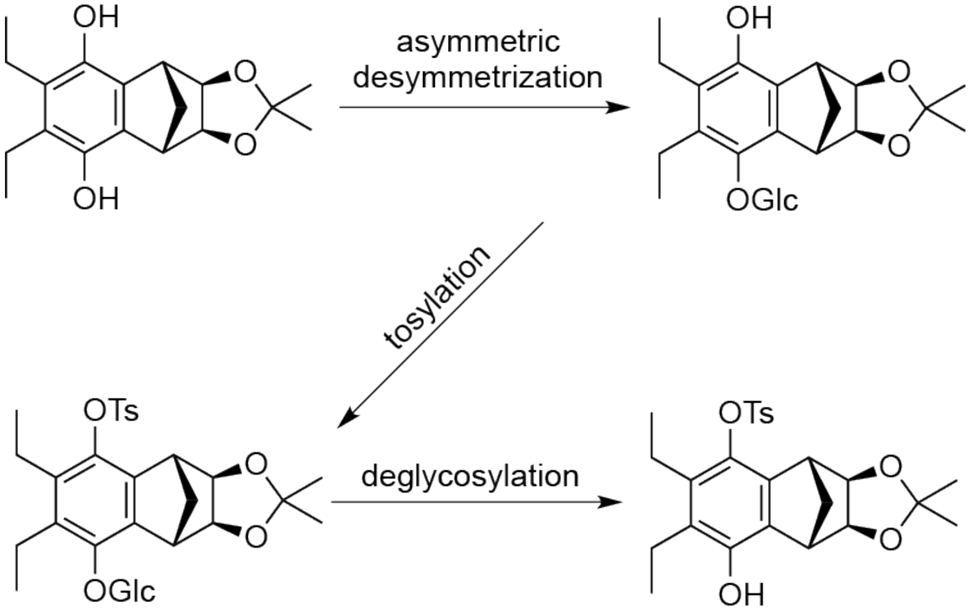
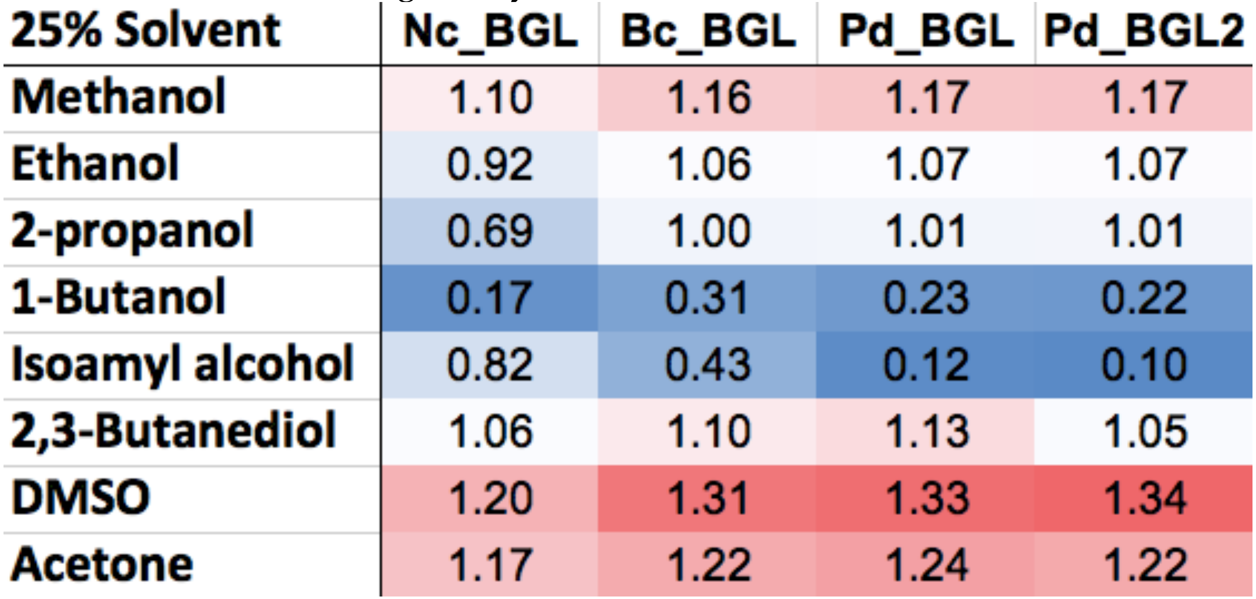


Figure 4-18. Desymmetrization of a *meso* precursor of herbindole. Desymmetrization can be achieved by glucosylation of just one of the free hydroxyls in the starting material. Once desymmetrized, the remaining hydroxyl must be tosylated, and then the original glucose removed.

Table 4-5. Protection of a *meso* compound for desymmetrization. DR indicates diasteromeric ratio.

Strategy	Yield	DR
racemic triflation	60%	1:1
Miller peptide-catalyzed acetylation	65%	4.7:1
superstoichiometric chiral isothiourea	42%	4.3:1
Pt_UGT2	57%	15:1
OleD	5%	1:2

Once the mono-glucosylated product was tosylated (**Figure 4-19**), it lost most of its solubility in water, and we were forced to re-evaluate our options for removing the glucosyl group. We had previously only used our BGLs in water, and did not expect much activity in high levels of organic solvent, although there is some precedent solvent tolerance in enzymes[249]. We tested four BGLs in high-solvent conditions and found, surprisingly, that they retained most of their activity even up to 25% solvent (**Figure 4-19**)! However, when we tested the BGLs in 25% solvent with the tosylated, glucosylated product, we did not observe any activity. Even so, the individual parts of this toolbox have been shown to be functional and may be useful for the next organic synthesis.



25% Solvent	Nc_BGL	Bc_BGL	Pd_BGL	Pd_BGL2
Methanol	1.10	1.16	1.17	1.17
Ethanol	0.92	1.06	1.07	1.07
2-propanol	0.69	1.00	1.01	1.01
1-Butanol	0.17	0.31	0.23	0.22
Isoamyl alcohol	0.82	0.43	0.12	0.10
2,3-Butanediol	1.06	1.10	1.13	1.05
DMSO	1.20	1.31	1.33	1.34
Acetone	1.17	1.22	1.24	1.22

Figure 4-19. Activities of BGLs remain in 25% organic solvent. Purified *Neurospora crassa* GH-1 (Nc_BGL) or *Bacillus circulans* BglA (Bc_BGL) or purchased *Prunus dulcis* BGL from supplier 1 (Pd_BGL) or supplier 2 (Pd_BGL2) were tested in 25% of the indicated solvent to determine whether activity remained. Numbers shown are conversion of substrate (4-methylumbelliferyl glucoside) into product (4-methylumbelliferone) after 5 hours of incubation, relative to no-solvent controls. Numbers exceeding 1 are likely due to the volume of 25 μL solvent + 75 μL water being less than 100 μL, concentrating and increasing the apparent signal to allow greater than 100% conversion.

4.4.6. Synergy with benzylisoquinoline alkaloids

Please refer to **Chapter 2** to see how protecting groups can enhance benzylisoquinoline alkaloid biosynthesis, stability, and enable new abuse-resistant bioproducts and organic syntheses.

4.4.7. Synergy with peroxisomal compartmentalization

Please refer to **Chapter 3** to see how biological protecting groups can be used in the peroxisome, both to provide a separate pool of donors such as acyl-CoA, UDP-sugars, or PAPS, as well as to trap intermediates and products in the peroxisome.

4.5. Materials and Methods

4.5.1. Plant growth and harvesting

P. tinctorium plants (from Companion Plants, Athens, Ohio) were grown indoors in a greenhouse or laboratory environment. Samples of leaf tissue were taken from live plants at several points during the day and frozen in liquid nitrogen. These samples were mixed, kept frozen with liquid nitrogen, and crushed using 5 mm diameter steel beads in a beadbeater at 30 Hz for 2 min. Total RNA was extracted from the 100 mg of powdered frozen leaf tissue using the Qiagen RNEasy Plant Mini kit obtained from Qiagen GMBH (Hilden, Germany), in accordance with the manufacturer recommended protocol. The UC Berkeley Functional Genomics Laboratory completed the rest of the cDNA preparation as follows: The mRNA was extracted from the total RNA using magnetic beads coated with oligo $(dT)_{25}$. The mRNA was then sheared to approximately 550 base pairs in length using a Covaris S2 ultrasonicator from Covaris Inc. (Woburn, MA). A cDNA library was generated using the Apollo 324 Next-Gen Library Prep System from Wafergen Biosystems Inc. (Fremont, CA) using the manufacturer-supplied PrepX RNA-Seq Library Preparation Kit. The cDNA library was then clustered using the cBot from Illumina Inc. (San Diego, CA) and the clustered sample was loaded onto an Illumina HiSeq2500 courtesy of the UC Berkeley Vincent J. Coates Genomics Sequencing Laboratory and sequenced using the Rapid Run reagent kit for 150 base, paired-end reads.

4.5.2. Transcriptome sequencing and assembly

Paired-end reads received from the Illumina HiSeq2500 sequencer were first trimmed to remove low-quality reads using the Trimmomatic software package[250] in paired-end mode to remove Illumina adapter sequences and using a sliding quality window of 30 or greater, where reads with under 36 acceptable bases are dropped. Overlapping paired-end reads were then merged using the FLASH software package[251] with a minimum overlap size of 15 bases and an expected fragment length of 350 bases. The remaining merged and unmerged reads were pooled and digitally normalized to remove redundant data using the khmer software package[252] with options set to paired-end, k-mer size of 19, culling count of 20, and 4 hash tables of 4 GiB each. Following digital normalization, reads were assembled into transcript scaffolds using the Trinity RNA-seq assembly package in paired-end mode[206,207] or OASES using k-mer sizes between 17 and 31[208]. Scaffolds were then annotated using BLASTX against a library of known plant glycosyltransferases retrieved from Uniprot with an E-value threshold of 1e-60 to identify glucosyltransferase candidates. These candidates were then translated into peptides using the Trinity package, Transdecoder, using the default settings[206,207]. This work was done on the XSEDE Blacklight system[253,254].

4.5.3. Isolation of indican synthase and cDNA identification

The predicted UDP-glucose:glucosyltransferase was purified from *P. tinctorium* following the protocol of Minami et al[189]. 200 grams of fresh leaves were flash frozen with liquid nitrogen and ground into a fine powder using a mortar and pestle. The fine powder was suspended in 400 mL of extraction buffer (2 mM EDTA, 20 mM β-mercaptoethanol, 100 mM H_2KPO_4, 1X cOmplete protease inhibitor, pH 7.0 in water) and centrifuged at 24500 rcf for 30 min while refrigerated at 4°C. The supernatant was transferred to a new tube and precipitated by adding 600 mL of extraction buffer with 50% w/v polyethylene glycol (M_w 6000 Da) to make 1 L total solution with 30% w/v PEG 6000. This solution was centrifuged at 143000 rcf for 30 min at 4°C. The supernatant of this process was mixed with 25 mL DEAE-Sepharose beads. The beads were washed with 125 mL of wash buffer A (1 mM EDTA, 10 mM β-mercaptoethanol, 50 mM HEPES-NaOH pH 7.0), then washed with 50 mL of wash buffer A + 50 mM NaCl, and then protein was eluted with 50 mL of wash buffer A + 100 mM NaCl. This 50 mL elution was loaded on a 10 mL hydroxyapatite column, which was washed with 50 mL buffer A1 (1 mM EDTA, 5 mM dithiothreitol, 10% v/v glycerol, 50 mM HEPES-NaOH pH 7.0) and eluted with 98% A1 / 2% B1 where buffer B1 is buffer A1 + 500 mM H_2KPO_4 pH 7.0. This eluate was loaded onto a Mono Q 5/50 GL column and washed with 5 column volumes of buffer A2 (1 mM EDTA, 5 mM dithiothreitol, 10% v/v glycerol, 10 mM H_2KPO_4, 50 mM HEPES-NaOH pH 7.0). Protein was eluted with a gradient from 0-100% Buffer B2 (Buffer A2 + 200 mM NaCl), at a 1 mL per minute flow rate over 30 min. The protein eluate from 5-9 min was collected and concentrated tenfold using a Millipore 10 kDa molecular weight-cutoff spin column. This concentrate was run on a Superdex 200 size-exclusion column, eluting with 5 mM dithiothreitol, 50 mM HEPES-NaOH pH 7.0. Fractions were lyophilized, resuspended in water, dialyzed in water, and then separated using SDS-PAGE. Proteins trapped in the polyacrylamide gel were electroblotted onto a PVDF membrane and extracted for trypsin digestion and column separation followed by tandem mass spectrometry carried out by the UC Berkeley Vincent J. Coates Proteomics/Mass Spectrometry Laboratory to identify protein fragments. These protein fragments were correlated to the Transdecoder-predicted sequences using DTAselect[255] and SEQUEST[256]. The sequences with the most complete coverage of matching peptide fragments were selected for further study.

Another sample of total RNA was prepared as described above, but was converted to adaptor-tagged cDNA using the GeneRacer® Kit with SuperScript® III RT and TOPO TA Cloning® Kit for Sequencing from Life Technologies (Carlsbad, CA). Using gene specific oligonucleotide primers, a polymerase chain reaction was used on the adaptor-tagged cDNA library to clone the sequences of interest identified previously. These genes were then Sanger sequenced to confirm their identity and nucleotide sequence.

4.5.4. β-glucosidase expression and purification

BL21 DE3 *E. coli* cells were transformed with recombinant pET vectors for inducible expression of N-terminally hexahistidine-tagged β-glucosidase from *Bacillus circulans* (Uniprot accession: Q03506)[201,202]. A 2 L flask containing 0.5 L LB media supplemented with 100 μg/mL ampicillin was inoculated with the *E. coli* by the addition of 5 mL of saturated culture, a 1:100 dilution. This was kept shaking at 250 rpm and 37 °C until the *E. coli* had grown to an OD600 of about 0.5. Once grown, the culture was induced to produce protein by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. The flask was then allowed to continue shaking at 250 rpm and 37 °C for 16 h. Once incubation was complete, cells were harvested by centrifugation at 4800 rcf for 5 min. The supernatant was discarded and the cells were resuspended in 35 mL of ice-cold lysis buffer containing 150 mM NaCl, 25 mM imidazole, and 25 mM Tris pH 8.0 in water. Cells were kept at 4 °C and sonicated for 30 sec at maximum amplitude to achieve lysis. Lysate was centrifuged at 24500 rcf for 10 min, and the supernatant mixed with 4 mL nickel-nitriloacetic acid agarose beads for 10 min. The beads and supernatant were packed into a column and allowed to flow through by gravity. The column was washed twice with 30 mL of fresh lysis buffer to remove any unbound protein, and the hydrolase was eluted off of the column by running 10 mL of elution buffer (150 mM NaCl, 500 mM imidazole, and 25 mM Tris pH 8.0 in water) through the column. The eluate was concentrated using a Millipore 10 kDa molecular weight-cutoff spin column and glycerol was added to a final concentration of 10% v/v, then aliquots of the hydrolase were frozen at -80 °C.

4.5.5. *In vivo* Production of Indican in *E. coli* Expressing Pt_UGT1

E. coli BW25113 were transformed with a plasmid with expression cassettes for both the flavin monooxygenase (FMO) and Pt_UGT1 from *P. tinctorium*. A 24-well culture block containing 3 mL per well of EZ-Rich MOPS culture medium obtained from Teknova (Hollister, CA) and supplemented with 50 μg/mL spectinomycin, 5 mM L-tryptophan, and 2% w/v glucose was inoculated with these *E. coli* directly from colonies. This culture was incubated at 37 °C with shaking at 750 rpm for 36 h.

After incubation, indican was extracted by taking 360 μL of each reaction and adding 40 μL of 50% w/v aqueous NH_4OH to achieve a final volume of 400 μL supplemented with 5% w/v NH_4OH. This mixture was allowed to sit for 5 minutes. Samples were vortexed with 800 μL of methyl *tert*-butyl ether (MTBE) for 10 min, centrifuged at 20000 rcf for 5 min, and the aqueous phase separated by LC/MS on a C18 column. Indican was detected via time-of-flight mass spectrometry using a negative electrospray ion source.

4.5.6. Dyeing cloth with indican and BGL

Culture media from indican-producing bacteria was concentrated by boiling for several hours in a 100 mL beaker. A bear-shaped piece of undyed cotton cloth was then added to the beaker containing the indican solution. The cloth was allowed to soak in the liquid for 10 min at ambient temperature. The cloth pieces were removed from the solution and laid flat on wax paper. Then, 100 μL of a 4 mg/mL β-glucosidase solution was pipetted unevenly across each cloth piece. The cloth pieces were incubated at ambient temperature for 6 h, then washed repeatedly with water and allowed to dry.

4.6. Indigo Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE)[253], which is supported by National Science Foundation award number ACI-1053575. Specifically, it used the Blacklight system, which is supported by NSF award number ACI-1041726, at the Pittsburgh Supercomputing Center (PSC)[254]. The specific XSEDE grant was TG-MCB140082. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303. This work used the Vincent J. Proteomics/Mass Spectrometry Laboratory at UC Berkeley, supported in part by NIH S10 Instrumentation Grant S10RR025622. We would also like to acknowledge the NDSEG, Siebel Scholars Program, and the Bakar Fellowship for supporting this work.

Chapter 5. References

1. Clark, J. H. Green chemistry: challenges and opportunities. *Green Chem.* **1,** 1–8 (1999).
2. Clark, J. H. *et al.* Green chemistry and the biorefinery: a partnership for a sustainable future. *Green Chem.* **8,** 853–860 (2006).
3. Abelshauser, W., Von Hippel, W., Johnson, J. A. & Stokes, R. G. *German Industry and Global Enterprise. BASF: the History of a Company*. (Cambridge University Press, 2004).
4. de Meijere, A. Adolf von Baeyer: Winner of the Nobel Prize for Chemistry 1905. *Angew. Chem. Int. Ed.* **44,** 7836–7840 (2005).
5. Steingruber, E. in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH Verlag GmbH & Co. KGaA, 2000).
6. Wainwright, M. Dyes in the development of drugs and pharmaceuticals. *Dyes Pigments* **76,** 582–589 (2008).
7. Ensley, B. D. Microbial production of indigo. (1985).
8. Ensley, B. D. *et al.* Expression of naphthalene oxidation genes in Escherichia coli results in the biosynthesis of indigo. *Science* **222,** 167–169 (1983).
9. DeLoache, W. C. *et al.* An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* **11,** 465–471 (2015).
10. DeLoache, W. C., Russ, Z. N. & Dueber, J. E. Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nat. Commun.* **7,** 11152 (2016).
11. Leonard, E., Runguphan, W., O'Connor, S. & Prather, K. J. Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nat. Chem. Biol.* **5,** 292–300 (2009).
12. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **4,** 206–220 (2005).
13. Glenn, W. S., Runguphan, W. & O'Connor, S. E. Recent progress in the metabolic engineering of alkaloids in plant systems. *Curr. Opin. Biotechnol.* **24,** 354–365 (2013).
14. Paterson, I. & Anderson, E. A. The Renaissance of Natural Products as Drug Candidates. *Science* **310,** 451–453 (2005).
15. Mora-Pale, M., Sanchez-Rodriguez, S. P., Linhardt, R. J., Dordick, J. S. & Koffas, M. A. Biochemical strategies for enhancing the in vivo production of natural products with pharmaceutical potential. *Curr. Opin. Biotechnol.* **25,** 86–94 (2014).
16. Facchini, P. J. *et al.* Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* **30,** 127–131 (2012).

17. Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7,** 114 (2006).
18. Siddiqui, M. S., Thodey, K., Trenchard, I. & Smolke, C. D. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *FEMS Yeast Res.* **12,** 144–170 (2012).
19. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2013).
20. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. *Plant Cell Physiol.* **54,** 647–672 (2013).
21. Beaudoin, G. A. W. & Facchini, P. J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240,** 19–32 (2014).
22. Nakagawa, A. *et al.* A bacterial platform for fermentative production of plant alkaloids. *Nat. Commun.* **2,** 326 (2011).
23. Nakagawa, A. *et al.* (R,S)-Tetrahydropapaveroline production by stepwise fermentation using engineered Escherichia coli. *Sci. Rep.* **4,** (2014).
24. Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **4,** 564–573 (2008).
25. Fossati, E. *et al.* Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in Saccharomyces cerevisiae. *Nat. Commun.* **5,** 3283 (2014).
26. Thodey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10,** 837–844 (2014).
27. Fossati, E., Narcross, L., Ekins, A., Falgueyret, J.-P. & Martin, V. J. J. Synthesis of Morphinan Alkaloids in Saccharomyces cerevisiae. *PLoS ONE* **10,** e0124459 (2015).
28. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci.* **111,** E2149–E2156 (2014).
29. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33,** 377–383 (2015).
30. Lichman, B. R. *et al.* 'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile. *FEBS J.* **282,** 1137–1151 (2015).
31. Fitzpatrick, P. F. Tetrahydropterin-Dependent Amino Acid Hydroxylases. *Annu. Rev. Biochem.* **68,** 355–381 (1999).
32. Claus, H. & Decker, H. Bacterial tyrosinases. *Syst. Appl. Microbiol.* **29,** 3–14 (2006).

33. Halaouli, S., Asther, M., Sigoillot, J.-C., Hamdi, M. & Lomascolo, A. Fungal tyrosinases: new prospects in molecular characteristics, bioengineering and biotechnological applications. *J. Appl. Microbiol.* **100,** 219–232 (2006).
34. Gandía-Herrero, F., García-Carmona, F. & Escribano, J. Botany: Floral fluorescence effect. *Nature* **437,** 334–334 (2005).
35. Sasaki, N. *et al.* Detection of DOPA 4,5-Dioxygenase (DOD) Activity Using Recombinant Protein Prepared from Escherichia coli Cells Harboring cDNA Encoding DOD from Mirabilis jalapa. *Plant Cell Physiol.* **50,** 1012–1016 (2009).
36. Gandía-Herrero, F. & García-Carmona, F. Biosynthesis of betalains: yellow and violet plant pigments. *Trends Plant Sci.* **18,** 334–343 (2013).
37. Gandía-Herrero, F., García-Carmona, F. & Escribano, J. A novel method using high-performance liquid chromatography with fluorescence detection for the determination of betaxanthins. *J. Chromatogr. A* **1078,** 83–89 (2005).
38. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4,** 67–73 (1999).
39. Santos, C. N. S. & Stephanopoulos, G. Melanin-Based High-Throughput Screen for l-Tyrosine Production in Escherichia coli. *Appl. Environ. Microbiol.* **74,** 1190–1197 (2008).
40. Lezzi, C., Bleve, G., Spagnolo, S., Perrotta, C. & Grieco, F. Production of recombinant Agaricus bisporus tyrosinase in Saccharomyces cerevisiae cells. *J. Ind. Microbiol. Biotechnol.* **39,** 1875–1880 (2012).
41. Hernández-Romero, D., Sanchez-Amat, A. & Solano, F. A tyrosinase with an abnormally high tyrosine hydroxylase/dopa oxidase ratio. *FEBS J.* **273,** 257–270 (2006).
42. Hatlestad, G. J. *et al.* The beet R locus encodes a new cytochrome P450 required for red betalain production. *Nat. Genet.* **44,** 816–820 (2012).
43. Gandía-Herrero, F. & García-Carmona, F. Characterization of recombinant Beta vulgaris 4,5-DOPA-extradiol-dioxygenase active in the biosynthesis of betalains. *Planta* **236,** 91–100 (2012).
44. Koyanagi, T. *et al.* Eukaryotic-type aromatic amino acid decarboxylase from the root colonizer Pseudomonas putida is highly specific for 3,4-dihydroxyphenyl-l-alanine, an allelochemical in the rhizosphere. *Microbiology* **158,** 2965–2974 (2012).
45. Luttik, M. A. H. *et al.* Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: Quantification of metabolic impact. *Metab. Eng.* **10,** 141–153 (2008).

46. Gandía-Herrero, F., Escribano, J. & García-Carmona, F. Characterization of the Activity of Tyrosinase on Betanidin. *J. Agric. Food Chem.* **55,** 1546–1551 (2007).
47. Dohm, J. C. *et al.* The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* **505,** 546–549 (2014).
48. Pieper, U. *et al.* ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **42,** D336–D346 (2014).
49. Sentheshanmuganathan, S. & Elsden, S. R. The mechanism of the formation of tyrosol by Saccharomyces cerevisiae. *Biochem. J.* **69,** 210–218 (1958).
50. Pauli, H. H. & Kutchan, T. M. Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. *Plant J.* **13,** 793–801 (1998).
51. Xiao, M. *et al.* Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* **166,** 122–134 (2013).
52. Maresh, J. J. *et al.* Facile one-pot synthesis of tetrahydroisoquinolines from amino acids via hypochlorite-mediated decarboxylation and Pictet–Spengler condensation. *Tetrahedron Lett.* **55,** 5047–5051 (2014).
53. Minami, H., Dubouzet, E., Iwasa, K. & Sato, F. Functional Analysis of Norcoclaurine Synthase in Coptis japonica. *J. Biol. Chem.* **282,** 6274–6282 (2007).
54. Hazelwood, L. A., Daran, J.-M., Maris, A. J. A. van, Pronk, J. T. & Dickinson, J. R. The Ehrlich Pathway for Fusel Alcohol Production: a Century of Research on Saccharomyces cerevisiae Metabolism. *Appl. Environ. Microbiol.* **74,** 2259–2266 (2008).
55. Polturak, G. *et al.* Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants. *New Phytol.* **210,** 269–283 (2016).
56. Sunnadeniya, R. *et al.* Tyrosine Hydroxylation in Betalain Pigment Biosynthesis Is Performed by Cytochrome P450 Enzymes in Beets (Beta vulgaris). *PLOS ONE* **11,** e0149417 (2016).
57. Lamb, D. C., Kelly, D. E., Manning, N. J., Kaderbhai, M. A. & Kelly, S. L. Biodiversity of the P450 catalytic cycle: yeast cytochrome b5/NADH cytochrome b5 reductase complex efficiently drives the entire sterol 14-demethylation (CYP51) reaction. *FEBS Lett.* **462,** 283–288 (1999).

58. Kandel, S. E. & Lampe, J. N. Role of Protein–Protein Interactions in Cytochrome P450-Mediated Drug Metabolism and Toxicity. *Chem. Res. Toxicol.* **27,** 1474–1486 (2014).
59. Urban, P., Mignotte, C., Kazmaier, M., Delorme, F. & Pompon, D. Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related Arabidopsis thalianaNADPH-Cytochrome P450 Reductases with P450 CYP73A5. *J. Biol. Chem.* **272,** 19176–19186 (1997).
60. Waskell, L., Zhang, H., Im, S.-C. & Razeghifard, R. Cytochrome b5 and NADPH-Cytochrome P450 Reductase Compete for a Binding Site on the Proximal Surface of Cytochrome P450 2B4. *FASEB J.* **21,** A670 (2007).
61. Zhang, H., Hamdane, D., Im, S.-C. & Waskell, L. Cytochrome b5 Inhibits Electron Transfer from NADPH-Cytochrome P450 Reductase to Ferric Cytochrome P450 2B4. *J. Biol. Chem.* **283,** 5217–5225 (2008).
62. Zhang, H., Im, S.-C. & Waskell, L. Cytochrome b5 Increases the Rate of Product Formation by Cytochrome P450 2B4 and Competes with Cytochrome P450 Reductase for a Binding Site on Cytochrome P450 2B4. *J. Biol. Chem.* **282,** 29766–29776 (2007).
63. Yamazaki, H. *et al.* Roles of NADPH-P450 Reductase and Apo- and Holo-Cytochrome b5 on Xenobiotic Oxidations Catalyzed by 12 Recombinant Human Cytochrome P450s Expressed in Membranes of Escherichia coli. *Protein Expr. Purif.* **24,** 329–337 (2002).
64. Lah, L. *et al.* The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification. *Mol. Microbiol.* **81,** 1374–1389 (2011).
65. Lewis, D. F. V. & Hlavica, P. Interactions between redox partners in various cytochrome P450 systems: functional and structural aspects. *Biochim. Biophys. Acta BBA - Bioenerg.* **1460,** 353–374 (2000).
66. Bumpus, N. N. & Hollenberg, P. F. Cross-linking of human cytochrome P450 2B6 to NADPH-cytochrome P450 reductase: Identification of a potential site of interaction. *J. Inorg. Biochem.* **104,** 485–488 (2010).
67. Chun, Y.-J., Shimada, T. & Guengerich, F. P. Construction of a Human Cytochrome P450 1A1:Rat NADPH–Cytochrome P450 Reductase Fusion Protein cDNA and Expression inEscherichia coli,Purification, and Catalytic Properties of the Enzyme in Bacterial Cells and after Purification. *Arch. Biochem. Biophys.* **330,** 48–58 (1996).
68. Leonard, E., Yan, Y. & Koffas, M. A. G. Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in Escherichia coli. *Metab. Eng.* **8,** 172–181 (2006).
69. Pompon, D., Louerat, B., Bronine, A. & Urban, P. in *Methods in Enzymology* (ed. Waterman, E. F. J. and M. R.) **272,** 51–64 (Academic Press, 1996).

70. Wester, M. R. *et al.* The Structure of Human Cytochrome P450 2C9 Complexed with Flurbiprofen at 2.0-Å Resolution. *J. Biol. Chem.* **279,** 35630–35637 (2004).
71. Sueyoshi, T., Park, L. J., Moore, R., Juvonen, R. O. & Negishi, M. Molecular Engineering of Microsomal P450 2a-4 to a Stable, Water-Soluble Enzyme. *Arch. Biochem. Biophys.* **322,** 265–271 (1995).
72. Schoch, G. A., Attias, R., Belghazi, M., Dansette, P. M. & Werck-Reichhart, D. Engineering of a Water-Soluble Plant Cytochrome P450, CYP73A1, and NMR-Based Orientation of Natural and Alternate Substrates in the Active Site. *Plant Physiol.* **133,** 1198–1208 (2003).
73. Lamb, D. C. *et al.* Generation of a Complete, Soluble, and Catalytically Active Sterol 14α-Demethylase−Reductase Complex. *Biochemistry (Mosc.)* **38,** 8733–8738 (1999).
74. Lamb, D. C., Warrilow, A. G. S., Venkateswarlu, K., Kelly, D. E. & Kelly, S. L. Activities and Kinetic Mechanisms of Native and Soluble NADPH–Cytochrome P450 Reductase. *Biochem. Biophys. Res. Commun.* **286,** 48–54 (2001).
75. Abe, K., Kimura, S., Kizawa, R., Anan, F. K. & Sugita, Y. Amino Acid Sequences of Cytochrome b5 from Human, Porcine, and Bovine Erythrocytes and Comparison with Liver Microsomal Cytochrome b5. *J. Biochem. (Tokyo)* **97,** 1659–1668 (1985).
76. Yamamoto, K. *et al.* Probing the Transmembrane Structure and Topology of Microsomal Cytochrome-P450 by Solid-State NMR on Temperature-Resistant Bicelles. *Sci. Rep.* **3,** (2013).
77. Cosme, J. & Johnson, E. F. Engineering Microsomal Cytochrome P450 2C5 to Be a Soluble, Monomeric Enzyme MUTATIONS THAT ALTER AGGREGATION, PHOSPHOLIPID DEPENDENCE OF CATALYSIS, AND MEMBRANE BINDING. *J. Biol. Chem.* **275,** 2545–2553 (2000).
78. Lynge, M. E., Westen, R. van der, Postma, A. & Städler, B. Polydopamine—a nature-inspired polymer coating for biomedical science. *Nanoscale* **3,** 4916–4928 (2011).
79. Sasaki, N. *et al.* Amaranthin in feather cockscombs is synthesized via glucuronylation at the cyclo-DOPA glucoside step in the betacyanin biosynthetic pathway. *J. Plant Res.* **118,** 439–442 (2005).
80. Dajani, R. *et al.* Kinetic properties of human dopamine sulfotransferase (SULT1A3) expressed in prokaryotic and eukaryotic systems: comparison with the recombinant enzyme purified from Escherichia coli. *Protein Expr. Purif.* **16,** 11–18 (1999).
81. Itäaho, K. *et al.* Regioselective sulfonation of dopamine by SULT1A3 in vitro provides a molecular explanation for the preponderance of

dopamine-3-O-sulfate in human blood circulation. *Biochem. Pharmacol.* **74,** 504–510 (2007).

82. Sasaki, N. *et al.* Isolation and Characterization of cDNAs Encoding an Enzyme with Glucosyltransferase Activity for cyclo-DOPA from Four O'clocks and Feather Cockscombs. *Plant Cell Physiol.* **46,** 666–670 (2005).
83. Vogt, T., Grimm, R. & Strack, D. Cloning and expression of a cDNA encoding betanidin 5-O-glucosyltransferase, a betanidin- and flavonoid-specific enzyme with high homology to inducible glucosyltransferases from the Solanaceae. *Plant J.* **19,** 509–519 (1999).
84. Vogt, T. Substrate specificity and sequence analysis define a polyphyletic origin of betanidin 5- and 6-O-glucosyltransferase from Dorotheanthus bellidiformis. *Planta* **214,** 492–495 (2002).
85. Mantione, K. J., Goumon, Y., Esch, T. & Stefano, G. B. Morphine 6beta glucuronide: fortuitous morphine metabolite or preferred peripheral regulatory opiate? *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **11,** MS43-46 (2005).
86. Young, A. M., Havens, J. R. & Leukefeld, C. G. Route of administration for illicit prescription opioids: a comparison of rural and urban drug users. *Harm. Reduct. J.* **7,** 24 (2010).
87. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* **6,** e16765 (2011).
88. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLOS ONE* **4,** e5553 (2009).
89. Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science* **330,** 1355–1358 (2010).
90. Lee, H., DeLoache, W. C. & Dueber, J. E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* **14,** 242–251 (2012).
91. Fernández-Aguado, M., Ullán, R. V., Teijeira, F., Rodríguez-Castro, R. & Martín, J. F. The transport of phenylacetic acid across the peroxisomal membrane is mediated by the PaaT protein in Penicillium chrysogenum. *Appl. Microbiol. Biotechnol.* **97,** 3073–3084 (2012).
92. Hecht, K. A., O'Donnell, A. F. & Brodsky, J. L. The proteolytic landscape of the yeast vacuole. *Cell. Logist.* **4,** (2014).
93. Ayer, A. *et al.* Distinct Redox Regulation in Sub-Cellular Compartments in Response to Various Stress Conditions in Saccharomyces cerevisiae. *PLoS ONE* **8,** e65240 (2013).
94. Ma, C., Hagstrom, D., Polley, S. G. & Subramani, S. Redox-regulated Cargo Binding and Release by the Peroxisomal Targeting Signal Receptor, Pex5. *J. Biol. Chem.* **288,** 27220–27231 (2013).

95. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31,** 335–341 (2013).
96. Bayer, T. S. *et al.* Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *J. Am. Chem. Soc.* **131,** 6508–6515 (2009).
97. Purdue, P. E. & Lazarow, P. B. Peroxisome Biogenesis. *Annu. Rev. Cell Dev. Biol.* **17,** 701–752 (2001).
98. Gleeson, M. A. & Sudbery, P. E. The methylotrophic yeasts. *Yeast* **4,** 1–15 (1988).
99. Meijer, W. H. *et al.* Peroxisomes Are Required for Efficient Penicillin Biosynthesis in Penicillium chrysogenum. *Appl. Environ. Microbiol.* **76,** 5702–5709 (2010).
100. Subramani, S. Targeting of proteins into the peroxisomal matrix. *J. Membr. Biol.* **125,** 99–106 (1992).
101. Liu, F. *et al.* Making two organelles from one: Woronin body biogenesis by peroxisomal protein sorting. *J. Cell Biol.* **180,** 325–339 (2008).
102. Bhataya, A., Schmidt-Dannert, C. & Lee, P. C. Metabolic engineering of Pichia pastoris X-33 for lycopene production. *Process Biochem.* **44,** 1095–1102 (2009).
103. Poirier, Y., Erard, N. & MacDonald-Comber Petétot, J. Synthesis of polyhydroxyalkanoate in the peroxisome of Pichia pastoris. *FEMS Microbiol. Lett.* **207,** 97–102 (2002).
104. Meinecke, M. *et al.* The peroxisomal importomer constitutes a large and highly dynamic pore. *Nat. Cell Biol.* **12,** 273–7 (2010).
105. Kim, P. K. & Hettema, E. H. Multiple Pathways for Protein Transport to Peroxisomes. *J. Mol. Biol.* **427,** 1176–1190 (2015).
106. Smith, J. J. & Aitchison, J. D. Peroxisomes take shape. *Nat. Rev. Mol. Cell Biol.* **14,** 803–817 (2013).
107. Maynard, E. L. & Berg, J. M. Quantitative Analysis of Peroxisomal Targeting Signal Type-1 Binding to Wild-type and Pathogenic Mutants of Pex5p Supports an Affinity Threshold for Peroxisomal Protein Targeting. *J. Mol. Biol.* **368,** 1259–1266 (2007).
108. Noguchi, M., Okumoto, K. & Fujiki, Y. System to quantify the import of peroxisomal matrix proteins by fluorescence intensity. *Genes Cells* **18,** 476–492 (2013).
109. Hagen, S. *et al.* Structural insights into cargo recognition by the yeast PTS1-receptor. *J. Biol. Chem.* jbc.M115.657973 (2015). doi:10.1074/jbc.M115.657973
110. Nyathi, Y. *et al.* The Arabidopsis peroxisomal ABC transporter, comatose, complements the Saccharomyces cerevisiae pxa1 pxa2Delta

mutant for metabolism of long-chain fatty acids and exhibits fatty acyl-CoA-stimulated ATPase activity. *J. Biol. Chem.* **285,** 29892–902 (2010).

111. Antonenkov, V. D. & Hiltunen, J. K. Transfer of metabolites across the peroxisomal membrane. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1822,** 1374–1386 (2012).
112. Duve, C. D. & Baudhuin, P. Peroxisomes (microbodies and related particles). *Physiol. Rev.* **46,** 323–357 (1966).
113. Wolvetang, E. J., Tager, J. M. & Wanders, R. J. A. Latency of the peroxisomal enzyme acyl-coA:dihydroxyacetonephosphate acyltransferase in digitonin-permeabilized fibroblasts: the effect of ATP and ATPase inhibitors. *Biochem. Biophys. Res. Commun.* **170,** 1135–1143 (1990).
114. Antonenkov, V. D., Sormunen, R. T. & Hiltunen, J. K. The rat liver peroxisomal membrane forms a permeability barrier for cofactors but not for small metabolites in vitro. *J. Cell Sci.* **117,** 5633–5642 (2004).
115. Antonenkov, V. D., Mindthoff, S., Grunau, S., Erdmann, R. & Hiltunen, J. K. An involvement of yeast peroxisomal channels in transmembrane transfer of glyoxylate cycle intermediates. *Int. J. Biochem. Cell Biol.* **41,** 2546–2554 (2009).
116. Antonenkov, V. D. & Hiltunen, J. K. Peroxisomal membrane permeability and solute transfer. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1763,** 1697–1706 (2006).
117. Balibar, C. J. & Walsh, C. T. In Vitro Biosynthesis of Violacein from l-Tryptophan by the Enzymes VioA−E from Chromobacterium violaceum†. *Biochemistry (Mosc.)* **45,** 15444–15457 (2006).
118. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* (2015). doi:10.1021/sb500366v
119. Tam, Y. Y. C. *et al.* Pex11-related Proteins in Peroxisome Dynamics: A Role for the Novel Peroxin Pex27p in Controlling Peroxisome Size and Number in Saccharomyces cerevisiae. *Mol. Biol. Cell* **14,** 4089–4102 (2003).
120. Glover, J. R., Andrews, D. W. & Rachubinski, R. A. Saccharomyces cerevisiae peroxisomal thiolase is imported as a dimer. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10541–10545 (1994).
121. Chen, R. *et al.* A Barcoding Strategy Enabling Higher-Throughput Library Screening by Microscopy. *ACS Synth. Biol.* 150715145000006 (2015). doi:10.1021/acssynbio.5b00060
122. Galazka, J. M. *et al.* Cellodextrin Transport in Yeast for Improved Biofuel Production. *Science* **330,** 84–86 (2010).

123. Grunau, S. *et al.* Channel-forming activities of peroxisomal membrane proteins from the yeast Saccharomyces cerevisiae. *FEBS J.* **276,** 1698–1708 (2009).
124. Kuga, S. Pore size distribution analysis of gel substances by size exclusion chromatography. *J. Chromatogr. A* **206,** 449–461 (1981).
125. Kurath, S. F. & Bump, D. D. Hydrodynamic friction coefficients for cellodextrins in water. *J. Polym. Sci. A* **3,** 1515–1526 (1965).
126. Ihnat, M. & Goring, D. A. I. Shape of the cellodextrins in aqueous solution at 25 °C. *Can. J. Chem.* **45,** 2353–2361 (1967).
127. Roermund, C. W. T. van *et al.* The peroxisomal lumen in Saccharomyces cerevisiae is alkaline. *J. Cell Sci.* **117,** 4231–4237 (2004).
128. Liu, X., Ma, C. & Subramani, S. Recent advances in peroxisomal matrix protein import. *Curr. Opin. Cell Biol.* **24,** 484–489 (2012).
129. Williams, C., Bener Aksam, E., Gunkel, K., Veenhuis, M. & van der Klei, I. J. The relevance of the non-canonical PTS1 of peroxisomal catalase. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1823,** 1133–1141 (2012).
130. Soga, O., Kinoshita, H., Ueda, M. & Tanaka, A. Evaluation of Peroxisomal Heme in Yeast. *J. Biochem. (Tokyo)* **121,** 25–28 (1997).
131. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25,** 117–124 (2007).
132. Rokka, A. *et al.* Pxmp2 Is a Channel-Forming Protein in Mammalian Peroxisomal Membrane. *PLoS ONE* **4,** e5090 (2009).
133. Halbach, A. *et al.* Targeting of the tail-anchored peroxisomal membrane proteins PEX26 and PEX15 occurs through C-terminal PEX19-binding sites. *J. Cell Sci.* **119,** 2508–2517 (2006).
134. Kim, H., Lee, W.-H., Galazka, J. M., Cate, J. H. D. & Jin, Y.-S. Analysis of cellodextrin transporters from Neurospora crassa in Saccharomyces cerevisiae for cellobiose fermentation. *Appl. Microbiol. Biotechnol.* 1–8 (2013). doi:10.1007/s00253-013-5339-2
135. Ryan, O. W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife* **3,** e03703 (2014).
136. Eriksen, D. T., Hsieh, P. C. H., Lynn, P. & Zhao, H. Directed evolution of a cellobiose utilization pathway in Saccharomyces cerevisiae by simultaneously engineering multiple proteins. *Microb. Cell Factories* **12,** 1–11 (2013).
137. Lian, J., Li, Y., HamediRad, M. & Zhao, H. Directed evolution of a cellodextrin transporter for improved biofuel production under

anaerobic conditions in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* n/a-n/a (2014). doi:10.1002/bit.25214

138. Znameroski, E. A. *et al.* Evidence for Transceptor Function of Cellodextrin Transporters in Neurospora crassa. *J. Biol. Chem.* **289,** 2610–2619 (2014).
139. Ha, S.-J. *et al.* Energetic benefits and rapid cellobiose fermentation by Saccharomyces cerevisiae expressing cellobiose phosphorylase and mutant cellodextrin transporters. *Metab. Eng.* **15,** 134–143 (2013).
140. Halbach, A. *et al.* Function of the PEX19-binding Site of Human Adrenoleukodystrophy Protein as Targeting Motif in Man and Yeast PMP TARGETING IS EVOLUTIONARILY CONSERVED. *J. Biol. Chem.* **280,** 21176–21182 (2005).
141. Halbach, A., Rucktäschel, R., Rottensteiner, H. & Erdmann, R. The N-domain of Pex22p Can Functionally Replace the Pex3p N-domain in Targeting and Peroxisome Formation. *J. Biol. Chem.* **284,** 3906–3916 (2009).
142. Yuan, X., Protchenko, O., Philpott, C. C. & Hamza, I. Topologically Conserved Residues Direct Heme Transport in HRG-1-related Proteins. *J. Biol. Chem.* **287,** 4914–4924 (2012).
143. Mindthoff, S. *et al.* Peroxisomal Pex11 is a pore-forming protein homologous to TRPM channels. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* (2015). doi:10.1016/j.bbamcr.2015.11.013
144. Idiris, A., Tohda, H., Kumagai, H. & Takegawa, K. Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* **86,** 403–417 (2010).
145. Wang, P. *et al.* Enhanced secretion of a methyl parathion hydrolase in Pichia pastoris using a combinational strategy. *Microb. Cell Factories* **14,** 123 (2015).
146. Costantini, L. M., Subach, O. M., Jaureguiberry-bravo, M., Verkhusha, V. V. & Snapp, E. L. Cysteineless non-glycosylated monomeric blue fluorescent protein, secBFP2, for studies in the eukaryotic secretory pathway. *Biochem. Biophys. Res. Commun.* **430,** 1114–1119 (2013).
147. Erdmann, R. & Gould, S. J. in *Methods in Enzymology* (ed. Christine Guthrie, G. R. F.) **Volume 351,** 365–381 (Academic Press, 2002).
148. Distel, B. & Kragt, A. Purification of yeast peroxisomes. *Methods Mol. Biol. Clifton NJ* **313,** 21–6 (2006).
149. Cramer, J., Effelsberg, D., Girzalsky, W. & Erdmann, R. Small-Scale Purification of Peroxisomes for Analytical Applications. *Cold Spring Harb. Protoc.* **2015,** pdb.prot083717 (2015).

150. Wang, Y., Taylor, T. H. & Arriaga, E. A. Analysis of the bioactivity of magnetically immunoisolated peroxisomes. *Anal. Bioanal. Chem.* **402,** 41–49 (2011).
151. Zipor, G., Brocard, C. & Gerst, J. in *RNA Detection and Visualization* (ed. Gerst, J. E.) 323–333 (Humana Press, 2011).
152. Güther, M. L. S., Urbaniak, M. D., Tavendale, A., Prescott, A. & Ferguson, M. A. J. High-Confidence Glycosome Proteome for Procyclic Form Trypanosoma brucei by Epitope-Tag Organelle Enrichment and SILAC Proteomics. *J. Proteome Res.* **13,** 2796–2806 (2014).
153. Ru, Y. *et al.* A micropreparation of mitochondria from cells using magnetic beads with immunoaffinity. *Anal. Biochem.* **421,** 219–226 (2012).
154. Franko, A. *et al.* Efficient Isolation of Pure and Functional Mitochondria from Mouse Tissues Using Automated Tissue Disruption and Enrichment with Anti-TOM22 Magnetic Beads. *PLoS ONE* **8,** e82392 (2013).
155. Schmidt, T. G. & Skerra, A. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.* **2,** 1528–1535 (2007).
156. Shai, N., Schuldiner, M. & Zalckvar, E. No peroxisome is an island — Peroxisome contact sites. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* (2015). doi:10.1016/j.bbamcr.2015.09.016
157. Sagt, C. M. J. *et al.* Peroxicretion: a novel secretion pathway in the eukaryotic cell. *BMC Biotechnol.* **9,** 48–48 (2009).
158. Hall, M. P. *et al.* Engineered Luciferase Reporter from a Deep Sea Shrimp Utilizing a Novel Imidazopyrazinone Substrate. *ACS Chem. Biol.* **7,** 1848–1857 (2012).
159. Schmitz, U. K., Lonsdale, D. M. & Jefferson, R. A. Application of the β-glucuronidase gene fusion system to Saccharomyces cerevisiae. *Curr. Genet.* **17,** 261–264 (1990).
160. Broothaerts, W. *et al.* Gene transfer to plants by diverse species of bacteria. *Nature* **433,** 629–633 (2005).
161. Troy AA, H. A Simplified Method for Measuring Secreted Invertase Activity in Saccharomyces cerevisiae. *Biochem. Pharmacol. Open Access* **3,** (2014).
162. Marten, M. R. & Seo, J.-H. Localization of cloned invertase in Saccharomyces cerevisiae directed by the SUC2 and MFα1 signal sequences. *Biotechnol. Bioeng.* **34,** 1133–1139 (1989).
163. Tammi, M., Ballou, L., Taylor, A. & Ballou, C. E. Effect of glycosylation on yeast invertase oligomer stability. *J. Biol. Chem.* **262,** 4395–4401 (1987).

164. Esmon, P. C., Esmon, B. E., Schauer, I. E., Taylor, A. & Schekman, R. Structure, assembly, and secretion of octameric invertase. *J. Biol. Chem.* **262,** 4387–4394 (1987).
165. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).
166. Dinh, T. & Bernhardt, T. G. Using Superfolder Green Fluorescent Protein for Periplasmic Protein Localization Studies. *J. Bacteriol.* **193,** 4984–4987 (2011).
167. Rizo, J. & Xu, J. Synaptic Vesicle Fusion without SNARE Transmembrane Regions. *Dev. Cell* **27,** 124–126 (2013).
168. Zhou, P., Bacaj, T., Yang, X., Pang, Z. P. & Südhof, T. C. Lipid-Anchored SNAREs Lacking Transmembrane Regions Fully Support Membrane Fusion during Neurotransmitter Release. *Neuron* **80,** 470–483 (2013).
169. Rossi, G., Salminen, A., Rice, L. M., Brünger, A. T. & Brennwald, P. Analysis of a Yeast SNARE Complex Reveals Remarkable Similarity to the Neuronal SNARE Complex and a Novel Function for the C Terminus of the SNAP-25 Homolog, Sec9. *J. Biol. Chem.* **272,** 16610–16617 (1997).
170. Yu, I.-M. & Hughson, F. M. Tethering Factors as Organizers of Intracellular Vesicular Traffic. *Annu. Rev. Cell Dev. Biol.* **26,** 137–156 (2010).
171. Lipatova, Z., Hain, A. U., Nazarko, V. Y. & Segev, N. Ypt/Rab GTPases: Principles learned from yeast. *Crit. Rev. Biochem. Mol. Biol.* **50,** 203–211 (2015).
172. Girzalsky, W. & Erdmann, R. Give what you can and keep what you need! *EMBO J.* **32,** 2424–2426 (2013).
173. van der Zand, A., Gent, J., Braakman, I. & Tabak, H. F. Biochemically Distinct Vesicles from the Endoplasmic Reticulum Fuse to Form Peroxisomes. *Cell* **149,** 397–409 (2012).
174. David, C. *et al.* A Combined Approach of Quantitative Interaction Proteomics and Live-cell Imaging Reveals a Regulatory Role for Endoplasmic Reticulum (ER) Reticulon Homology Proteins in Peroxisome Biogenesis. *Mol. Cell. Proteomics* **12,** 2408–2425 (2013).
175. Knoblach, B. *et al.* An ER-peroxisome tether exerts peroxisome population control in yeast. *EMBO J.* **32,** 2439–2453 (2013).
176. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 31–34 (2007).

177. Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J. & Dueber, J. E. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res.* **41,** 10668–10678 (2013).
178. Gatto, G. J., Geisbrecht, B. V., Gould, S. J. & Berg, J. M. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat. Struct. Mol. Biol.* **7,** 1091–1095 (2000).
179. Roehrl, M. H. A., Wang, J. Y. & Wagner, G. A General Framework for Development and Data Analysis of Competitive High-Throughput Screens for Small-Molecule Inhibitors of Protein−Protein Interactions by Fluorescence Polarization†. *Biochemistry (Mosc.)* **43,** 16056–16066 (2004).
180. Franssen, M. C. R., Kircher, M. & Wohlgemuth, R. in *Industrial Biotechnology* (eds. Soetaert, W. & Vandamme, E. J.) 323–350 (Wiley-VCH Verlag GmbH & Co. KGaA, 2010).
181. Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol.* **14,** 454–459 (2003).
182. Yim, H. *et al.* Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7,** 445–452 (2011).
183. Murdock, D., Ensley, B. D., Serdar, C. & Thalen, M. Construction of Metabolic Operons Catalyzing the De Novo Biosynthesis of Indigo in Escherichia coli. *Nat. Biotechnol.* **11,** 381–386 (1993).
184. Berry, A., Dodge, T. C., Pepsin, M. & Weyler, W. Application of metabolic engineering to improve both the production and use of biotech indigo. *J. Ind. Microbiol. Biotechnol.* **28,** 127–133 (2002).
185. Straathof, A. J. J. & Adlercreutz, P. *Applied Biocatalysis*. (CRC Press, 2003).
186. State Council of the People's Republic of China & EU-China Environmental Governance Programme. *Action Plan for Water Pollution Prevention 2015*. (2015).
187. Paul, R. *Denim: Manufacture, Finishing and Applications*. (Elsevier, 2015).
188. Paul, S. A., Chavan, S. K. & Khambe, S. D. Studies on characterization of textile industrial waste water in Solapur city. *Int. J. Chem. Sci.* **10,** 635–642 (2012).
189. Minami, Y., Nishimura, O., Hara-Nishimura, I., Nishimura, M. & Matsubara, H. Tissue and Intracellular Localization of Indican and the Purification and Characterization of Indican Synthase from Indigo Plants. *Plant Cell Physiol.* **41,** 218–225 (2000).

190. Minami, Y. *et al.* Cloning, sequencing, characterization, and expression of a β-glucosidase cDNA from the indigo plant. *Plant Sci.* **142,** 219–226 (1999).
191. Angelini, L. G., Tozzi, S. & Nassi o Di Nasso, N. Environmental Factors Affecting Productivity, Indican Content, and Indigo Yield in Polygonum tinctorium Ait., a Subtropical Crop Grown under Temperate Conditions. *J. Agric. Food Chem.* **52,** 7541–7547 (2004).
192. Shakoor, N. *et al.* A Sorghum bicolor expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums. *BMC Plant Biol.* **14,** 35 (2014).
193. Choi, H. S. *et al.* A novel flavin-containing monooxygenase from Methylophaga sp. strain SK1 and its indigo synthesis in Escherichia coli. *Biochem. Biophys. Res. Commun.* **306,** 930–936 (2003).
194. Han, G. H., Shin, H.-J. & Kim, S. W. Optimization of bio-indigo production by recombinant E. coli harboring fmo gene. *Enzyme Microb. Technol.* **42,** 617–623 (2008).
195. Han, G. H. *et al.* Bio-indigo production in two different fermentation systems using recombinant Escherichia coli cells harboring a flavin-containing monooxygenase gene (fmo). *Process Biochem.* **46,** 788–791 (2011).
196. Ruettinger, R. T., Wen, L. P. & Fulco, A. J. Coding nucleotide, 5′ regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450 reductase from Bacillus megaterium. *J. Biol. Chem.* **264,** 10987–10995 (1989).
197. Li, H., Mei, L., Urlacher, V. B. & Schmid, R. D. Cytochrome P450 BM-3 Evolved by Random and Saturation Mutagenesis as an Effective Indole-Hydroxylating Catalyst. *Appl. Biochem. Biotechnol.* **144,** 27–36 (2008).
198. Park, S.-H. *et al.* Engineering Bacterial Cytochrome P450 (P450) BM3 into a Prototype with Human P450 Enzyme Activity Using Indigo Formation. *Drug Metab. Dispos.* **38,** 732–739 (2010).
199. Dietrich, J. A. *et al.* A Novel Semi-biosynthetic Route for Artemisinin Production Using Engineered Substrate-Promiscuous P450BM3. *ACS Chem. Biol.* **4,** 261–267 (2009).
200. van Vugt-Lussenburg, B. M. A. *et al.* Identification of Critical Residues in Novel Drug Metabolizing Mutants of Cytochrome P450 BM3 Using Random Mutagenesis. *J. Med. Chem.* **50,** 455–461 (2007).
201. Hakulinen, N., Paavilainen, S., Korpela, T. & Rouvinen, J. The Crystal Structure of β-Glucosidase from Bacillus circulans sp. alkalophilus: Ability to Form Long Polymeric Assemblies. *J. Struct. Biol.* **129,** 69–79 (2000).

202. Paavilainen, S., Hellman, J. & Korpela, T. Purification, characterization, gene cloning, and sequencing of a new beta-glucosidase from Bacillus circulans subsp. alkalophilus. *Appl. Environ. Microbiol.* **59,** 927–932 (1993).
203. Kim, J.-Y., Lee, J.-Y., Shin, Y.-S. & Kim, G.-J. Characterization of an indican-hydrolyzing enzyme from Sinorhizobium meliloti. *Process Biochem.* **45,** 892–896 (2010).
204. Opassiri, R. *et al.* Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. *BMC Plant Biol.* **6,** 33 (2006).
205. Marcinek, H., Weyler, W., Deus-Neumann, B. & Zenk, M. H. Indoxyl-UDPG-glucosyltransferase from Baphicacanthus cusia. *Phytochemistry* **53,** 201–207 (2000).
206. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8,** 1494–1512 (2013).
207. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29,** 644–652 (2011).
208. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28,** 1086–1092 (2012).
209. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
210. Hefner, T., Arend, J., Warzecha, H., Siems, K. & Stöckigt, J. Arbutin synthase, a novel member of the NRD1β glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg. Med. Chem.* **10,** 1731–1741 (2002).
211. Brazier-Hicks, M. & Edwards, R. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in Arabidopsis thaliana. *Plant J.* **42,** 556–566 (2005).
212. Brazier-Hicks, M. *et al.* Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc. Natl. Acad. Sci.* **104,** 20238–20243 (2007).
213. Zangoui, P., Vashishtha, K. & Mahadevan, S. Evolution of Aromatic β-Glucoside Utilization by Successive Mutational Steps in Escherichia coli. *J. Bacteriol.* **197,** 710–716 (2015).
214. Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D. & Martín, H. G. Synthetic and systems biology for microbial production of commodity chemicals. *Npj Syst. Biol. Appl.* **2,** 16009 (2016).
215. Vogt, T. & Jones, P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci.* **5,** 380–386 (2000).

216. Patnaik, R., Zolandz, R. R., Green, D. A. & Kraynie, D. F. L-Tyrosine production by recombinant Escherichia coli: Fermentation optimization and recovery. *Biotechnol. Bioeng.* **99,** 741–752 (2008).
217. Hansen, E. H. *et al.* De Novo Biosynthesis of Vanillin in Fission Yeast (Schizosaccharomyces pombe) and Baker's Yeast (Saccharomyces cerevisiae). *Appl. Environ. Microbiol.* **75,** 2765–2774 (2009).
218. Minami, Y., Sarangi, B. K. & Thul, S. T. Transcriptome analysis for identification of indigo biosynthesis pathway genes in Polygonum tinctorium. *Biologia (Bratisl.)* **70,** 1026–1032 (2016).
219. Mao, Z., Shin, H.-D. & Chen, R. R. Engineering the E. coli UDP-Glucose Synthesis Pathway for Oligosaccharide Synthesis. *Biotechnol. Prog.* **22,** 369–374 (2006).
220. Simkhada, D., Kurumbang, N. P., Lee, H. C. & Sohng, J. K. Exploration of glycosylated flavonoids from metabolically engineered E. coli. *Biotechnol. Bioprocess Eng.* **15,** 754–760 (2010).
221. Williams, G. J., Yang, J., Zhang, C. & Thorson, J. S. Recombinant E. coli Prototype Strains for in Vivo Glycorandomization. *ACS Chem. Biol.* **6,** 95–100 (2011).
222. Xu, Z.-S. *et al.* Characterization of a Bifunctional O- and N-Glucosyltransferase from Vitis vinifera in Glucosylating Phenolic Compounds and 3,4-dichloroaniline in Pichia pastoris and Arabidopsis thaliana. *PLoS ONE* **8,** e80449 (2013).
223. Sepúlveda-Jiménez, G., Rueda-Benítez, P., Porta, H. & Rocha-Sosa, M. A red beet (Beta vulgaris) UDP-glucosyltransferase gene induced by wounding, bacterial infiltration and oxidative stress. *J. Exp. Bot.* **56,** 605–611 (2005).
224. Sarry, J.-E. & Günata, Z. Plant and microbial glycoside hydrolases: Volatile release from glycosidic aroma precursors. *Food Chem.* **87,** 509–521 (2004).
225. Ahmad, S. A. & Hopkins, T. L. Phenol β-glucosyltransferases in six species of insects: properties and tissue localization. *Comp. Biochem. Physiol. Part B Comp. Biochem.* **104,** 515–519 (1993).
226. Schweiger, W. *et al.* Validation of a Candidate Deoxynivalenol-Inactivating UDP-Glucosyltransferase from Barley by Heterologous Expression in Yeast. *Mol. Plant. Microbe Interact.* **23,** 977–986 (2010).
227. Quirós, L. M., Aguirrezabalaga, I., Olano, C., Méndez, C. & Salas, J. A. Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by Streptomyces antibioticus. *Mol. Microbiol.* **28,** 1177–1185 (1998).

228. Bolam, D. N. *et al.* The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc. Natl. Acad. Sci.* **104,** 5336–5341 (2007).
229. Ahmad, S. A., Hopkins, T. L. & Kramer, K. J. Tyrosine β-Glucosyltransferase in the tobacco hornworm, Manduca sexta (L.): properties, tissue localization, and developmental profile. *Insect Biochem. Mol. Biol.* **26,** 49–57 (1996).
230. Agatsuma, S., Sekino, H., Nagoshi, T. & Watanabe, H. Indoxyl-beta-D-glucuronide, the primary emitter of low-level chemiluminescence in plasma of hemodialysis patients. *Clin. Chem.* **40,** 1580–1586 (1994).
231. BANOGLU, E. & KING, R. S. Sulfation of indoxyl by human and rat aryl (phenol) sulfotransferases to form indoxyl sulfate. *Eur. J. Drug Metab. Pharmacokinet.* **27,** 135–140 (2002).
232. Laffy, P. W. Evolution, gene expression and enzymatic production of Tyrian purple: A molecular study of the Australian muricid Dicathais orbita (Neogastropoda: Muricidae). (Flinders University, School of Biological Sciences., 2011).
233. Lunan, K. D. & Mitchell, H. K. The metabolism of tyrosine-O-phosphate in Drosophila. *Arch. Biochem. Biophys.* **132,** 450–456 (1969).
234. Biswas, T., Houghton, J. L., Garneau-Tsodikova, S. & Tsodikov, O. V. The structural basis for substrate versatility of chloramphenicol acetyltransferase CATI. *Protein Sci.* **21,** 520–530 (2012).
235. Zenser, T. V. *et al.* Human N-Acetylation of Benzidine: Role of NAT1 and NAT2. *Cancer Res.* **56,** 3941–3947 (1996).
236. Dang, T.-T. T., Chen, X. & Facchini, P. J. Acetylation serves as a protective group in noscapine biosynthesis in opium poppy. *Nat. Chem. Biol.* **11,** 104–106 (2015).
237. Taguchi, G. *et al.* Malonylation is a key reaction in the metabolism of xenobiotic phenolic glucosides in Arabidopsis and tobacco. *Plant J.* **63,** 1031–1041 (2010).
238. Holt, S. J. & Withers, R. F. J. Studies in Enzyme Cytochemistry. V. An Appraisal of Indigogenic Reactions for Esterase Localization. *Proc. R. Soc. Lond. Ser. B - Biol. Sci.* **148,** 520–532 (1958).
239. Horwitz, J. P. *et al.* SUBSTRATES FOR CYTOCHEMICAL DEMONSTRATION OF ENZYME ACTIVITY. I. SOME SUBSTITUTED 3-INDOLYL-BETA-D-GLYCOPYRANOSIDES. *J. Med. Chem.* **7,** 574–575 (1964).
240. Beil, S. *et al.* Purification and Characterization of the Arylsulfatase Synthesized by Pseudomonas aeruginosa PAO During Growth in Sulfate-Free Medium and Cloning of the Arylsulfatase Gene (atsA). *Eur. J. Biochem.* **229,** 385–394 (1995).

241. Berger, I., Guttman, C., Amar, D., Zarivach, R. & Aharoni, A. The Molecular Basis for the Broad Substrate Specificity of Human Sulfotransferase 1A1. *PLoS ONE* **6,** e26794 (2011).
242. Raman, R., Myette, J., Venkataraman, G., Sasisekharan, V. & Sasisekharan, R. Identification of Structural Motifs and Amino Acids within the Structure of Human Heparan Sulfate 3-O-Sulfotransferase That Mediate Enzymatic Function. *Biochem. Biophys. Res. Commun.* **290,** 1214–1219 (2002).
243. Møldrup, M. E., Geu-Flores, F., Olsen, C. E. & Halkier, B. A. Modulation of sulfur metabolism enables efficient glucosinolate engineering. *BMC Biotechnol.* **11,** 12 (2011).
244. Kim, J. y., Lee, K.Kim, Y., Kim, C.-K. & Lee, K. Production of dyestuffs from indole derivatives by naphthalene dioxygenase and toluene dioxygenase. *Lett. Appl. Microbiol.* **36,** 343–348 (2003).
245. Zhang, X. *et al.* Cloning and expression of naphthalene dioxygenase genes from Comamonas sp. MQ for indigoids production. *Process Biochem.* **48,** 581–587 (2013).
246. Nakamura, K., Martin, M. V. & Guengerich, F. P. Random Mutagenesis of Human Cytochrome P450 2A6 and Screening with Indole Oxidation Products1. *Arch. Biochem. Biophys.* **395,** 25–31 (2001).
247. Zhang, Z.-G. *et al.* Identification of amino acid residues involved in 4-chloroindole 3-hydroxylation by cytochrome P450 2A6 using screening of random libraries. *J. Biotechnol.* **139,** 12–18 (2009).
248. Rioz-Martínez, A. *et al.* Exploring the biocatalytic scope of a bacterial flavin-containing monooxygenase. *Org. Biomol. Chem.* **9,** 1337–1341 (2011).
249. Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409,** 241–246 (2001).
250. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* btu170 (2014). doi:10.1093/bioinformatics/btu170
251. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
252. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4,** 900 (2015).
253. Towns, J. *et al.* XSEDE: Accelerating Scientific Discovery. *Computing in Science and Engineering* **16,** 62–74 (2014).
254. Nick Nystrom, Joel Welling, Phil Blood & EngLim Goh. in *Contemporary High Performance Computing* 421–440 (Chapman and Hall/CRC, 2013).

255. Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *J. Proteome Res.* **1,** 21–26 (2002).
256. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).