Title
Biochemical Protecting Groups for a Sustainable Indigo Dyeing Method

Author
Hsu, Tammy Melody

Publication Date
2019

Peer reviewed|Thesis/dissertation

Biochemical Protecting Groups for a Sustainable Indigo Dyeing Method

by

Tammy Melody Hsu

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 Tanja Kortemme
Professor David F. Savage

Spring 2019

Biochemical Protecting Groups for a Sustainable Indigo Dyeing Method

Abstract

Biochemical Protecting Groups for a Sustainable Indigo Dyeing Method

by

Tammy Melody Hsu

Joint Doctor of Philosophy
with University of California, San Francisco
in Bioengineering

University of California, Berkeley

Professor John E. Dueber, Chair

Indigo is a widely used dye in the textile industry, but current industry practice uses toxic chemicals to produce and solubilize the compound. I am developing a more environmentally friendly dye process by using microbial fermentation to produce indigo dye. Unlike the industrial procedure, microbes can produce indigo from renewable sources, and many of the hazardous chemicals can be eliminated from the dye synthesis and dye application processes. First, I describe a method to dye with indican, the glucosylated form of an unstable indigo precursor found in indigo-producing plants. I identified the gene sequence for a key biosynthetic enzyme from the indigo plant, and I expressed that gene in an *Escherichia coli* strain to produce indican. The glucosyl group on indican can be removed by a β-glucosidase, forming indigo. I show that applying indican and β-glucosidase to cotton can dye it an indigo color. Second, I extend this strategy of biological protecting groups towards producing a second-generation dye molecule, isatan B. Although chemically similar to indican, isatan B is hydrolyzed at a basic pH, a desirable property for the indigo dyeing process. I isolated the isatan B molecule from plant leaves and used the purified molecule to characterize isatan B's stability in the *E. coli* host. I then describe efforts to construct a biosynthetic pathway for isatan B in *E. coli*.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to thank my advisor John Dueber for always believing in me. Even in the darker days of my grad school journey, he always pushed me to reach for higher goals and take measurable steps to achieve them. His mentorship has been invaluable, and he taught me how to ask good scientific questions, how to think creatively and rigorously, and how to craft an impactful story.

I would also like to thank all the members of the Dueber Lab, past and present, for being such a supportive family that helped me grow in so many ways. In particular, I would like to thank Zach Russ, Luke Latimer, and Ryan Protzko for answering my incessant questions and providing guidance on how to design smart experiments. I would also like to thank Judy Savitskaya, Parry Grewal, and Jen Samson for many lunches on the 5th floor balcony, musing on science, complaining about cloning, sharing stories about local startups, and planning our next climbing session.

I owe a great debt to Zach and the iGEM 2013 team for sowing the seeds for the indigo project. Without their work, who knows if the project would have taken off like it did. Helping in the effort over the years were also a host of hardworking undergrads, each working on a corner of the puzzle but all contributing to the whole. I would also like to thank the residents of the Energy Biosciences Building for teaching me how to use a variety of equipment (and how to fix them when I broke them), and I would like to thank my collaborator, Ditte Welner, for her help in protein crystallography.

Lastly, I would like to thank my friends and family for always being there for me. My housemates have been a great support system, always down for a practice quals talk or a hot pot date. My parents and brother never fail to give me perspective, urge me to take care of myself, and remind me that there is life beyond grad school. I am truly grateful for their support.

Chapter 1. Introduction

1.1. Motivation

For thousands of years, humans have turned to plants for their materials, remedies, fragrances, dyes, and more. The natural world has a wealth of products that we have used for our comfort and survival – cotton for textiles, quinine for treating malaria, madder for red dye, and much more. These plant-derived products get the job done, but they can be hard to scale, either because the plant takes time to cultivate or because the plant only produces minute quantities of the desired product.

In the past century – and especially in recent decades – organic chemistry has replaced many of these plant-derived products. We can now make plastics with a dizzying array of properties, all sorts of synthetic drugs, and fragrances to mimic most natural scents. Unlike plant products, organic chemistry is readily scalable – it does not typically require large amounts of time or space to produce large quantities of material. It is also more flexible than relying on existing biosynthetic pathways in pathways, allowing chemists to produce compounds with desirable qualities not found in nature. However, chemical synthesis is not the most sustainable, with petroleum being a common feedstock.

As synthetic biologists, we aim to bring together the best of both the plant world and the organic chemistry world. By moving product biosynthesis to microbes, we are working to produce these molecules at high titer, without requiring large fields and long growth cycles. We also use more renewable feedstocks, that are not petroleum-derived. Currently, it is still a feat to produce natural compounds in microbes using biosynthetic pathways that are similar to those found in nature, but in the future, we hope to be able to swap out enzymes with others to produce molecules that are not found in nature. With protein engineering techniques, we can evolve enzymes to act on substrates different from their natural ones, expanding the reaction space possible in the biosynthetic toolbox.

One example of this progression from plant-sourced, to chemically synthesized, to biosynthesized in microbes, is indigo dye. For millennia, indigo has been extracted from a number of indigo-producing plants, and the dye was so coveted that fields of indigo were grown in tropical places around the world[1]. The chemical synthesis of indigo was developed in 1890, and from then on, the chemical method scaled to keep up with demand for the dye, while plant indigo could not compete and was replaced[2]. Unfortunately, the synthesis of indigo requires numerous toxic chemicals and is derived from the petroleum product aniline. This led to the development of bio-indigo by Amgen

and Genencor in the 1990s[3,4]. Though their work producing indigo in *E. coli* was groundbreaking, I want to build on their developments by creating a better dye process, that replaces even more of the hazardous chemicals used in industrial indigo dyeing.

1.2. Organization

In this dissertation, I propose two related strategies for dyeing with indigo using microbially-produced dye precursors. In **Chapter 2**, I describe the production of indican, an indigo dye precursor, and its use as a textile dye[5]. To build the biosynthetic pathway for indican in *E. coli*, I identified the amino acid sequence of a glucosyltransferase from an indigo-producing plant that naturally produces indican. My collaborator Ditte Welner used protein crystallography to structurally characterize the glucosyltransferase, and I expressed it in *E. coli* along with another heterologous enzyme from bacteria to produce indican. I further improved indican production with strain engineering and media optimization. I then used indican to dye a sample of cotton fabric.

In **Chapter 3**, I extend this strategy to isatan B, a potential second-generation indigo dye precursor molecule. Isatan B is also a natural product from an indigo plant, and I developed a method to purify the molecule from plant leaves for use as a chemical standard. I then characterized the stability of isatan B in the presence of *E. coli* cells, and I tested a few strains that degraded isatan B to varying degrees. I then searched for an oxidase that would act on indican to produce isatan B. Candidates from bacteria, fungi, and the woad plant *Isatis tinctoria* were tested, using a variety of methods to assess the activity of the enzymes.

Chapter 2. Employing a biochemical protecting group for a sustainable indigo dyeing strategy

2.1. Introduction

Indigo is a dye prized since antiquity for its vibrancy and deep blue hue, with evidence of its use dating back 6,000 years[6]. For more than a century, its unique properties have been leveraged to produce the popular textile blue denim. Historically, indigo was extracted from dye-producing plants that were important cash crops on European farms. Production later expanded to plantations in India, the Spice Islands, Central America, and South Carolina[1]. This critical economy was supplanted at the turn of the 20th century when a chemical synthesis was developed.

Other blue dyes are now synthesized and can be used on a variety of textiles, but none have replaced indigo as a denim dye. Unlike most dyes, indigo does not covalently bind to cotton but instead adsorbs to the fibers. The adsorbed indigo is robust to strong detergents used for laundering, yet it flakes off with persistent abrasion to expose the internal white yarn cord, yielding the desired worn-in look that personalizes an individual's pair of jeans. This unique combination of robustness to detergents, yet susceptibility to abrasion, makes indigo irreplaceable as a denim dye and contributes to the enduring popularity of blue denim. As of 2011, 50,000 tons of indigo were synthesized per year, 95% of which were used to dye the over 4 billion denim garments manufactured annually[7,8].

The high demand for indigo presents a serious sustainability problem for two reasons. First, indigo is chemically synthesized on an industrial scale from aniline, a toxic compound derived from the petroleum product benzene. The synthesis involves hazardous chemicals including formaldehyde, hydrogen cyanide, sodamide, and strong bases (**Figure 2-1**)[2,9]. Second, indigo is insoluble in water, and it must be reduced to the water-soluble leuco form by an excess of a reducing agent to be used as a dye. In the current industrial process, sodium dithionite is the preferred reducing agent due to its low cost and short reduction time[2,10]. However, sodium dithionite decomposes to form sulfate and sulfite, which can corrode equipment and pipes in dye mills and wastewater treatment facilities[10]. Many dye mills avoid the additional cost of wastewater treatment by dumping the spent dye materials in rivers, where they have negative ecological impacts.

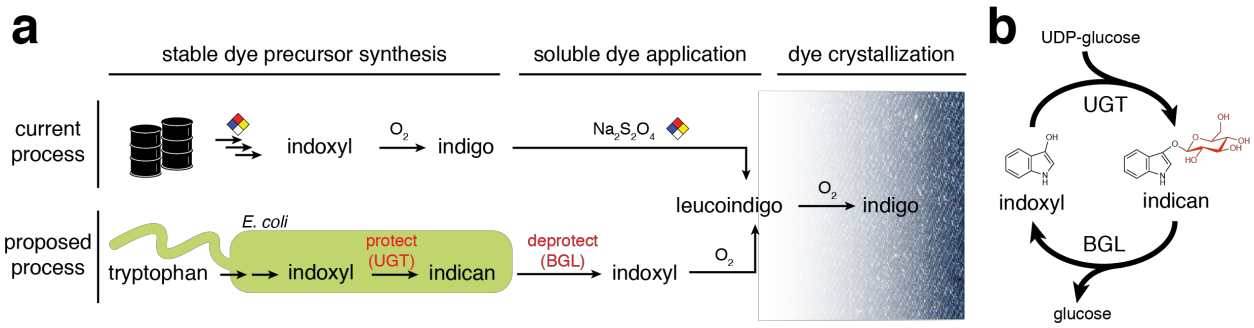

Figure 2-1. A glucosyl protecting group enables control over the timing and location of indigo dyeing. a) The current industrial process involves chemically synthesizing indigo and adding a reducing agent (typically sodium dithionite) to the indigo vat for reduction to dye-competent, soluble leucoindigo. In the proposed microbial process, *E. coli* biosynthesizes indoxyl and glucosylates at the C3 hydroxyl before its spontaneous air oxidation to indigo. The glucoside, indican, is stable in air and can be stored. The glucosyl group is removed only at the point of dyeing, allowing the regenerated indoxyl to oxidize to indigo crystals in cotton fibers. No reducing agent is required when dyeing with indican. b) Glucose acts as a protecting group for indoxyl inside the production host and in the fermenter until removed by co-application with β-glucosidase to cotton.

To combat environmental concerns, microbial production of indigo has previously been targeted as a replacement for chemical synthesis[3,4,11,12]. These efforts offered considerable environmental improvements by removing the need for the harmful chemicals used in indigo synthesis; however, reducing agent was still required to reduce the insoluble indigo to the soluble leucoindigo for dyeing. A number of other strategies (e.g. bacterial fermentation[13,14], catalytic hydrogenation[15], and electrochemical reduction[16]) have been used to reduce indigo, but none are as fast-acting and cost-effective as sodium dithionite.

Here I present an alternative indigo dyeing process, which mimics the natural biochemical protecting group strategy employed by the Japanese indigo plant *Polygonum tinctorium* (**Figure 2-2**). In the leaf, indole is oxygenated at the C3 position to produce the highly reactive indigo precursor indoxyl. Instead of further oxidizing to form indigo, the hydroxyl group is glucosylated, protecting this reactive functional group and generating the colorless molecule indican. Only upon damage to the leaf tissue is blue coloration observed: the intracellular endomembranes of the vacuole (containing indican)[17] and the chloroplast (containing β-glucosidase)[18] are lysed, allowing their contents to mix. β-glucosidase hydrolyzes indican to indoxyl, which spontaneously oxidizes to form crystalline indigo through a leucoindigo intermediate. Thus, the glucose moiety serves as a biochemical protecting group, stabilizing the reactive indoxyl in the reduced state until treated with a β-glucosidase. Accordingly, indican provides many desired attributes for a dye: in addition to water solubility and broad pH and thermostability, indican provides spatial and temporal control over indoxyl formation for direct dyeing of cotton fibers.

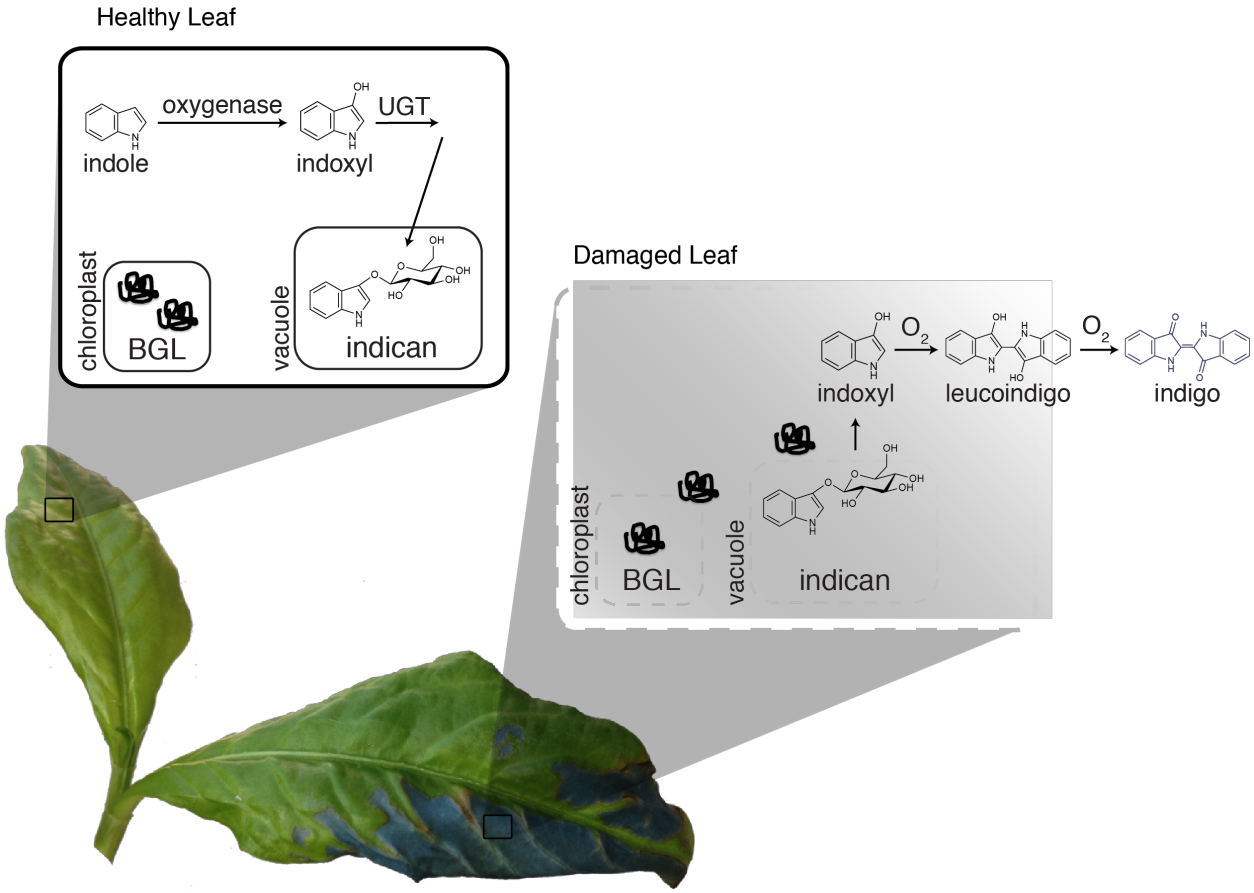

Figure 2-2. *P. tinctorium* indigo biosynthesis pathway. In the indigo plant *P. tinctorium*, the healthy leaf produces and efficiently glucosylates the reactive indoxyl with a glucosyltransferase (UGT) to generate indican, which is stored in the vacuole. Thus, healthy *P. tinctorium* leaves do not show blue coloration. Upon damage to the leaf tissue, induced here by spraying 70% ethanol on the lower right part of the leaf, the vacuole lyses, resulting in indican hydrolysis by a β-glucosidase (BGL) similarly liberated from its chloroplast compartment. The regenerated indoxyl spontaneously oxidizes to form indigo through a leucoindigo intermediate.

The reversible nature of biochemical protecting groups can be a powerful tool in metabolic engineering for controlling metabolite activity, solubility, and recognition by enzymes and transporters. Addition and removal of the protecting group are catalyzed by separate enzymes; thus, these processes can be independently regulated. Several biochemical groups (e.g., sulfyl, acetyl, methyl, malonyl, and glycosyl groups) can serve as protecting groups, with each relating their own chemical properties, such as the charge of a sulfyl group or the hydrophobicity of an acetyl group. Transferase enzymes (e.g., sulfyl-, methyl-, acetyl-, and glycosyl-transferases) can site-selectively protect a functional group while another set of enzymes (e.g., sulfatases, demethylases, esterases, β-glucosidases) can catalyze the removal of these protecting groups. An interesting example is the use of acetylation in the noscapine biosynthetic pathway in *Papaver somniferum L.* in which one hydroxyl group is selectively acetylated before hydroxylation

at another carbon[19]. The acetyl protecting group prevents off-pathway reactivity and is subsequently hydrolyzed by an esterase. Similarly, biochemical protecting group strategies can be used by the metabolic engineer to separately control the properties of a metabolite inside the cell versus outside the bioreactor.

In this specific application, I sought to gain temporal control over the regeneration of the indoxyl intermediate to achieve indigo crystallization within the cotton fibers, eliminating the need for a reducing agent (**Figure 2-1**). I leverage the work of previous groups to biosynthesize the indigo precursor indoxyl from tryptophan in *Escherichia coli*[3,4,11,12]. However, I also express a glucosyltransferase to stabilize indoxyl, preventing spontaneous dimerization by protecting the reactive hydroxyl with a glucosyl group. This glucoside, indican, is secreted from the cell into the fermentation broth and is sufficiently stable for long-term storage and concentration via gentle boiling. At the point of dyeing, indican can be enzymatically hydrolyzed (or deprotected) with a β-glucosidase to form indoxyl. Indoxyl spontaneously oxidizes to leucoindigo, and, as in the chemical process, leucoindigo crystallizes to indigo directly in the cotton fibers of the textile, where it adsorbs and is trapped.

2.2. Results

2.2.1. Identification of glucosyltransferase from *P. tinctorium*

The described biochemical protecting strategy requires a UDP-glucose:indoxyl glucosyltransferase (UGIG) to produce indican; however, none had previously been reported. Furthermore, at the outset of this project, no genome or transcriptome sequence had been available for any indigo-producing plant. To my knowledge, though transcriptome sequencing data have now been reported for two indigo plants[20-22], no UGIG amino acid sequence had been identified. I chose *P. tinctorium* for my enzyme discovery efforts because this Japanese indigo plant has been reported to be among the highest indican yielding plants[23,24] and there exists a published method for purifying its UGIG, active on indoxyl *in vitro*[17]. I adapted this protocol to purify the glucosyltransferase from *P. tinctorium* leaves. In parallel, cDNA isolated from the leaves was sequenced and a transcriptome was assembled *de novo* using the Trinity algorithm[25]. The purified glucosyltransferase was then analyzed by fragmentation and mass spectrometry, and fragments were matched to transcriptome-predicted sequences to identify the UGIG gene sequence.

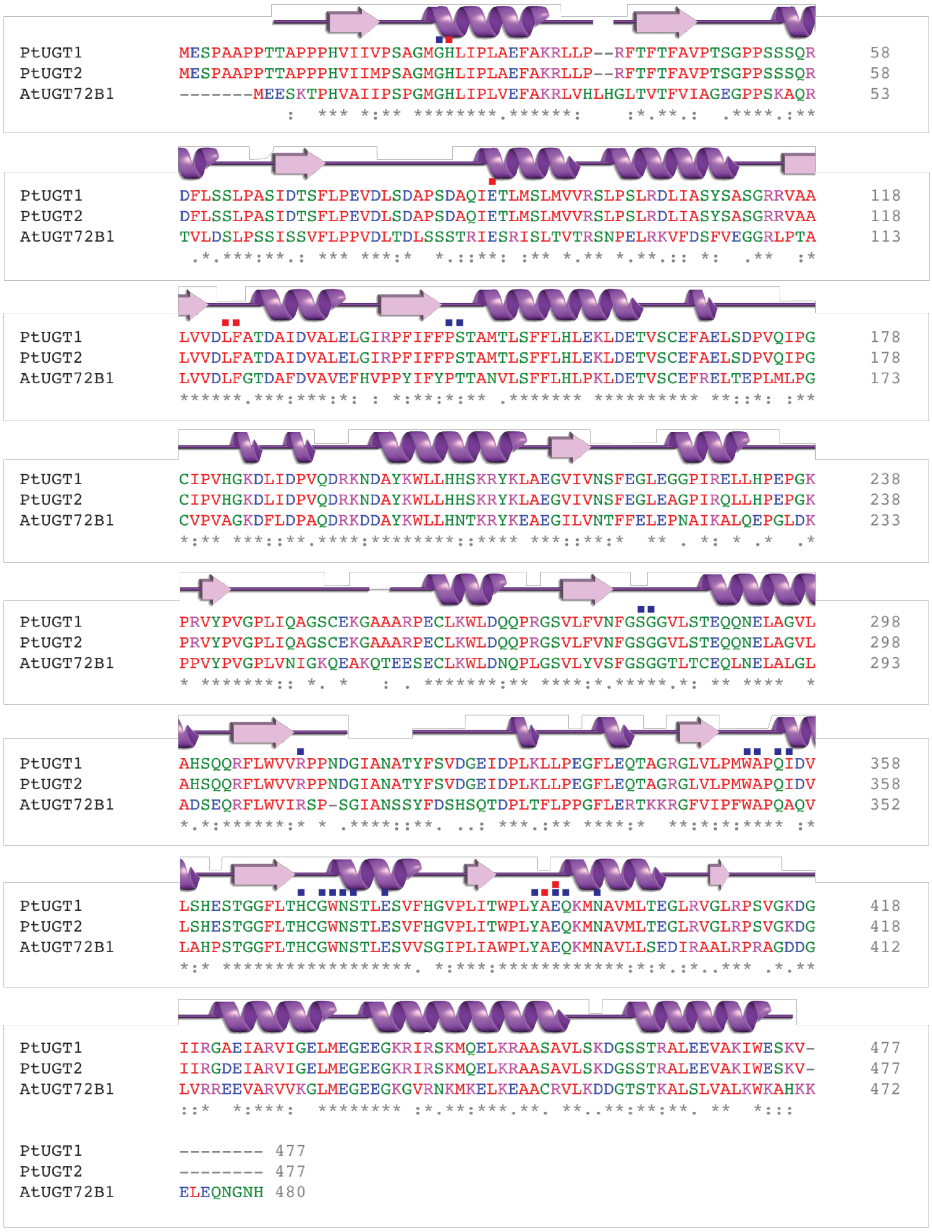

Figure 2-3. Multiple sequence alignment of UGT amino acid sequences. Clustal Omega[26] was used to generate a multiple sequence alignment of PtUGT1, PtUGT2, and the closest structurally characterized homolog, UGT72B1 from *Arabidopsis thaliana*. Red residues are small and/or hydrophobic; blue residues are acidic; magenta residues are basic; green residues have a hydroxyl, sulfhydryl, or amine. An asterisk (*) denotes a fully conserved residue; a colon (:) denotes a group of strongly similar residues; a period (.) denotes a group of weakly similar residues. PtUGT1 was used for structural studies and indican production in this work. The secondary structure of PtUGT1 (above alignment) was generated by PDBsum[27]. Red squares denote residues that contact indoxyl sulfate, based on the crystal structure. Blue squares denote residues that contact UDP-glucose, based on homology to other UGT crystal structures.

Sequence analysis showed that the identified glucosyltransferase (UGT72B29, referred to here as PtUGT1) belongs to the UDP-glycosyltransferase (UGT) superfamily[28] and family

GT1 of inverting glycosyltransferases in the Carbohydrate Active Enzymes database (CAZy) [29]. When amplifying the *PtUGT1* gene sequence from the cDNA, a second isoform of the glucosyltransferase gene was also identified, encoding a protein (UGT72B30, or PtUGT2) that differs from PtUGT1 by four amino acid residues (V19M, G225A, E230Q, A423D) (**Figure 2-3**, **Figure 2-5a**). When expressed and purified from *E. coli*, PtUGT1 and PtUGT2 were both active on indoxyl and had similar kinetic parameters *in vitro* (PtUGT1 k_{cat} = 9.1 ± 0.5 sec^{-1}, PtUGT2 k_{cat} = 12 ± 0.8 sec^{-1}) (**Figure 2-4**). PtUGT1 was used for the rest of this work because it was the predominant isoform in the transcriptome sequencing.

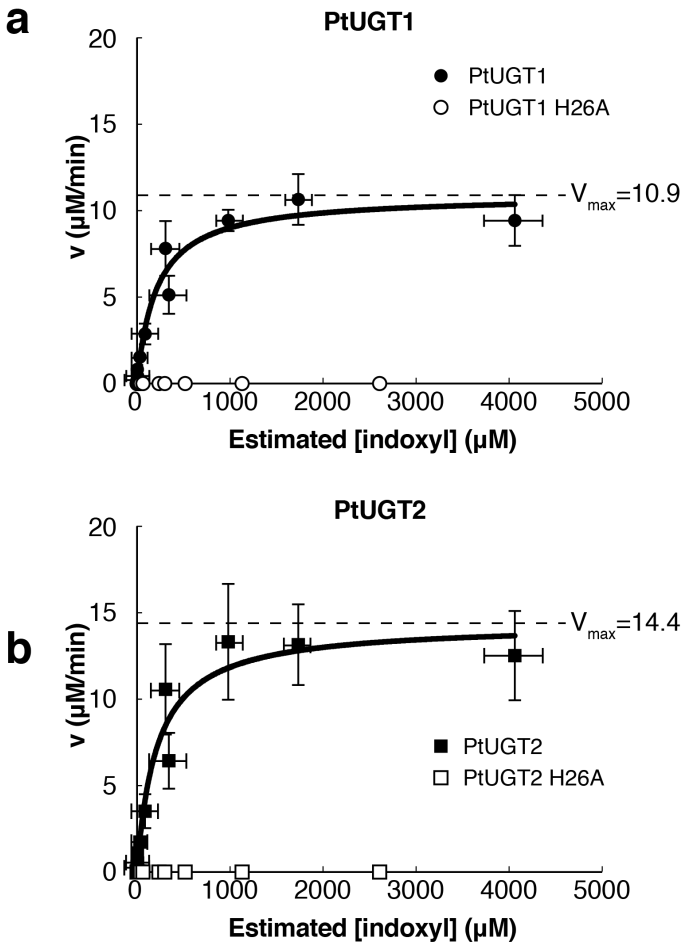

Figure 2-4. Michaelis-Menten curves. PtUGT1 (closed circles) has a k_{cat} of 9.1 ± 0.5 sec^{-1}, and PtUGT2 (closed squares) has a k_{cat} of 12 ± 0.8 sec^{-1}. The corresponding H26A mutants (open circles and squares) show no catalytic activity. The K_M of the enzymes cannot be accurately reported because the concentrations of the indoxyl substrate cannot be determined with high certainty due to the instability of the molecule. Error bars represent the mean ± s.d. of three technical replicates.

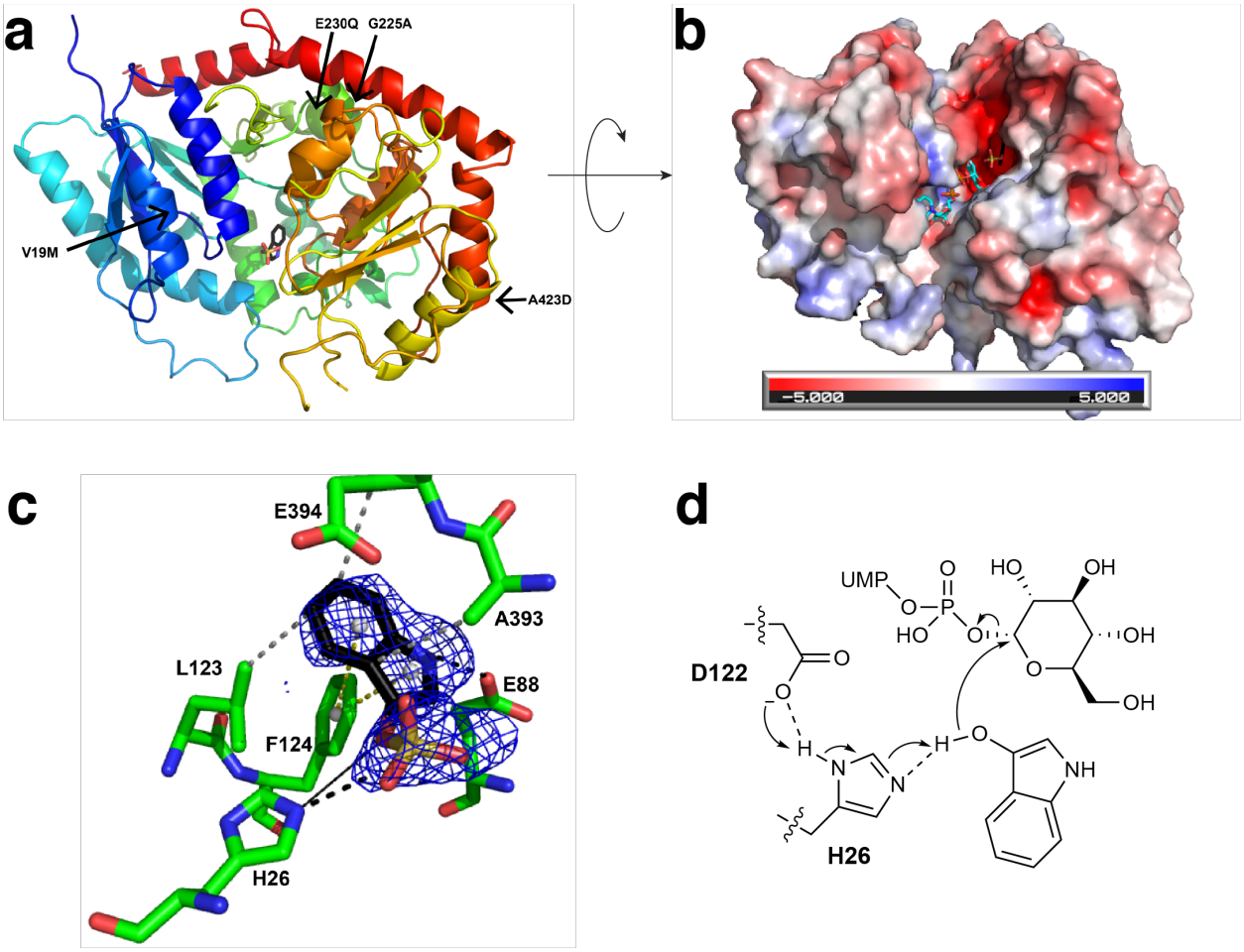

Figure 2-5. The crystal structure of PtUGT1 with bound indoxyl sulfate. Overall GT-B fold of PtUGT1 (PDB ID: 5NLM). The N-terminal Rossmann domain (blue) consists of a 7-stranded parallel β-sheet surrounded by 9 α-helical segments, and the C-terminal Rossmann domain (red) consists of a 6-stranded parallel β-sheet and 5 α-helices. The indoxyl sulfate bound in the active site is shown in stick representation. The structure is annotated with the four amino acids differing between PtUGT1 and PtUGT2 (arrows). b) Electrostatic surface potential of PtUGT1. The view is rotated ~90° around the x-axis with respect to panel a. Bound indoxyl sulfate is depicted in sticks (black) together with the donor substrate UDP-glucose superposed from the homologous AtUGT72B1 complex structure (cyan, PDB ID: 2VCE). c) The acceptor binding site with the bound indoxyl sulfate and interacting residues shown in stick representation. The indoxyl sulfate omit map is displayed as a blue grid contoured at 3.0 σ. Interactions are depicted as dashed lines (black = salt bridge/hydrogen bond, grey = hydrophobic interaction, yellow = π-stacking). A black solid line indicates the distance from the catalytic histidine to the glucose-accepting oxygen. d) In the proposed catalytic mechanism based on homology to other characterized UGTs[30,31], H26 deprotonates the indoxyl hydroxyl group, which then performs an S_N2 attack on the anomeric carbon of glucose. The conserved D122 is believed to balance the charge on the catalytic histidine. Consistent with this hypothesis, D122 forms a 2.6 Å hydrogen bond to H26 in the present structure.

2.2.2. Structural characterization of PtUGT1

To investigate the structural basis for indoxyl glucosylation by PtUGT1, the PtUGT1 crystal structure was solved at 2.14-Å resolution with indoxyl sulfate, a glycosylation incompetent indoxyl analog, in the active site (PDB ID: 5NLM, **Figure 2-5a**). PtUGT1 displays the canonical GT-B fold[32], consisting of two Rossmann domains connected with a linker. Of the four amino acids that differ between PtUGT1 and PtUGT2, the one closest to the active site (V19) is 12 Ångstrøm from indoxyl sulfate (**Figure 2-5a**). It is located in the core β-sheet of the N-terminal Rossmann domain. The remaining three substitutions are solvent exposed residues. In line with the kinetic observations, this structural analysis suggests none of the substitutions are likely to affect catalysis. Indoxyl sulfate binds in one end of the catalytic cleft between the two domains, similarly to other UGT acceptors[33] (**Figure 2-5b**). This rather large cavity excessively accommodates indoxyl rather than fits it snugly as in the classical lock-and-key model. Hence, structural determinants of indoxyl specificity can be speculated to be limited to those amino acids in direct contact with the indoxyl molecule (**Figure 2-5c**). In addition to hydrophobic interactions with residues L123, A393, and E394 and perpendicular π-stacking with F124, indoxyl sulfate interacts with the enzyme through a hydrogen bond between the indoxyl sulfate ring nitrogen and the conserved E88, which could therefore play a major role in indoxyl specificity and turnover. By analogy to other UGTs, H26 is expected to be the catalytic Brønsted base[30] (**Figure 2-5d**). This residue is oriented towards the glucose-accepting oxygen at a distance of 3.9 Ångstrøm. It is possibly displaced with respect to the catalytically competent conformation by the bulky sulfate group on the acceptor substrate mimic, with which it interacts through a salt bridge in the current structure. Indeed, the H26A mutant produces no detectable indican in an enzyme kinetics assay (**Figure 2-4**).

A search for PtUGT1 structural homologs reveals significant similarity (Z-scores above 30) to several plant UGTs belonging to GT1. The closest structural homolog is the AtUGT72B1 enzyme (Z-score 53.6, r.m.s.d. = 1.6 Å for 361 aligned residues), which is a bifunctional N- and O-glycosyltransferase implicated in xenobiotic metabolism in *Arabidopsis thaliana*[34]. Its structure has been solved in complex with a donor analog, UDP-2F-glucose (PDB ID: 2VCE)[31], revealing a donor binding mode very similar to that of other characterized UGTs. An extensive hydrogen bonding network is formed between all parts of the nucleotide sugar (i.e. the nucleobase, pentose, phosphates and sugar moiety) and enzyme residues primarily from the C-terminal Rossmann domain. In AtUGT72B1, 21 residues participate in this network. 19 of these are conserved in PtUGT1 and align well (r.m.s.d. = 0.6 for 131 aligned atoms) with the corresponding AtUGT72B1 residues (**Figure 2-6**). It is therefore likely that the donor binding sites in PtUGT1 and AtUGT72B1 are similar.

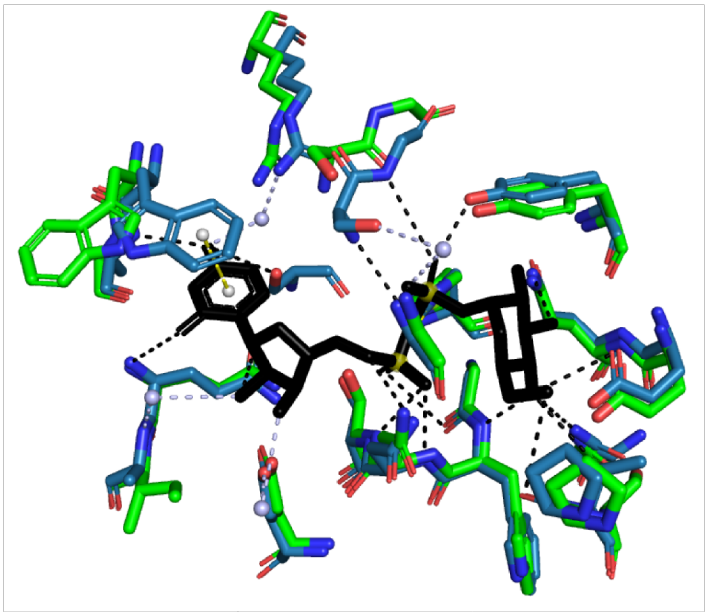

Figure 2-6. UDP-glucose binding pocket inferred from AtUGT72B1. The donor binding site of AtUGT72B1 (PDB ID: 2VCE). The bound UDP-2F-glucose (black), interacting residues (teal) and water molecules (grey) are shown, together with aligned PtUGT1 residues (green), illustrating the putative UDP-glucose binding site of PtUGT1. Interactions are shown as dashed lines (black = hydrogen bond, grey = water bridge, yellow = π-stacking).

2.2.3. Production of indican in *E. coli*

E. coli was chosen as the production host due to its short generation time, its ability to support high flux through engineered metabolic pathways, its use as an industrial production host for metabolic productsd[35-37], and its ability to efficiently glucosylate small molecules[38,39]. Furthermore, indoxyl production in *E. coli* has previously been accomplished at high titer for indigo biosynthesis[3,4,12].

An appropriate *E. coli* background strain and growth conditions were developed in order to test heterologous expression of PtUGT1 *in vivo*. Previously, the flavin-dependent monooxygenase (FMO) from *Methylophaga aminisulfidivorans* has been shown to efficiently convert indole to indoxyl[40], the precursor to indican and indigo (**Figure 2-10a**). The *FMO* gene was expressed on a plasmid under the control of the constitutive bacterial promoter J2310038 (ref. [41]) to test indoxyl production under different carbon sources. Previous reports have shown that the expression and activity of TnaA, required for conversion of tryptophan to indole in *E. coli*, is repressed by glucose[11,42] but not by glycerol[43]. Indeed, a glycerol-fed culture produced 400 mg/L indigo after 24 h growth in EZ Rich defined medium when an optimal concentration of tryptophan (over 15 mM, or 3.1 g/L) was co-fed, whereas a culture grown with glucose instead of glycerol produced

almost no indigo (**Figure 2-7**). Tryptophan supplementation is required for visible indigo production, as native tryptophan production is insufficient to sustain indoxyl production in the laboratory strain MG1655 grown in EZ Rich medium (containing only 0.1 mM tryptophan).

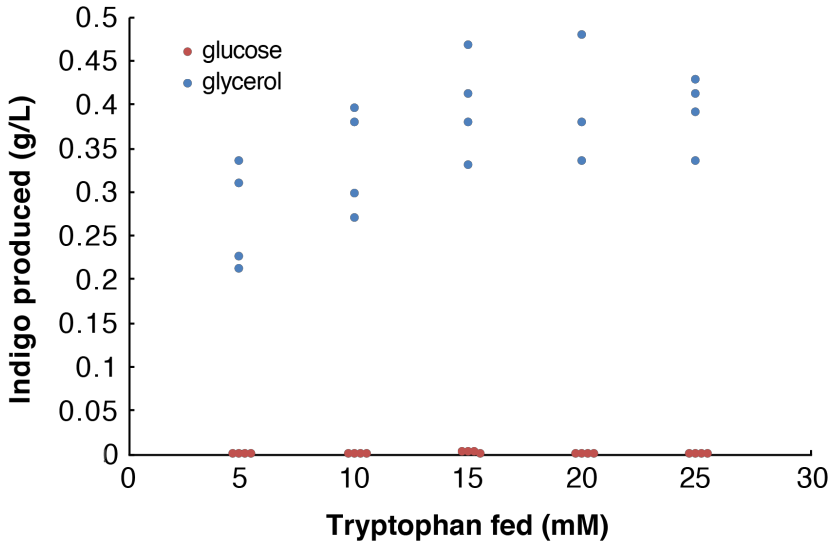

Figure 2-7. Carbon source affects indican production. When cells that constitutively express FMO are fed glucose, they produce significantly less indigo than when fed glycerol. This is most likely due to repression of TnaA expression and activity by glucose. FMO was expressed under a strong constitutive promoter (plasmid pTMH561) in strain background MG1655(DE3). Strains were grown for 24 h in EZ Rich media with 2% w/v glucose or 2% v/v glycerol carbon source. Data points represent independent cell cultures.

Before biosynthesizing indican, I also examined whether this product would be stable in the *E. coli* production host. Indican has previously been shown to be hydrolyzed by a number of β-glucosidases[44-46], and since *E. coli* contains several native β-glucosidases, indican could be hydrolyzed after production. To measure background hydrolysis, 1 g/L indican was fed to MG1655(DE3) *E. coli* and the loss of indican was quantified over 4 days (**Figure 2-8a**). I did observe background hydrolysis and hypothesized that it was due to one or more endogenous *E. coli* β-glucosidases. A panel of single-gene glucosidase knockouts from the Keio collection (background strain BW25113)[47] was tested for stability of fed indican (**Figure 2-9**). The only knockout that noticeably reduced hydrolysis was Δ*bglA*, and its effect was confirmed in my bacterial host, MG1655(DE3). There was considerably less indican degradation in the Δ*bglA* strain than in wild type, and even after 5 days of incubation at 37°C, there was no visible indigo formation in the Δ*bglA* strain (**Figure 2-8b**).

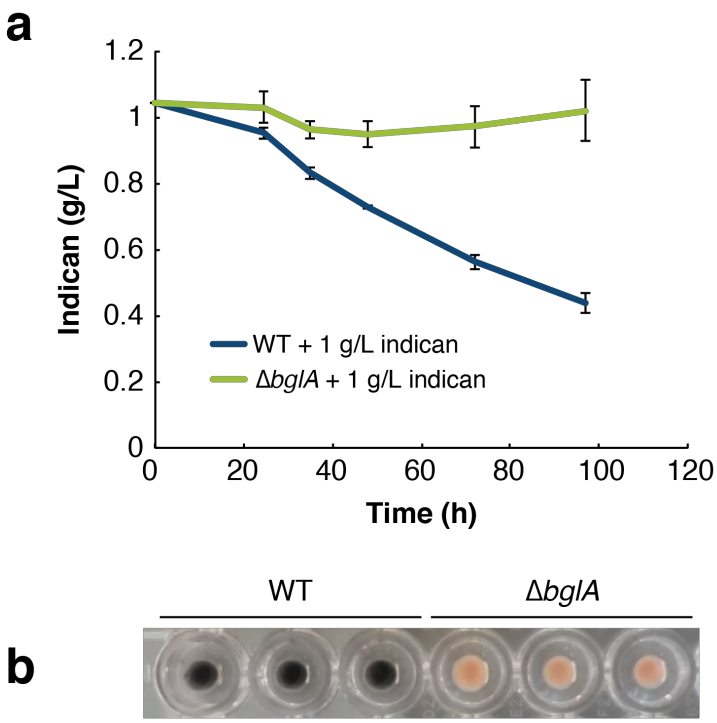

Figure 2-8. Background hydrolysis of indican by wild type vs Δ*bglA* *E. coli*. a) After 97 h incubation with 1 g/L indican at 37 °C, wild type (WT) MG1655(DE3) *E. coli* grown in EZ Rich media showed considerable indican hydrolysis, whereas the same strain with *bglA* knocked out did not appear to reduce indican titers. Error bars represent the mean ± s.d. of three biological replicates. b) Even after 120 h incubation, very low amounts of indigo were visible in the *ΔbglA* strain (right 3), whereas ample color was observed in the WT strain (left 3).

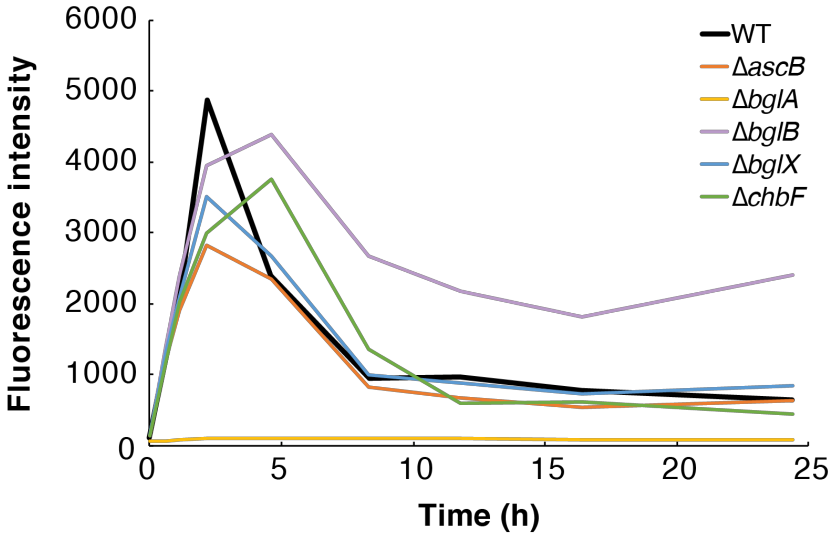

Figure 2-9. Panel of *E. coli* β-glucosidase knockouts. A panel of five *E. coli* β-glucosidase knockouts from the Keio collection (background BW25113; ref. [47]) were tested for their ability to hydrolyze 1 mM indican over 24 h. Cultures were grown to saturation and incubated with 1 mM indican in 1x PBS + 20% DMSO. Indoxyl is fluorescent[48], and its fluorescence intensity was measured with excitation and emission wavelengths of 410 nm and 490 nm, respectively.

The *PtUGT1* gene and the *FMO* gene were then co-expressed in MG1655Δ*bglA* for *in vivo* indican production (**Figure 2-10a**, **Figure 2-11**). The two-gene operon was driven by P_{T7} for high gene expression and integrated into the chromosome to minimize copy number variability (strain TMH011). In addition to producing indican, the strain also visibly showed dramatically lower indigo production compared to the same MG1655Δ*bglA* background strain expressing only *FMO* (strain TMH003, **Figure 2-10b**), consistent with the prediction that glucosylation prevents indoxyl dimerization and oxidation into indigo crystals. Indican production, as quantified by LC-MS, was extremely efficient, with 97% of tryptophan converted into indican on a molar basis (**Figure 2-10c**). Addition of purified β-glucosidase BglA from *Bacillus circulans*[49] to the extracellular medium resulted in visible indigo precipitation (**Figure 2-10b**). Thus, indican is both produced and secreted from the engineered *E. coli* strain TMH011.

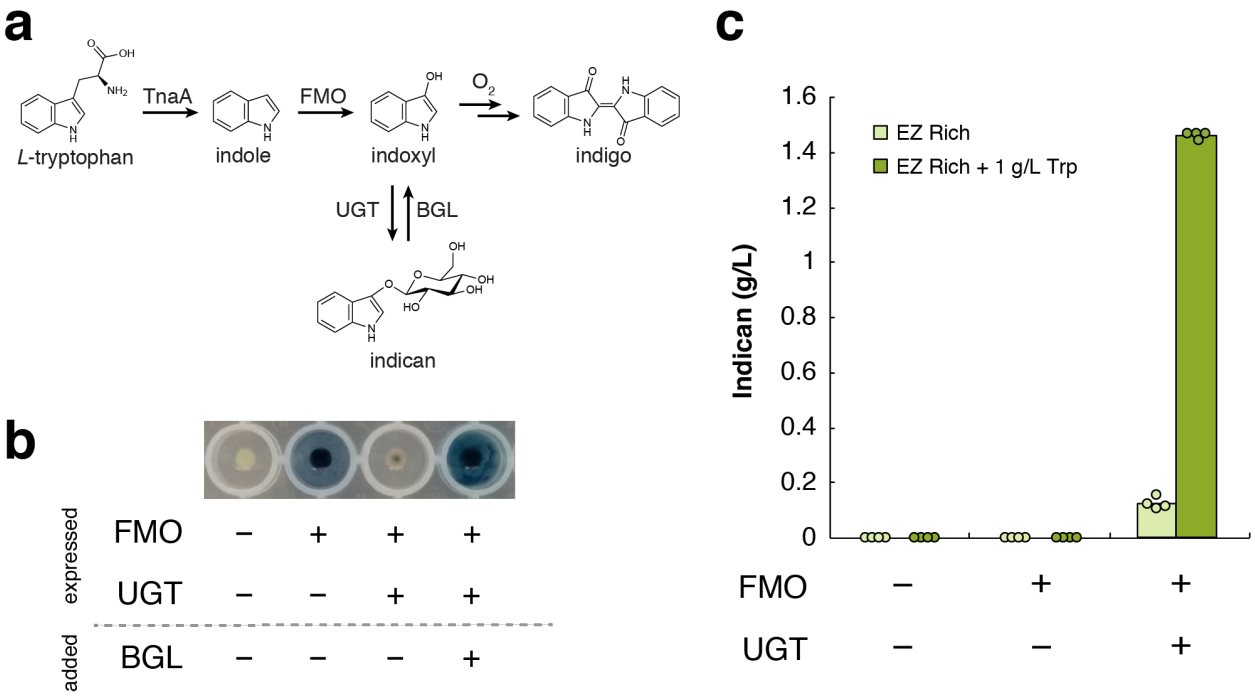

Figure 2-10. Heterologous expression of PtUGT1 stabilizes indoxyl before it dimerizes, producing indican. a) Tryptophan is converted to indole by the native *E. coli* enzyme TnaA, oxygenated to indoxyl by FMO, and glucosylated by PtUGT1. A β-glucosidase hydrolyzes indican into indoxyl, which spontaneously oxidizes to indigo. b) Expression of PtUGT1 reduces indigo formation in an FMO-expressing Δ*bglA* strain, but upon addition of β-glucosidase indigo precipitate forms. c) Indican can be produced in a Δ*bglA* strain by heterologous expression of both FMO and PtUGT1, and titers improve with exogenous feeding of more tryptophan than the 20 mg/L present in EZ Rich media. Bars represent the mean of four biological replicates, and each replicate is represented by a circle.

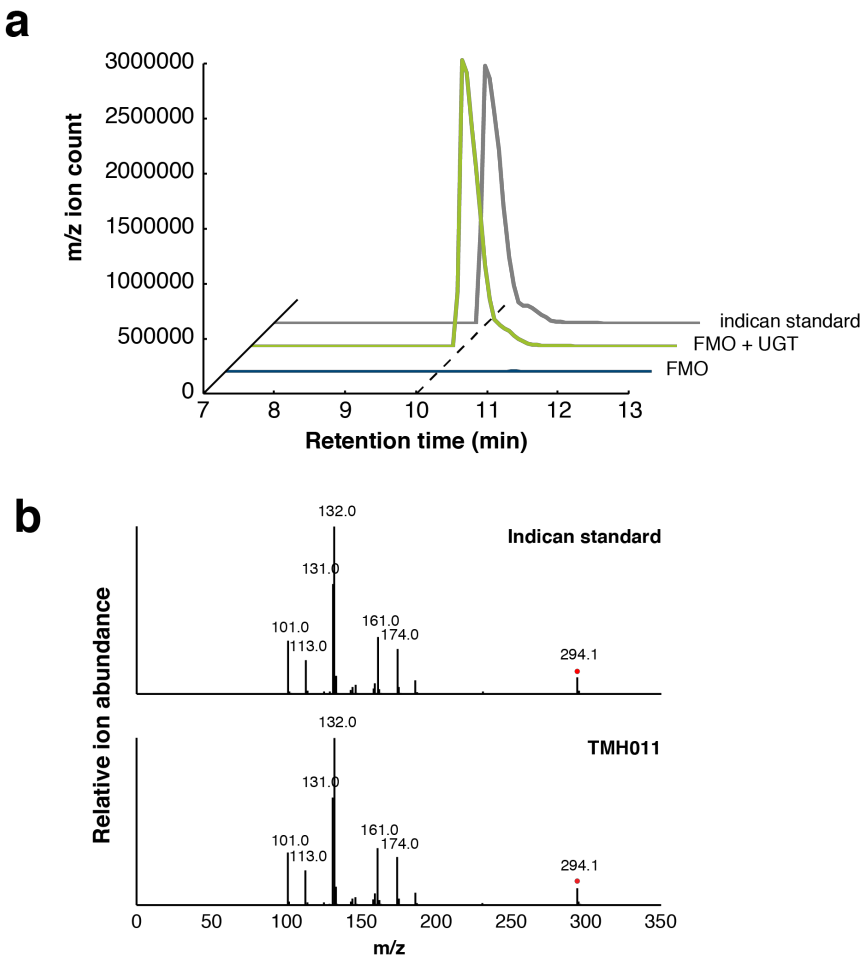

Figure 2-11. Detection of indican by LC-MS. a) Biosynthesized indican made from FMO and UGT has the same retention time as the commercial standard. Indican is only formed in the presence of both FMO and UGT. b) The MS/MS spectra of biosynthesized indican (obtained by collision induced dissociation at 10 V) matches that of the commercial standard. The red dot denotes the parent $[M-H]^-$ ion.

The efficiency of PtUGT1 can also be seen in comparing titers of indican in TMH011 (9.8 mM, or 2.9 g/L) to titers of indigo in TMH003 (5.3 mM, or 1.4 g/L) when grown for 24 h in media with excess (26 mM, or 5.4 g/L) fed tryptophan (**Figure 2-12a**, **Figure 2-12b**). Since indigo is an indoxyl dimer and thus requires twice as many moles of indoxyl substrate as indican, these titers are approximately equivalent on an indoxyl molarity basis, with a 38% molar conversion of tryptophan to indican and 41% molar conversion to indigo. Thus, indoxyl glucosylation appears to be competitively more efficient than indoxyl oxidation up to the titers achieved, consistent with the observation that only low amounts of indigo byproduct are made when expressing PtUGT1 (**Figure 2-12b**) while tryptophan is fully utilized (**Figure 2-12c**). Under the current production regime, the P_{T7}-driven expression of FMO and PtUGT1, as well as their metabolic products, do not appear to cause toxicity to the cells (**Figure 2-12d**).

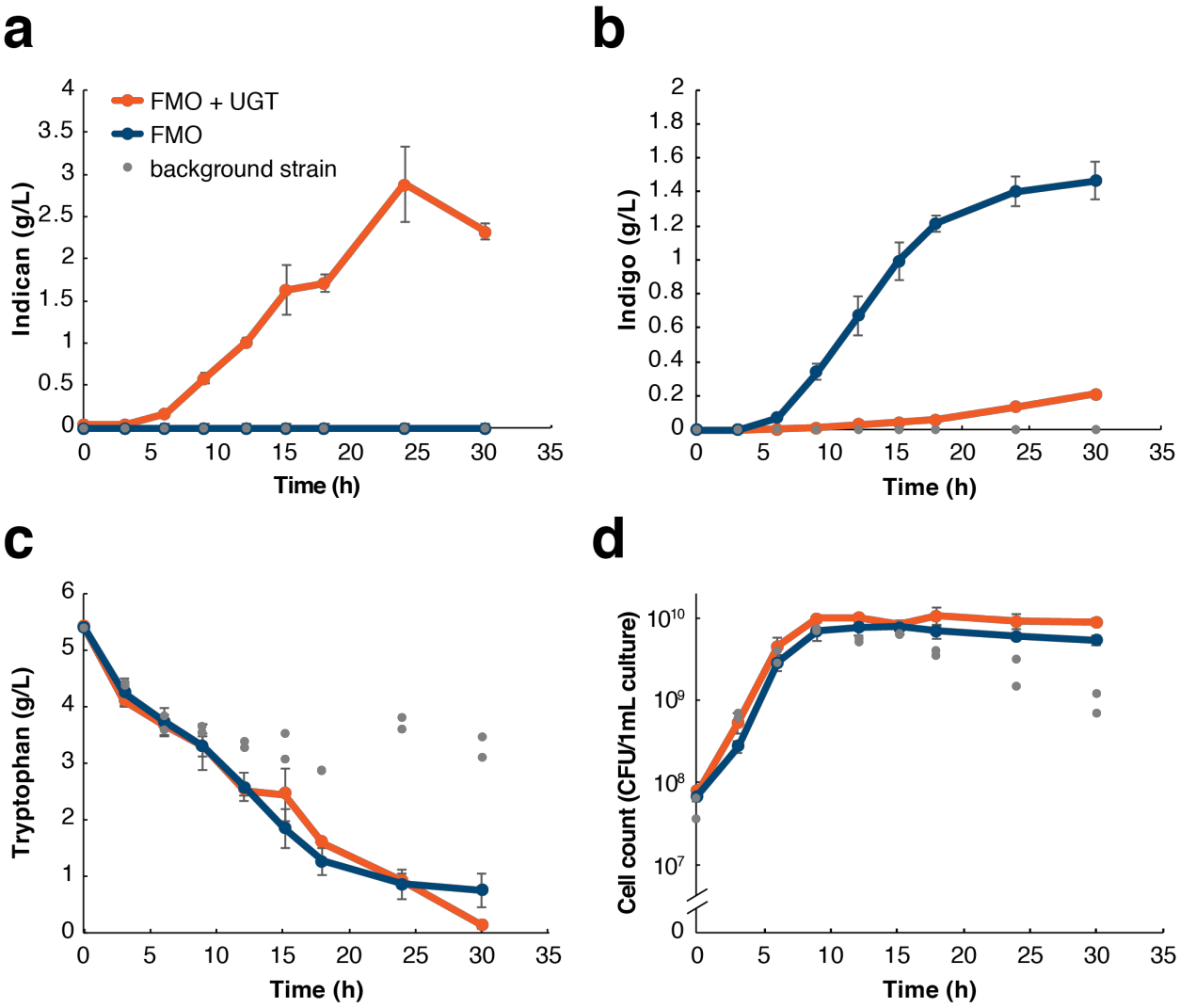

Figure 2-12. Production and growth curves for indican and indigo production. Titers of a) indican, b) indigo, c) and tryptophan, and d) cell count (measured as colony-forming units) over a 37 h growth period. Cells expressing FMO and PtUGT1 are represented by orange; cells expressing solely FMO are represented by blue; cells expressing neither gene are represented by grey. Error bars represent the mean ± s.d. of six (FMO + UGT) or four (FMO) biological replicates; circles represent individual biological replicates.

2.2.4. Use of biosynthesized indican as a textile dye

The current indigo dye mill process consists of repeatedly dipping ropes of cotton yarn into multiple dye baths containing an aqueous solution of leucoindigo, sodium hydroxide, and sodium dithionite. Exposure to air between dips allows spontaneous oxidation of the leucoindigo, depositing indigo in the fibers[2]. I propose a dyeing process in which an indican solution and a β-glucosidase solution are alternately applied to the textile, resulting in indican hydrolysis. The resulting indoxyl spontaneously oxidizes and

dimerizes in air to form leucoindigo and, subsequently, indigo. In this methodology, I apply indican as a spray to prevent contamination of the indican solution with β-glucosidase, which would result in premature hydrolysis (**Figure 2-1**).

This dyeing process requires the biosynthesis of indican and a β-glucosidase capable of hydrolyzing indican in cotton fibers. The gene encoding the *Bacillus circulans* β-glucosidase BglA was expressed under the control of the T7 promoter in the *E. coli* Rosetta BL21(DE3) strain to produce lysate with an indican hydrolysis activity of 1.8 ± 0.8 μmol/min/mg. In parallel, to produce a sufficient amount of indican to test the dyeing process, a 14 L fermenter was used to grow 5 L of indican-producing *E. coli*. The fermentation broth was used to dye fabric without the need for indican purification. Since higher dye concentrations are correlated with darker shades of indigo, the broth was gently boiled to increase the indican concentration 2.9-fold from 1.1 g/L to 3.2 g/L.

An unexpected orange color was produced in the media when biosynthesizing indican. This orange color required FMO expression and was more intense with co-expression of PtUGT1 (**Figure 2-13a**). However, this orange color is water-soluble as it is efficiently removed in both water and detergent washes (**Figure 2-13b**) and, accordingly, does not affect the final dyed product.

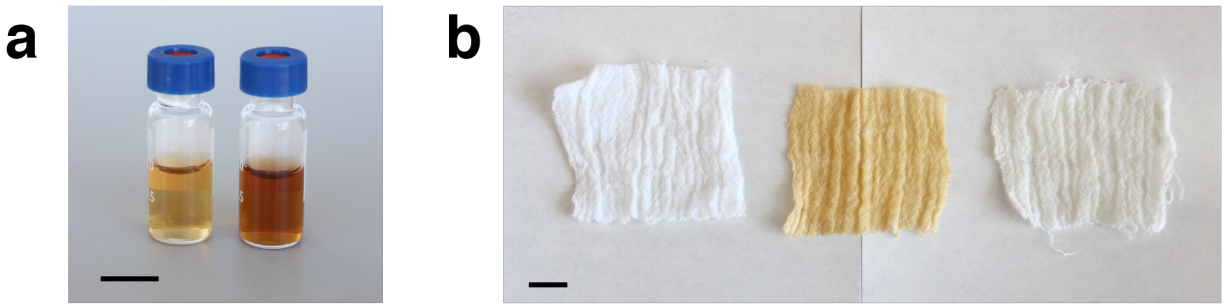

Figure 2-13. Orange color co-produced with indican washes off of cotton. a) An unexpected orange color is produced when FMO is expressed, and the color intensifies when PtUGT1 is co-expressed. Left: supernatant of TMH003 24h after induction with IPTG. Right: supernatant of TMH011 24h after induction with IPTG. b) The media shows an orange color when FMO and PtUGT1 are expressed, but it washes out of fabric. Left: white cotton gauze with EZ Rich media applied. Middle: white cotton gauze with indican broth applied (see "Indican production in bioreactor" in **Materials and Methods**). Right: gauze fabric applied with indican broth was allowed to air dry, and it was then washed in a cold-water laundry cycle with All detergent. Scale bars, 1 cm.

To demonstrate the indican dye strategy and compare it to the current, chemically-reduced indigo dyeing process, swatches of white cotton denim cloth were dyed using concentrated fermentation broth containing 3.2 g/L indican (**Figure 2-14a**). A solution of 3.2 g/L purchased indican standard was also used to examine the effect of indican purity

on dyeing. The swatches were dyed by spraying with indican, dipping into β-glucosidase cell lysate, and oxidizing in air. An equimolar chemically reduced indigo dye bath of 1.4 g/L indigo (considering that two indoxyl molecules are required to make one leucoindigo molecule) was used to mimic the conventional industrial process for comparison. An unreduced indigo vat did not visibly dye cloth, but the reduced-indigo dye was effective, as were the indican dyes after deprotection. Although the indican-dyed cloths did not dye as dark as the conventionally-dyed cloths, they remained blue even after a water wash (**Figure 2-14a**). The difference in color intensity is likely attributable to the higher pH used for the chemical dyeing process, as indigo dyeing is more efficient under alkaline conditions than at the neutral pH used for the indican dyeing.

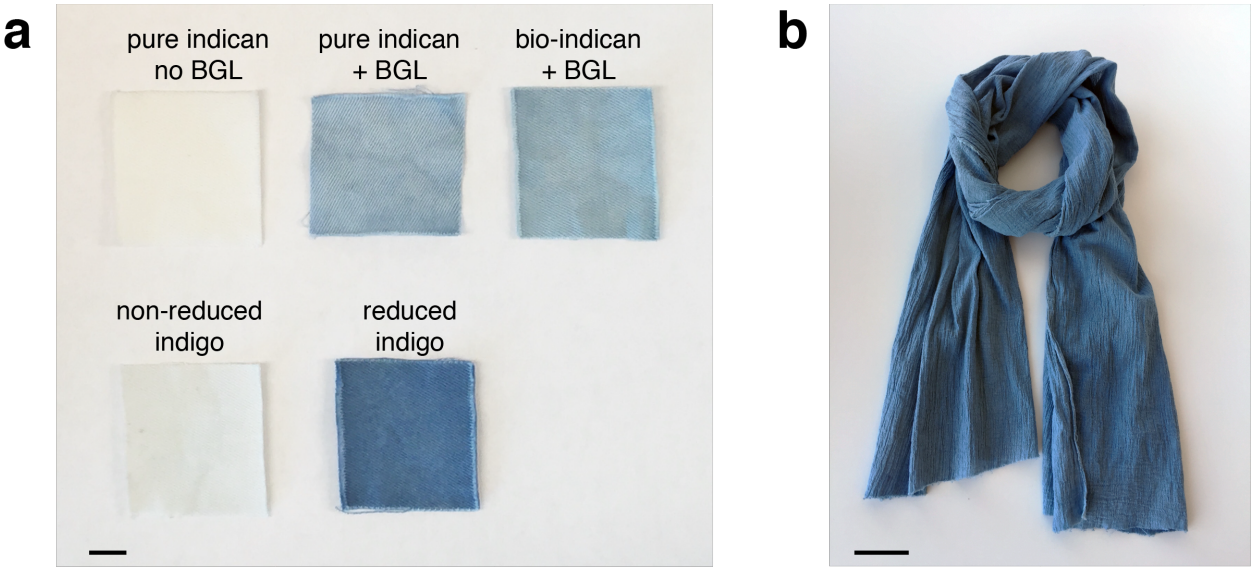

Figure 2-14. Bio-indican can be used as an effective, reductant-free cotton textile dye. a) Top row: Pure indican with no β-glucosidase (BGL); pure indican with β-glucosidase; bio-indican with β-glucosidase. Bottom row: Indigo, non-reduced; indigo, reduced with sodium dithionite. All swatches are dyed on white cotton denim. Scale bar, 1 cm. b) 100% white cotton gauze fabric, dyed with indican. All samples were photographed after numerous, vigorous water washes. Scale bar, 5 cm.

The concentrated bio-indican media and β-glucosidase lysate were used to dye a garment as a proof-of-concept: a cotton gauze fabric scarf. A total of three dye cycles were applied to each side of the fabric to achieve a deep indigo color, where one cycle is defined as an indican broth spray application followed by a β-glucosidase lysate spray application. Color lasted even after the fabric was vigorously rinsed in water until the runoff water was clear (**Figure 2-14b**). The scarf was subsequently laundered in a commercial laundry machine with detergent, and although some intensity was lost, ample color was maintained (**Figure 2-14b** for pre-machine wash, **Figure 2-15** for post-machine wash).

Figure 2-15. Indican dyed scarf retains blue color after laundry wash. Although some surface-adsorbed indigo is lost after washing (compare to **Figure 2-14b**), much of the blue color is retained. Scale bar, 5 cm.

2.3. Discussion

The indigo dyeing process has undergone minimal changes for a hundred years, yet demand for the dye is higher than ever before, making its ecological consequences unsustainable. Here, I report a promising novel strategy for dyeing that not only removes the need for the hazardous chemicals required for indigo synthesis, as previously addressed with indoxyl production in *E. coli*[3,4,12], but also eliminates the requirement for the environmentally-unfriendly reducing agent used for solubilization by instead utilizing a glucosyl biochemical protecting group. I identified the gene sequence of a novel glucosyltransferase from *P. tinctorium* (PtUGT1) with activity on indoxyl, characterized the enzyme structurally, and successfully co-expressed it with the oxygenase FMO in *E. coli* to produce up to 2.9 g/L indican *in vivo* from fed tryptophan. This bio-indican was shown to be an effective fabric dye when hydrolyzed by a β-glucosidase.

Addition of the glucosyl protecting group onto indoxyl confers several favorable properties. First, indican is stable in air until incubated with a β-glucosidase, providing control over when and where the indoxyl molecule will be deprotected for dyeing. Second, the hydrophilic glucosyl group keeps indican soluble in water, the preferred solvent of dye mills. Third, the stability of the glucoside bond allows not only long-term storage but also concentration via gentle boiling of the crude fermentation broth. Thus, indican can be produced by a batch, or eventually continuous, fermentation. Product could be purified or even used directly as concentrated spent media for dyeing cotton. A conventional leucoindigo bath is typically around 4 g/L (ref. [2]), so a fermentation producing an equivalent concentration, based on indoxyl molarity, of ~9 g/L indican would not require additional concentration.

Glucosylation has previously been employed as a metabolic engineering tool for changing the chemical and biological properties of the aglycone. Vanillin titers in *Schizosaccharomyces pombe* have been improved through glucosylation. Vanillin-glucoside has improved solubility and lower toxicity; consequently, it can be produced at considerably higher titers in *S. pombe* without growth defects[50]. In fact, several glucosides have been shown to be less toxic than their aglycone forms[50,51]. The glucose moiety can increase water solubility of many metabolites; in addition, this modification may block either spontaneous or enzyme-catalyzed reactions (by obstructing substrate recognition) that produce toxic metabolites. Thus, glucosylation promises to provide the metabolic engineer a valuable tool for achieving desired biochemical properties of a molecule in the cell, similar to how an organic chemist utilizes protecting groups to modify specific functional groups in an appropriate sequential manner. The ability to hydrolyze the glycosidic bond with a β-glucosidase enables reversible control, as the glucosyl group can be removed enzymatically when and where desired.

A glucosyl protecting group poses some practical challenges for incorporation into a dye mill process that must be addressed. Although indican's stability is an advantage for storing and shipping, it requires an enzyme or strong acid for hydrolysis. A strong acid is challenging to incorporate into the mill process as the dyeing properties of leucoindigo are optimal at a basic pH, conventionally around pH 11 (refs. [10] and [52]). Although the use of β-glucosidase to hydrolyze the glucosyl protecting group provides many advantages of control, there are a couple of challenges for adapting it into an industrial process. First, use of an enzyme increases the cost and process time for the dyeing procedure. Second, the use of a six-carbon glucosyl moiety as the protecting group means that a large portion of the feedstock carbon will not go into the final dye molecule, instead comprising a moiety that will ultimately be removed. Commercialization will require further improvements; however, it is reasonable to expect these to be achievable. First, a more

active and cheaply produced β-glucosidase must be obtained. Since β-glucosidases are mass-produced by filamentous fungi hosts such as *Aspergillus niger* for biofuel production from cellulosic material[53,54], inexpensive industrial production of an engineered β-glucosidase seems feasible. Furthermore, engineering a high activity β-glucosidase for faster hydrolysis and activity under higher pH conditions should be possible by utilizing the formation of visible indigo color for high-throughput screens. Similar screens can be employed for screening strain modifications increasing flux to indoxyl directly from glucose rather than relying on bioconversion from tryptophan.

The use of enzymes to add biochemical protecting groups site-specifically to small molecules in the cell, with subsequent removal either inside or outside the cell, could prove similarly powerful to synthetic biologists as chemical protecting groups are for synthetic organic chemists. Indigo provides a motivating case study where nature has provided a blueprint. Following this model has provided a promising new dyeing strategy offering a much-needed update to the historic, but unsustainable, indigo dyeing process.

2.4. Materials and Methods

cDNA library construction

P. tinctorium plants (Companion Plants, Athens, OH) were grown in a laboratory environment. Samples of leaf tissue were taken from live plants at several points during the day and frozen in liquid nitrogen. These samples, a total of 300 mg, 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 100 mg of powdered frozen leaf tissue using the Qiagen RNeasy Plant Mini kit obtained from Qiagen GMBH (Hilden, Germany). 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.

Transcriptomics

Paired-end reads received from the Illumina HiSeq2500 sequencer were first trimmed to remove low-quality reads using the Trimmomatic software package[55] 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[56] 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[57] 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[25,58] in paired-end mode or Oases using k-mer sizes between 17 and 31[59]. 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[25,58]. This work was done on the XSEDE Blacklight system[60,61].

UGIG purification

The UGIG purification protocol was adapted from Minami *et al.*[17]. 200 g leaves were harvested from the *P. tinctorium* plant grown in a greenhouse, flash frozen in liquid nitrogen, and ground to a fine powder. 400 mL extraction buffer (100 mM potassium phosphate pH 7.0, 2 mM (EDTA), 20 mM β-mercaptoethanol (BME), 1 tablet cOmplete protease inhibitor (obtained from Roche)) was added to the ground leaves and the mixture was thawed at 4° C for 30 min. The slurry was then centrifuged twice at 24,000 × g for 30 min at 4° C. 3 parts PEG solution (50% (w/v) PEG 6000, 100 mM potassium phosphate pH 7.0, 2 mM EDTA, 20 mM BME) was added to 2 parts supernatant, and the mixture was centrifuged at 143,000 × g. The supernatant was dialyzed against equilibration buffer (50 mM HEPES pH 7.0, 1 mM EDTA, 10 mM BME).

The supernatant was run through DEAE Sepharose Fast Flow beads (GE Healthcare) and eluted with a solution of 50 mM HEPES pH 7.0, 100 mM NaCl, 1 mM EDTA, and 10 mM BME. Fractions were assayed for the presence of a UGIG (see "Glucosyltransferase assay for purification"), as they were for all subsequent chromatography steps. The desired fractions were then run through hydroxyapatite beads (Sigma), eluting with buffer containing 10 mM potassium phosphate pH 7.0, 50mM HEPES-NaOH pH 7.0, 1 mM

EDTA, 5 mM DTT, and 10% v/v glycerol. The fractions showing glucosyltransferase activity were then applied to a Mono Q 5/50 GL column (GE Healthcare) and eluted with a NaCl gradient from 0 to 200 mM. Lastly, the protein was concentrated using a 10,000 MWCO spin concentrator and run through a Superdex 200 10/300 GL column (GE Healthcare), eluting with a buffer made of 50 mM HEPES pH 7.0 and 5 mM DTT.

The protein preparation was electrophoresed on a NuPAGE Novex 4-12% Bis-Tris Protein Gel (Life Technologies) at 140 V for 75 min. The sequential rounds of chromatography yielded a highly enriched protein around 52 kDa in weight, similar to the previous report[17]. The band on the gel corresponding to the glucosyltransferase was extracted for trypsin digestion and HPLC 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, BLASTX-limited sequences (see "Transcriptomics") using SEQUEST[62] and DTAselect[63]. The sequences with the most complete coverage of matching peptide fragments were selected for further study.

Glucosyltransferase assay for purification

After each purification step, a glycosyltransferase assay was performed, using a protocol adapted from that of Minami *et al*[17]. 100 μL protein sample was incubated in 100 mM glycine-NaOH pH 9 with 2 mM UDP-glucose, 2 mM indoxyl phosphate, 0.1 mg/mL potato acid phosphatase, 100 mM DTT, and 10 mM ascorbic acid, in a total volume of 200 uL total volume. The mixture was incubated at 37°C for 3 h. The reaction was terminated by adding 140 μL cold 40% (w/v) trichloroacetic acid, centrifuging at max speed for 5 min, and then adding 200 μL 500 mM NaOH. The sample was assayed for the presence of indican by LC-MS (see "Determination of metabolite concentrations").

Plasmids, strains, and growth media

Plasmids were constructed using a MoClo Golden Gate Assembly[64] and propagated using *E. coli* strain TG1 (Lucigen). Strains for plasmid construction were grown in Luria Broth (LB) selected on 34 mg/L chloramphenicol, 100 mg/L ampicillin, and/or 25 mg/L kanamycin.

Unless otherwise noted, final production strains are based on *E. coli* strain MG1655 containing the DE3 prophage (F⁻ *ilv*G⁻ *rfb-50 rph-1 λ*(DE3 [*lac*I *lac*UV5-T7 *gene 1 ind*1 *sam7 nin*5])), provided by the Michelle Chang laboratory (University of California, Berkeley)[65]. Genomic integrations to introduce heterologous genes were done by *λ* red

recombination[66] immediately downstream of base 44344920 in the *E. coli* genome. Oligos 1 and 2 (**Table 2-1**) were used to amplify the gene of interest for integration. *bglA* was knocked out of the genome by replacing the open reading frame with a KanR antibiotic marker. KanR, flanked by FRT sites, was amplified from the Δ*bglA* Keio knockout strain (background: BW25113)[47] using Oligos 3 and 4. This PCR product was integrated into MG1655(DE3) and selected on LB with kanamycin. The KanR gene was subsequently removed using pFLP2[67], and the plasmid was cured from the strain with 5% sucrose. All strains and plasmids used in this work are listed in **Table 2-2** and **Table 2-3**.

Oligo #	Oligo Name	Description	Sequence
1	ET55	Integrate at base 44344920 – forward	CCGGATAAGGAATTCACGCCGCATCCGGCAT CAACAAAGCACTGGCCGATAATTGCAGAC
2	ET56	Integrate at base 44344920 – reverse	AGCTACGGCGCTTTGGCTTGATAACCGGATAA CAACTTGCCTGATCCTTCAACTCAGCAA
3	CA78	Knock out *bglA* – forward	CACCAGCCCAACGATACC
4	DG75	Knock out *bglA* – reverse	ATCAGCAAAACCTCCACGCG

Table 2-1. List of oligos used in this work.

Strain Name	Parent Strain	Accession Number	Description	Antibiotic Marker	Used in Figure
MG1655(DE3)			from Bond-Watts et al., 2011 (ref. [65])	N/A	Figure 2-7, Figure 2-8
MG1655(DE3) ΔbglA	1			N/A	Figure 2-8, Figure 2-10, Figure 2-12
TMH003	2	MF688770	pT7-FMO-tT7	Kan^R	Figure 2-10, Figure 2-11, Figure 2-12
TMH011	2	MF688771	pT7-FMO-PtUGT1-tT7	Kan^R	Figure 2-10, Figure 2-11, Figure 2-12, Figure 2-13, Figure 2-14, Figure 2-15

Table 2-2. List of strains used in this work. Gene cassette was integrated into TMH003 and TMH011 at base 44344920 in the *E. coli* genome.

Plasmid Name	Accession Number	Description	Antibiotic Marker	Purpose	Used in Figure
pTMH307	MF688772	pT7-6xHis-PtUGT1-tT7	Amp^R	Purification of PtUGT1	Figure 2-4, Figure 2-5, Figure 2-6, Figure 2-16
pTMH308	MF688773	pT7-6xHis-PtUGT2-tT7	Amp^R	Purification of PtUGT2	Figure 2-4, Figure 2-16
pTMH634		pT7-6xHis-PtUGT1_H26A-tT7	Amp^R	Purification of PtUGT1 H26A	Figure 2-4, Figure 2-16
pTMH635		pT7-6xHis-PtUGT2_H26A-tT7	Amp^R	Purification of PtUGT2 H26A	Figure 2-4, Figure 2-16

pRLP121	MF688774	pT7-6xHis-BglA-tT7	Amp^R	Expression of *B. circulans* glucosidase BglA	Figure 2-10, Figure 2-14, Figure 2-15
pTMH561	MF688775	BBa_J23100-BBa_B0034-FMO-tSPY	Kan^R	Constitutive expression of FMO	Figure 2-7

Table 2-3. List of plasmids used in this work. All plasmids have the ColE1 origin of replication. BBa_J23100: ref. [41]. tSPY: ref. [68].

UGT crystallographic structure solution

Expression & purification of PtUGT1 for crystallization

Rosetta™(DE3) cells (Novagen) were transformed with pTMH307 and selected on LB agar plates containing chloramphenicol and ampicillin. Overnight cultures were diluted into 4 L Terrific Broth (Fisher) with ampicillin selection, grown at 37° C in an Innova 44 shaker (New Brunswick Scientific) at 200 rpm, and induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} ~3. Cultures were then grown at 18° C for 21 h for protein expression and the cells were harvested by centrifugation. The cell pellets were resuspended in 50 mM HEPES pH 7.0, 300 mM NaCl, 25 mM imidazole pH 8.0, and 1 mM DTT. The cell suspension was lysed by freeze/thaw and sonication. Lysate was purified using Ni-NTA agarose beads (Qiagen), and the protein was dialyzed against 25 mM HEPES pH 7.0, 50 mM NaCl, and 1 mM DTT. The N-terminal 6xHis tag was TEV-cleaved, and the final purified protein was concentrated to 15 mg/mL using a 30,000 MWCO Amicon Ultra-15 Centrifugal Filter Unit (EMD Millipore). Protein concentration was determined using the Bradford assay in triplicate (Bio-Rad).

Crystallographic structure determination and analysis

PtUGT1 was co-crystallized with indoxyl sulfate (Sigma-Aldrich I3875) in sitting drops consisting of 0.2 μL protein solution (15 mg/mL in 25 mM HEPES pH 7.0, 50 mM NaCl, 1 mM DTT, 1 mM indoxyl sulfate) and 0.2 μL crystallization buffer (MCSG-1 screen (Anatrace) solution 9: 0.2M $MgCl_2 \bullet 6H_2O$, 0.1 M HEPES pH 7.5, 25% PEG 3350). The drops were set up with a Phoenix crystallization robot in 3-drop IntelliPlate 96-well crystallization plates (Art Robbins Instruments). Crystals appeared after 2 days and grew to their final size after ~4 days. These crystals were cryoprotected in 10% glycerol and mounted in a nylon cryoloop (Hampton Research). 270° of data were collected at 100 K, 1.0000 nm, at the Berkeley Center for Structural Biology beamline 5.0.3 of the Advanced Light Source in Berkeley, California with a 1° oscillation and 5 s exposure time and an ADSC Q315r CCD detector. The data were processed with Xia2[69] and XDS[70]. The structure was solved by molecular replacement using PDB ID 2VCE[31] as a search model and Phaser[71] from the Phenix software package[72]. The structure was rebuilt and refined using

phenix.refine[73] and Coot[74] to a final R-factor of 22.5 and an R_{free} of 25.2, with 94.74% of the residues in the favored region of the Ramachandran plot, 5.04% in the allowed region, and 0.22% outliers. The final structure was validated with MolProbity[75] and deposited in the Protein Data Bank with PDB ID: 5NLM. Data collection and refinement statistics are reported in **Table 2-4**. Figures were prepared with the MacPyMOL Molecular Graphics System, Version 1.0 Schrödinger, LLC. PtUGT1:indoxyl sulfate interactions were analyzed with the Protein-Ligand Interaction Profiler (PLIP)[76] using default distance cut-offs. Structural homology searches were performed with DaliLite v.3[77].

	PtUGT1:indoxyl sulfate*
Data collection	
Space group	P 21 21 2
Cell dimensions	
a, *b*, *c* (Å)	121, 172.82, 48.41
α, β, γ (°)	90, 90, 90
Resolution (Å)	48.41 - 2.14 (2.216 - 2.14)**
R_{merge}	0.09 (1.52)
$I / \sigma I$	15.58 (0.86)
Completeness (%)	97.97 (84.44)
Redundancy	8.9 (3.9)
Refinement	
Resolution (Å)	2.14
No. reflections	55920
R_{work} / R_{free}	0.23 / 0.25
No. atoms	7131
Protein	6994
Ligand/ion	33
Water	104
B-factors	
Protein	73.81
Ligand/ion	128.8
Water	49.41
R.M.S. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.52

*One crystal was used to solve the structure.
**Values in parentheses are for highest-resolution shell.

Table 2-4. Crystallographic data collection and refinement statistics.

Glucosyltransferase assay for enzymology

Rosetta™(DE3) cells were transformed with pTMH307, pTMH308, pTMH634, or pTMH635, and selected on LB agar plates containing chloramphenicol and ampicillin. PtUGT1 and PtUGT2 were subsequently expressed and purified as described in "Expression & purification of PtUGT1 for crystallization" (see **Figure 2-16** for SDS-PAGE

gel). The enzymes were stored in the final buffer with 10% glycerol and flash frozen in liquid nitrogen for storage.

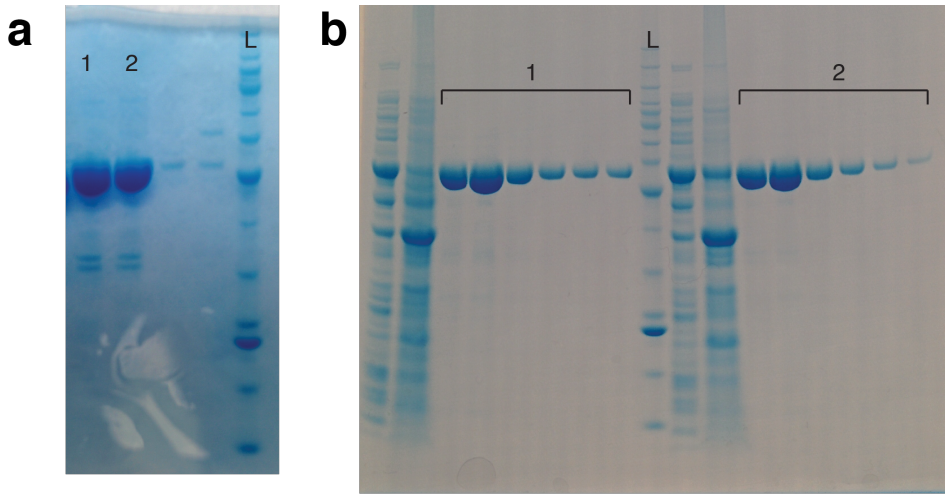

Figure 2-16. SDS-PAGE of UGTs. The proteins show high purity with a strong band at 52 kDa on a NuPAGE Novex 4-12% Bis-Tris Protein Gel (Life Technologies). a) Lane 1 is pTMH307, lane 2 is pTMH308, lane L is BenchMark Protein Ladder. b) Lanes in section 1 are pTMH634 fractions, lane L is BenchMark Protein Ladder, lanes in section 2 are pTMH635 fractions.

The activity assay consisted of two parts: first, indoxyl was produced and quantified; second, the glucosyltransferase's activity on indoxyl was quantified. To produce and quantify indoxyl, 100 mM indoxyl phosphate and 10 mg/mL potato acid phosphatase (both obtained from Sigma) were mixed 2:1 in an anaerobic chamber and allowed to react for 75 min at 30° C. Twofold serial dilutions of the indoxyl mixture were then made, and 50 µL of each indoxyl dilution was taken in triplicate to run a malachite green assay using the Glycosyltransferase Activity Kit (R&D Systems)[78], This assay determines the concentration of free phosphate, and thus, free indoxyl. The indoxyl dilutions were first further diluted by powers of 10 in water as necessary to keep the malachite green signal within the linear range of the assay. The, was used: 30 µL malachite green reagent A (containing ammonium molybdate in 3 M sulfuric acid), 100 µL water, and 30 µL malachite green reagent B (containing malachite green oxalate and polyvinyl alcohol) were sequentially added to 50 µL of each indoxyl dilution. The reactions were incubated for 20 min, and absorbance at 620 nm was measured on an Infinite M1000 PRO microplate reader (Tecan) and converted to indoxyl concentration by comparison against a six-point standard curve of twofold dilutions of potassium phosphate monobasic (range: 3.1-100 µM).

The dilutions of indoxyl not used for the malachite green assay were used in the indoxyl glucosyltransferase activity assay, taking place in an anaerobic chamber. A 100-µL reaction mixture was prepared in triplicate per dilution, consisting of 50 mM HEPES pH

7.8, 50 mM NaCl, 10 mM DTT, 5 mM UDP-glucose, and 20 nM PtUGT1 or PtUGT2 enzyme, and including 15 μL of the indoxyl solution. The reactions were carried out for 10 min at 30° C, and they were quenched with 100 μL 100 mM NaOH.

The reactions were spun down and run on a 6520 Accurate-Mass Q-TOF LC-MS (Agilent). Indican was detected by LC-MS as described in "LC-MS detection and quantification of indican," and indican concentration was determined by comparison against a six-point indican standard curve (Sigma, ≥97% pure) ranging from 8.3 to 265 μM. Michaelis-Menten graphs and turnover numbers were generated by Prism (GraphPad Software Inc.). The K_M for the two PtUGT enzymes could be determined with accuracy because it is not possible to quantify with certainty the concentration of the indoxyl substrate due to low levels of oxidation, even under anaerobic conditions.

Small-scale assays for indican and indigo

Cultures were grown to saturation in MOPS EZ Rich Defined Medium (Teknova, ingredients listed in **Table 2-5**), with kanamycin selection and 2% glycerol in place of the typical 0.2% glucose as the carbon source. Cells were diluted to OD_{600} 0.1 in fresh media containing 1 mM IPTG (if strain contains T7 promoter), and indican or tryptophan as described in the text. They were grown at 37° C in a Multitron shaker (INFORS HT) at 750 rpm for 24 h (or as described in the text, if otherwise). Cultures were filtered through a 0.2 μm filter and stored at -20° C until HPLC analysis.

Component	Concentration
MOPS	40 mM
Tricine	4.0 mM
Iron Sulfate	0.01 mM
Ammonium Chloride	9.5 mM
Potassium Sulfate	0.276 mM
Calcium Chloride	5×10^{-7} M
Magnesium Chloride	0.525 mM
Sodium Chloride	50 mM
Ammonium Molybdate	3×10^{-9} M
Boric Acid	4×10^{-7} M
Cobalt Chloride	3×10^{-8} M
Cupric Sulfate	10^{-8} M
Manganese Chloride	8×10^{-8} M
Zinc Sulfate	10^{-8} M
Potassium Phosphate Dibasic	1.32 mM
Potassium Hydroxide	1.5 mM
Adenine	0.2 mM
Cytosine	0.2 mM
Uracil	0.2 mM
Guanine	0.2 mM
L-Alanine	0.8 mM
L-Arginine	5.2 mM

L-Asparagine	0.4 mM
L-Aspartic Acid, Potassium Salt	0.4 mM
L-Glutamic Acid, Potassium Salt	0.66 mM
L-Glutamine	0.6 mM
L-Glycine	0.8 mM
L-Histidine HCl H_2O	0.2 mM
L-Isoleucine	0.4 mM
L-Proline	0.4 mM
L-Serine	10 mM
L-Threonine	0.4 mM
L-Tryptophan	0.1 mM
L-Valine	0.6 mM
L-Leucine	0.8 mM
L-Lysine	0.4 mM
L-Methionine	0.2 mM
L-Phenylalanine	0.4 mM
L-Cysteine HCl	0.1 mM
L-Tyrosine	0.2 mM
Thiamine	0.01 mM
Calcium Pantothenate	0.01 mM
para-Aminobenzoic Acid	0.01 mM
para-Hydroxybenzoic Acid	0.01 mM
Dihydroxybenzoic Acid	0.01 mM
Glucose	2 g/L

Table 2-5. Components of EZ Rich Defined Medium (Teknova). Any substitutions made were described in the text.

β-glucosidase purification

Rosetta™(DE3) cells were transformed with pRLP121 and selected on LB agar plates containing chloramphenicol and ampicillin. BglA was subsequently expressed and purified as described in "Expression & purification of PtUGT1 for crystallization." The enzyme was stored in the final buffer with 10% glycerol at a concentration of 2 mg/mL and flash frozen in liquid nitrogen for storage.

Shake-flask time course fermentations

Cultures were grown to saturation in MOPS EZ Rich Defined Medium (Teknova), with kanamycin selection and 2% glycerol in place of the typical 0.2% glucose as the carbon source. Cells were diluted to OD_{600} 0.05 in 50 mL fresh media containing 20 mM (4 g/L) tryptophan and 1 mM IPTG. They were grown for 37 h at 37° C in 250-mL baffled shake flasks, in an Innova 44 shaker (New Brunswick Scientific) set to 200 rpm. Aliquots were taken at time points, to determine metabolite concentration or cell count.

Determination of metabolite concentrations

LC-MS detection and quantification of indican

Indican was detected by LC-MS/MS using a 6520 Accurate-Mass Q-TOF LC-MS (Agilent). Five microliters of sample were injected onto a ZORBAX Eclipse Plus C18 4.6 × 100 mm 3.5 μm column (Agilent) using a flow rate of 0.5 mL/min. The solvents were water with 0.05% ammonium hydroxide and acetonitrile with 0.05% ammonium hydroxide. The column was flushed with 95% water/5% acetonitrile plus 0.05% ammonium hydroxide for 3 min, before an elution with a linear gradient to 2% water/98% acetonitrile plus 0.05% ammonium hydroxide over 11 min. Indican (m/z 294.098 $[M-H]^-$, R_t 10.0 min) was ionized by electrospray ionization in negative mode using a fragmentor voltage of 100 V, and for MS/MS, fragmented again using collision-induced dissociation (CID) at 10 V. The LC-MS trace was integrated in MZmine2 (http://mzmine.github.io/). For quantification of indican, traces were compared against a seven-point indican standard curve (Sigma, ≥97%) ranging from 0 to 2.5 g/L.

HPLC quantification of indican and tryptophan

Indican and tryptophan concentrations were quantified by absorbance at 280 nm via diode array detector on a 1260 Infinity LC System (Agilent). Five microliters of sample were injected onto the column described in "LC-MS detection and quantification of indican", using the same solvents and gradient as before (tryptophan R_t 5 min, indican R_t 9.7 min). Integrations were done in ChemStation (Agilent). Indican content in the samples were compared against a 10-point standard curve made of twofold dilutions of indican (Sigma, ≥97%), ranging from 0.01 to 5 g/L. Tryptophan content in the samples were compared against a ten-point standard curve made of twofold dilutions of tryptophan (Sigma, ≥99%), ranging from 0.05 to 25 mM.

Extraction and quantification of indigo

Cultures were spun down at max speed and the supernatant removed. The pellet was resuspended in 100% DMSO (Sigma) and pelleted again. Absorbance at 620 nm was measured on an Infinite M1000 PRO microplate reader (Tecan) and converted to indigo concentration by an 8-point standard curve of commercially available indigo (Spectrum Chemical) dissolved in DMSO, ranging from 2 to 250 μM.

Determination of cell count

Indigo absorbs light at 600 nm, so OD_{600} was not an appropriate method to determine cell density. Instead, 10 μL of culture underwent repeated 10-fold dilutions in LB broth, and 5 μL of each dilution was spotted onto an LB agar plate with kanamycin selection. The most densely populated spot with distinct colonies was counted, and the colony forming units per 1 μL original culture was back-calculated.

Production of dye reagents

β-glucosidase lysate production

Rosetta™(DE3) cells transformed with pRLP121 were grown in 4 L Terrific Broth and induced with 1 mM IPTG as described in "Expression & purification of GT for crystallization." The culture was spun down, and the cell pellet was resuspended in 250 mL 25 mM HEPES pH 7.8, 50 mM NaCl, 1 mM DTT, and 5 tablets cOmplete protease inhibitor (Roche). The cells were lysed by freeze/thaw and sonication. Lysate was spun down at 4° C at 4,800 g, and the crude soluble protein was directly used for dyeing.

Indican production in bioreactor

TMH011 cells were grown to saturation in MOPS EZ Rich Defined Medium (Teknova), with kanamycin selection and 2% glycerol in place of the typical 0.2% glucose as the carbon source. Cells were diluted to OD_{600} 0.1 in 5 L fresh media containing 20 mM (4 g/L) tryptophan and grown in a 14-L reactor vessel on a BioFlo/CelliGen 115 bioreactor (New Brunswick Scientific). The temperature was maintained at 37° C and the dissolved oxygen level was maintained at 20%. The pH was maintained at 6.8, using 14% ammonium hydroxide as the base source. 1 mM IPTG was added at 3.5 h. Aliquots were taken at time points, to monitor metabolite concentration and cell count.

At 40 h, the culture was evacuated from the bioreactor and spun down twice. The pellets were discarded, and the pH of the media was adjusted to 8 using NaOH, to minimize acid hydrolysis during the concentration process. The media was then boiled gently on a stir plate set to 150 C until the volume decreased from 5 L to 1 L.

Dyeing with indican vs indigo

Swatches

Concentrated indican media was sprayed onto both sides a 4 cm × 4 cm square of white cotton denim fabric (Levi's) until the cloth was wet. 500mL BglA lysate was applied to the swatch for 1 min, then the cloth was let drip dry and allowed to oxidize in air. A sodium dithionite-reduced indigo vat was made as a comparison, using a protocol adapted from McKee and Zanger[79]. 3 pellets of NaOH were dissolved with 140 mg indigo (Spectrum Chemical) in 10 mL water and heated to boiling on a stir plate. 2 mL 10% sodium dithionite was added, and the mixture was stirred until the indigo reduced and dissolved. This solution was poured into 88 mL room temperature water and stirred to mix. The denim fabric swatches were dipped into this dye bath for 1 min, let drip dry, and let oxidize in air until dry. All swatches were vigorously washed in water and dried before photographing.

Scarf

Concentrated indican media and BglA lysate were placed in separate spray bottles. A 23″ x 72″ piece of white cotton gauze scarf was draped crosswise on a string and pre-wet by lightly spraying with water. Indican was sprayed onto the scarf until both sides were evenly covered, and then BglA was similarly sprayed onto the fabric until covered. The indican was allowed to hydrolyze and oxidize for 30 min before repeating the sprays. After three such dye cycles, the scarf was flipped on the string so the inside surface was on the outside, and three more dye cycles took place. The scarf was dried overnight and washed vigorously in water (**Figure 2-14b**). The scarf was then laundered in a commercial laundry machine with a full capsule of All Laundry Detergent (Sun Products) (**Figure 2-15**).

Statistics

Sample sizes are described in the figure legend of the appropriate experiment. Error bars represent the mean ± 1 s.d. Individual data points represent biological replicates.

Data availability

All DNA sequences in this work have been deposited in GenBank (www.ncbi.nlm.nih.gov/genbank), and their accession codes are listed in **Table 2-2** and **Table 2-3**. The crystal structure of PtUGT1 has been deposited in the RCSB Protein Data Bank (www.rcsb.org) with ID code 5NLM.

Chapter 3. Development of isatan B as a second generation indican replacement

3.1. Introduction

Thus far, I have been using indican as the indigo dye material of choice, and the glucosyl protecting group is removed via an enzymatic hydrolysis to induce indigo formation. This glucosyl group is very well suited for this application in several ways. First, the glycosidic bond is relatively stable under a variety of conditions, including a wide range of temperatures and pHs. This can be useful when purifying indican – if the purification protocol involves heating up the molecule or changing the pH, there is little risk that indican will be destroyed in the process. Second, indican is relatively soluble to about 100 g/L in water. It is also well tolerated by *E. coli*; it takes more than 12 g/L indican to inhibit growth. Fortunately, a single gene knockout – Δ*bglA* – can essentially eliminate indican hydrolysis by the cell. Finally, indican is a natural product found in plants, so there already exists a biosynthetic pathway somewhere, even if the enzymes involved were not initially known. The fact that indican accumulates in indigo-producing plants also suggests that there are cellular contexts in which the molecule is stable and not metabolized by the cell. A major downside of indican, though, is that the protecting group must be removed using an enzyme, and a glucosidase can be slow to act and relatively expensive to produce. Strong acid has been reported to hydrolyze indican[80], but it is relatively slow, and it is incompatible with the rest of the indigo dyeing workflow, which requires a pH of about 11 for efficient indigo oxidation.

However, there are other protecting groups that can be used to stabilize indoxyl until hydrolysis. In fact, the protecting group strategy is not new to organic chemists, who have long used protecting groups to prevent reactivity at one chemical site while modifying another chemical site on the molecule. Most of these protecting groups are not biologically relevant, but some (such as acyl and methyl groups) could theoretically be used in the cell. There are also other functional groups not typically used as protecting groups in organic chemistry, but are chemical modifications found in nature. For example, the human liver has sulfotransferase enzymes that sulfonate small molecules as a detoxifying mechanism, and one of these products is indoxyl sulfate.

These other protecting groups face many challenges that the glucosyl group does not. Indoxyl acetate, for example, is quite toxic to *E. coli*, and it turns out that non-specific esterases in the cell hydrolyze it easily. It is also not a naturally produced molecule, and

since no known acetyltransferase could efficiently transfer the acetyl group to indoxyl, an acetyltransferase would have to be engineered for the purpose. Indoxyl sulfate did not face many of these problems, but since sulfonation is not a common modification in *E. coli*, the cell did not naturally produce a lot of the donor molecule, 3'-phosphoadenosine-5'-phosphosulfate (PAPS). Since *E. coli* regularly uses UDP-glucose in the biosynthesis of other UDP sugars and to make necessary glycoproteins and glycolipids on the cell surface[81], the cell can efficiently replenish its stock of UDP-glucose; however, the production of PAPS would have to be upregulated to produce high titers of indoxyl sulfate.

However, these alternative protecting groups have certain advantages over the glucosyl protecting group, at least when indigo dyeing is the application. While the glucosyl group contains six carbons, the acetyl group only has two carbons, and the sulfo group has none. This makes the glucosyl group more metabolically expensive to make. The cell has to produce an extra six-carbon moiety that will not be in the final product, since the protecting group is cleaved off prior to indigo formation. One other downside of the glucosyl protecting group is that unless it is being removed from the indigo wastewater, it will increase the biological oxygen demand in the dye effluent more than other smaller protecting groups, which will cause a depletion of dissolved oxygen in the water when microbes consume the carbon in the waste stream. Another key advantage of an acetyl protecting group is that it can be hydrolyzed by a strong base, since the protecting group is connected to indoxyl with the notoriously base-labile ester bond. Since indigo dyeing already must take place around pH 11, it would be convenient to use this basic pH to hydrolyze the protected indoxyl and form indigo without the need for an enzyme.

There is another protecting group that could work for this purpose. In the woad plant *Isatis tinctoria*, there exist three indigo precursors: indican, isatan B (indoxyl 3-ketoglucoside), and isatan A (isatan B, with the glucosyl group malonylated at the 6-carbon position)[80]. Isatan B is of particular interest in this application because its ketoglucosyl protecting group is base-labile, and it can be removed when the pH is increased to 11. Also, isatan B is naturally produced in *Isatis tinctoria*, so there already exist enzymes for its biosynthesis. In fact, since isatan B is one chemical modification away from indican (and isatan A is one chemical modification away from isatan B), I postulate that the three molecules are products of adjacent steps in a biosynthetic pathway (**Figure 3-1**). Given that the woad plant cell can stably produce isatan B, it is conceivable that a bacterial cell could be engineered to be a suitable host for isatan B production. In such an environment, isatan B would not be hydrolyzed by the cell, and the cell not would suffer toxicity from isatan B.

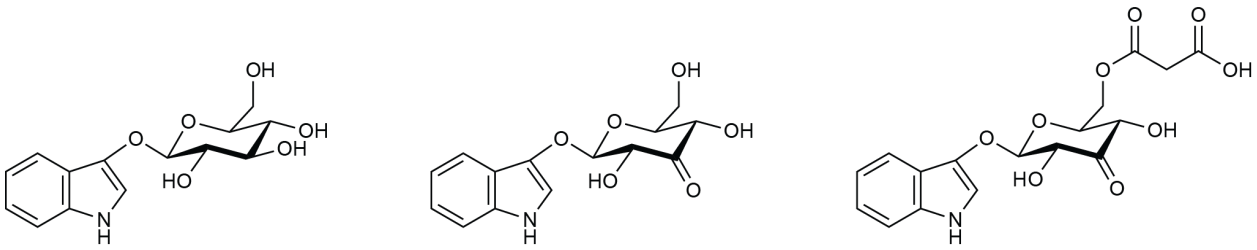

Figure 3-1. Chemical structures of indigo precursors in *Isatis tinctoria*. The indigo precursors found in *Isatis tinctoria* are indican, isatan B, and isatan A.

Although indigo is an ancient dye, and *I. tinctoria* has long been a source of indigo dye, there is little known about the natural biosynthesis of the indigo precursors in the plant. Isatan B from *I. tinctoria* was first characterized to be indoxyl-5-ketogluconate in 1967[82], partly because the base-lability of the molecule matched up with the behavior expected of an indoxyl ester. However, a study in 2004 showed using NMR and mass spectrometry that the structure of isatan B was instead consistent with indoxyl 3-ketoglucoside[80]. The authors also showed that *I. tinctoria* produced another indigo precursor named isatan A, which was the malonylated form of isatan B. Isatan A was too unstable to purify and was readily degraded to isatan B. Though we now have more clarity on the chemical structures of these indigo precursors, their biosynthesis in *I tinctoria* has not been elucidated fully.

3.2. Results

To produce isatan B in the *E. coli* host, I needed to 1) develop an analytical protocol to accurately quantify isatan B, 2) characterize its effects on *E. coli* cells, and if necessary, find conditions that would be compatible with isatan B production, and 3) build a biosynthetic pathway for isatan B production and implement it in *E. coli*.

3.2.1. Purifying isatan B from plants

One difficulty in working with isatan B is that there is no commercial standard available. A pure commercial standard is important to be able to determine with certainty whether isatan B is present in a sample, and how much there is. However, isatan B can be isolated from *I. tinctoria*, and the plant itself is easy to cultivate and relatively fast-growing in a greenhouse setting (**Figure 3-2**). Although the purity of the resulting sample was unknown, it could be used as a mass spectrometry standard, and the retention time of the molecule could thus be identified.

A hot water extraction protocol was first tested to extract isatan B from *I. tinctoria* leaves, and with that crude extract the effectiveness of alkaline hydrolysis was tested. Isatan B has been previously reported to hydrolyze immediately at pH 10.7 and higher[83], and this was corroborated by my results (**Figure 3-3**).

Figure 3-2. Woad plants. Woad plants grow robustly in a greenhouse setting.

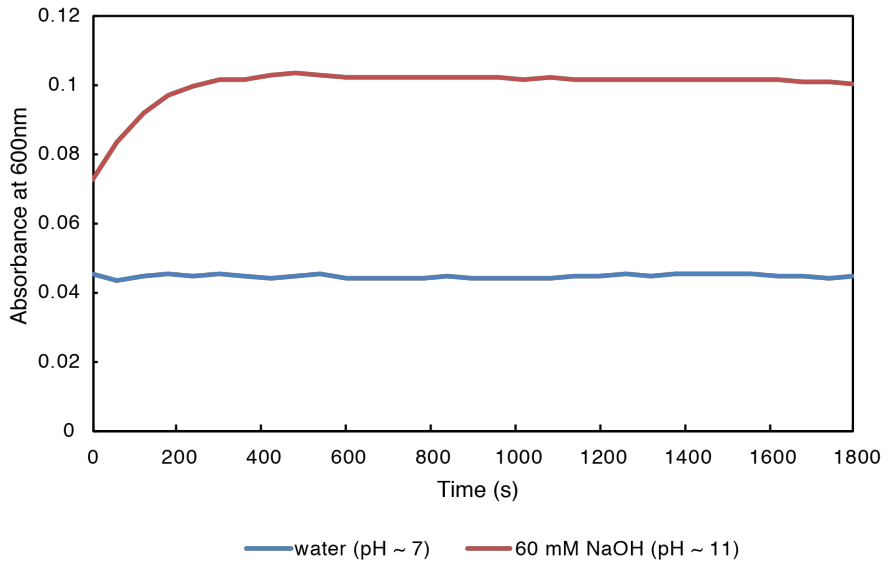

Figure 3-3. Alkaline hydrolysis of isatan B. Hydrolysis of isatan B is rapid under pH 11 conditions.

I subsequently tested a methanol extraction, and it seemed to be much more efficient at extracting isatan B from leaves. The hot water extraction method (using 0.1% formic acid as the extractant to reduce the chance of alkaline hydrolysis) was compared to blending the leaves with the same volume of methanol with 0.1% formic acid. The efficiency of isatan B extraction was determined by thin layer chromatography (**Figure 3-4**) on aluminum-backed silica gel 60 plates, and the plates were visualized with a 0.5% NaOH spray. A mixture of chloroform and methanol was used as the solvent system; various ratios were tested so that the R_f of isatan B was as close to 0.3 as possible. With an R_f around 0.3, isatan B could be separated effectively by column chromatography. The best solvent system was 85 parts chloroform to 15 parts methanol.

Figure 3-4. Hot water extraction vs methanol extraction of isatan B. Left: *I. tinctoria* leaves were extracted using 5 parts methanol with 0.1% formic acid. The blue spot at $R_f \sim 0.70$ is consistent with isatan B. Right: *I. tinctoria* leaves were extracted using 5 parts water with 0.1% formic acid.

The *I. tinctoria* leaf extract was then run on the mass spectrometer to see if the correct peak could be identified, at *m*/*z* 292.0821 in negative mode ($[M-H]^-$ for isatan B) or *m*/*z* 294.0978 in positive mode ($[M+H]^+$). Negative mode gave about 17 times better signal than positive mode, so negative mode was used for all subsequent LC-MS runs. There was a peak visible at R_t = 10.5 min, but a peak was also detected at R_t = 10.0 min, which is consistent with indican's retention time. When pure indican was run on the LC-MS, it also showed a *m*/*z* 292.0821 peak at R_t = 10.0 min. Therefore, it was clearly important for isatan B samples to be free of indican, otherwise it would be more difficult to differentiate isatan

B and indican on the LC-MS. The way to differentiate the two is to look for a $[M-H]^-$ peak at 294.0978; there should only be a peak at that *m*/*z* with indican, not isatan B.

Thus, a method was needed to separate indican from isatan B in mixed samples. First, I tried using adsorbent resins to selectively bind indican or isatan B, leaving the other compound, but none of my available resins were successful in this task (**Table 3-1**). Indeed, it is unlikely for an adsorbent resin to distinguish between indican and isatan B, given their structural similarities. I then turned to column chromatography to separate out indican. First, fractions were collected off a silica gel column with a solvent system of 85:15 chloroform to methanol, and samples of the fractions were treated with strong base to see which turn blue, i.e. which contained isatan B. These desired fractions were then applied to a Sephadex LH-20 column run with methanol, and fractions from this column were tested for isatan B in the same way. Using both columns in series, I was able to purify isatan B away from indican. This was verified by mass spec – at *m*/*z* 294.0978, there was no peak. So, the retention time was identified on mass spec, but since the standard was not fully pure, it could not be used to quantify isatan B.

No.	Resin
1	Amberlite XAD4
2	Amberlite XAD7HP
3	Amberlite XAD18
4	Amberlite FPX66
5	Dowex Optipore L-493
6	Relite SP411
7	Relite SP460
8	Relite SP490
9	Sepabeads SP70
10	Sepabeads SP700

Table 3-1. List of resins tested to separate indican from isatan B.

3.2.2. Characterizing isatan B stability in *E. coli*

Now that a method to was developed to quantify isatan B, at least on a relative level, I wanted to engineer a strain of *E. coli* that would be compatible with producing isatan B. This primarily meant making a strain where isatan B is stable. First, the stability of isatan B was tested in plain *E. coli* cells. The first strain tested was MG1655(DE3)Δ*bglA*, because that was the background strain the indican-producing *E. coli* was built off of, and the goal was to integrate isatan B production into that same strain. Preliminary results from incubating MG1655(DE3)Δ*bglA* cells in 400 μL LB media with 100 μL hot-water-extracted

woad lysate showed that the cells did seem to degrade isatan B quite a bit. Compared to lysate that was incubated at the same temperature without cells, the amount of isatan B was drastically reduced in samples with cells.

Since the degradation of isatan B appeared to be cell-dependent, I hypothesized that one or more enzymes in the cell was responsible for the degradation. Unfortunately, there were no great candidates for such an enzyme, since there are several reactions that could have happened to isatan B and a large number of enzymes that could have catalyzed those reactions. So, I tried to narrow down the possibilities by testing isatan B stability in an *E. coli* strain with a minimal genome, called DGF-298 (ref. [84]). Since DGF-298 has a 2.98 Mbp genome (compared to the 4.64 Mbp genome of MG1655), it is missing a large portion of its genes, and it would be possible to determine if the gene(s) responsible for isatan B degradation are in the shared set of genes, or are present in MG1655 and not in DGF-298. This time, LB media was tried alongside the EZ Rich growth medium used for the cultivation of indican-producing cells.

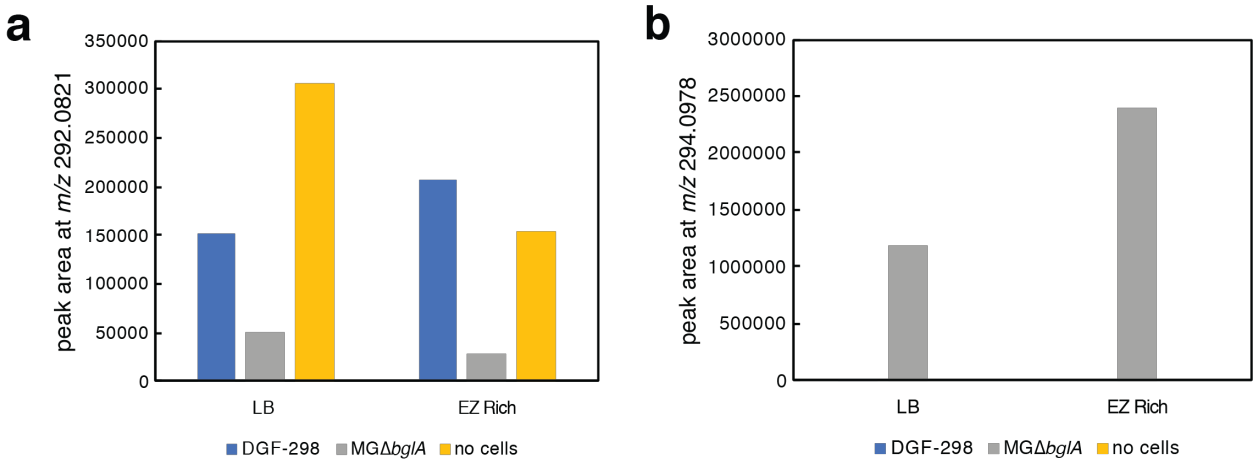

Figure 3-5. Isatan B stability in minimal genome *E. coli* strain. a) Isatan B was incubated with normal *E. coli* cells (MGΔ*bglA*), minimal genome cells (DGF-298), or no cells. After 24 h, isatan B was still largely present in DGF-298 cells, whereas incubation with MGΔ*bglA* cells caused significant degradation. b) Interestingly, the isatan B degraded by MGΔ*bglA* (signal at *m/z* 292.0821) appeared to become indican (signal at *m/z* 294.0978).

The isatan B appeared to be degraded by MG1655(DE3)Δ*bglA* cells after 24 h of incubation, but DGF-298 cells did not seem to have the same effect (**Figure 3-5a**). The effect of DGF-298 cells on isatan B was not totally consistent, compared to isatan B incubated without cells, but it was clear that the cells did not degrade isatan B nearly as much as MG1655(DE3)Δ*bglA* cells did. One possible alternative explanation for the results was that the putative indican transporter(s) present in MG1655 cells was absent in DGF-298, and the minimal genome cells were simply not taking up isatan B. However, indican, a

structurally similar molecule, was shown to be taken up by DGF-298 cells: cells expressing a BGL active on indican turned blue when incubated with indican, implying that the indican was transported into the cells and hydrolyzed by BGL.

Although a much smaller isatan B peak (*m*/*z* 292.0821) was seen with MG1655 cells than with DGF-298 cells, there was also a peak consistent with indican (*m*/*z* 294.0978, R_t = 10.0 min) (**Figure 3-5b**). This peak was not present in the no-cells condition, showing that there was no indican present in the original purified isatan B that was fed to the cells. It was also not present with DGF-298 cells, showing that it could be related to isatan B degradation. The presence of an indican peak suggested that isatan B degradation was due to it being reduced to indican.

I also tried growing cells in other types of media to see if media components can have an effect on the stability of isatan B (**Figure 3-6**). In addition to MG1655(DE3)Δ*bglA* and DGF-298, I tried another reduced genome strain called MDS42 (ref. [85]), which had a 4.0 Mbp genome. I also tested the version of all three strains that had *bglA* as well as the Δ*bglA* variant, since I suspected *bglA* might act on isatan B due to its structural similarity to indican, a known substrate.

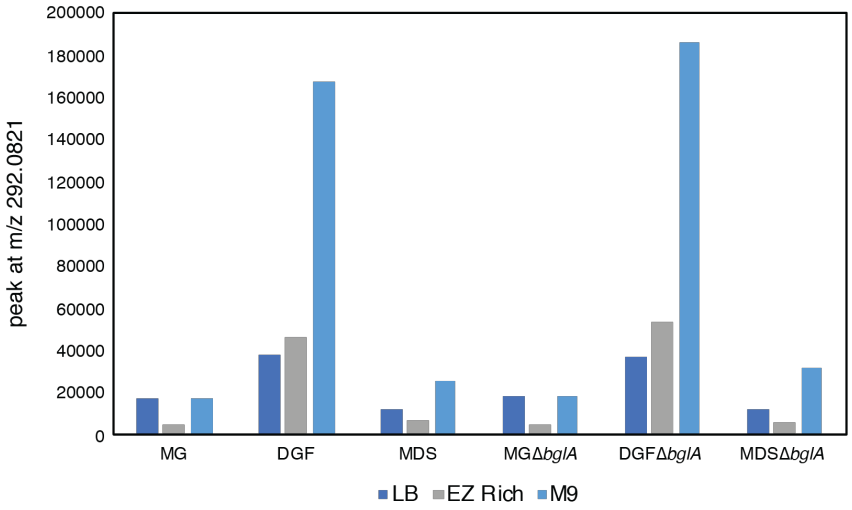

Figure 3-6. Isatan B stability in various media. Isatan B was incubated with MG1655(DE3), DGF-298, or MDS42 cells; or their Δ*bglA* counterparts. The cells were grown in LB, EZ Rich, or M9 minimal media.

MDS42 cells turned out to behave similarly to MG1655 cells in that both strains degraded isatan B to the same extent, and the presence or absence of *bglA* was generally irrelevant. DGF-298, like before, did not degrade isatan B as much as MG1655 did, and it did not matter if *bglA* was present. Similar to before, DGF-298 cells grown in LB and EZ Rich showed about the same amount of isatan B degradation. However, when DGF-298 cells were grown in M9 minimal media, the isatan B in the media showed a much larger peak,

suggesting that the most protective environment for isatan B production is in DGF-298 cells grown in M9 minimal media. However, further work is needed to determine what specific reductase(s) present in MG1655 and MDS42 but absent in DGF-298, are responsible for reducing isatan B to indican.

3.2.3. Finding an oxidase to convert indican to isatan B

Making an *E. coli* chassis for isatan B production is important, but a biosynthetic pathway was also needed to produce isatan B in the cell. Based on the chemical structure of isatan B, I hypothesized that there are two plausible routes that *I. tinctoria* uses: 1) a 3-ketoglucosyl group could be transferred directly to indoxyl (similar to how a glucosyl group is transferred to indoxyl to form indican), or 2) indican could be directly oxidized to isatan B via an enzyme that oxidizes the C3 hydroxyl of the glucoside to a carbonyl. While the first option is possible, it seems unlikely because UDP-3-ketoglucose is not a common sugar donor, and its existence has been speculated only rarely[86], and never in plants. There are also no known glycosyltransferases that can transfer a ketoglucosyl group. Thus, most of my efforts focused on finding an oxidase that can act on indican. Another indication that this is the right direction is that indican, isatan B, and isatan A are all found in *I. tinctoria*, and their chemical structures suggest that they are adjacent steps in a metabolic pathway (**Figure 3-1**). It is not unreasonable that indican would be oxidized to isatan B, and isatan B malonylated to produce isatan A.

Although there are no known enzymes that perform exactly the function I am looking for, there are examples of oxidases that act on glucose and other glucosides, oxidizing their C3 hydroxyls to carbonyls. These tend to fall into two categories: pyranose dehydrogenases (PDH, EC 1.1.99.29) from mushroom species, and glucoside 3-dehydrogenases (G3DH, EC 1.1.99.13) from various bacteria. I ended up testing five G3DHs (from *Agrobacterium tumefaciens*[87-92], *Stenotrophomonas maltophilia*[93,94], *Halomonas* sp. α-15[95,96], *Flavobacterium saccharophilum*[97-99], and *Sphingobacterium faecium*[100-102]) and two PDHs (from *Agaricus bisporus*[103-106] and *Agaricus meleagris*[107-109]). Both types of dehydrogenases have a broad substrate specificity, though they generally act on sugars or glucosides[110,111]. The G3DHs are specific for the C3 hydroxyl, while the PDHs oxidize the C1, C2, and/or C3 hydroxyl[112]. The PDHs are also known to have a 25-amino acid N-terminal signal sequence that is cleaved off on the mature protein, which tags the protein for secretion[107,108]. Thus, the protein is always found extracellularly in the native host. The PDHs are also heavily glycosylated in fungi[108,113], making them unlikely to function in bacteria. There are also two *A. bisporus* protein sequences available in UniProt, so I tested both - UniProt A0A0A7EMT3 was referred to as "*A. bisporus* 1" and UniProt Q3L1D1 was "*A. bisporus* 2."

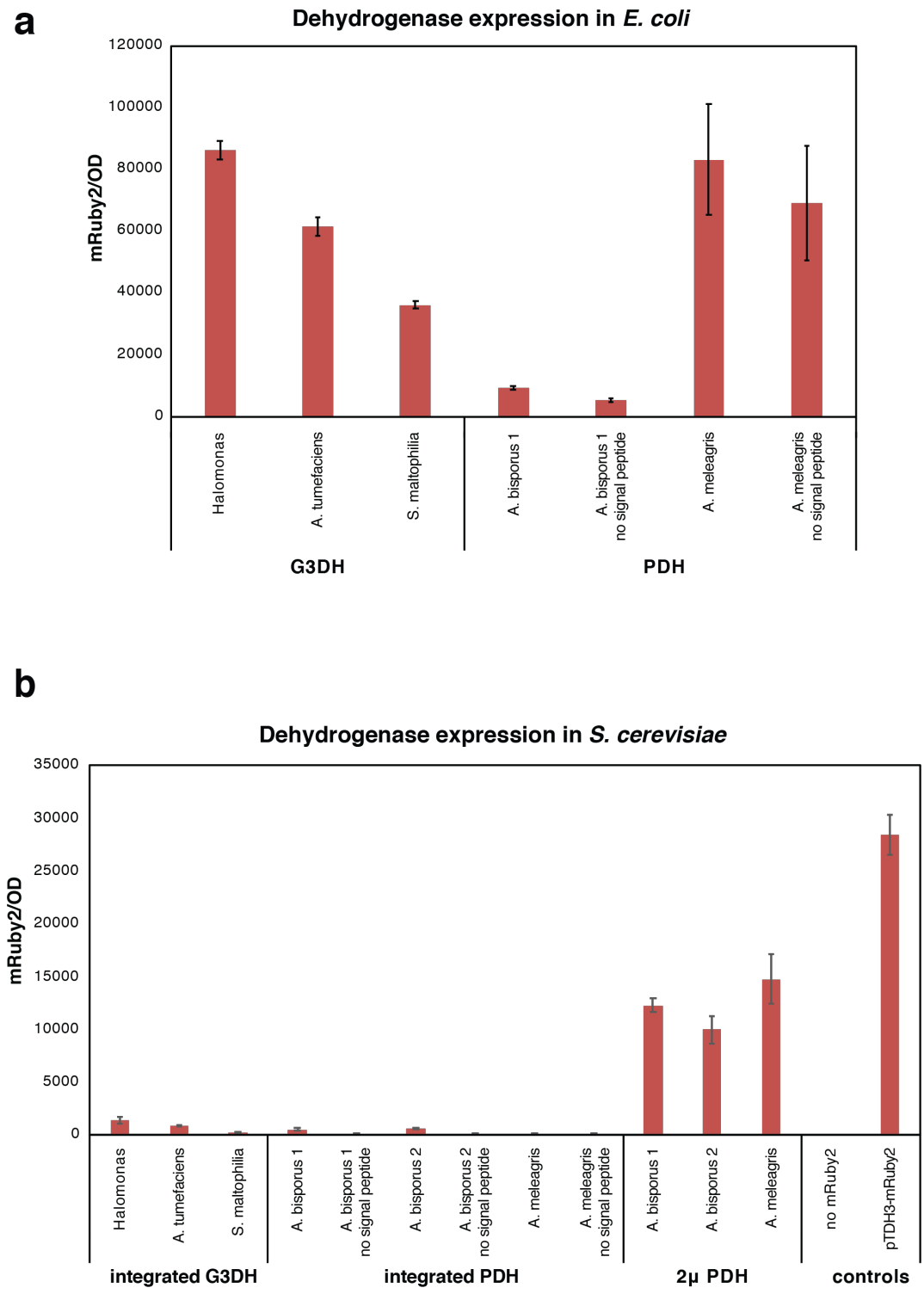

Figure 3-7. Dehydrogenase expression in *E. coli* and *S. cerevisiae*. G3DHs and PDHs were tagged with mRuby2 and expressed in *E. coli* or *S. cerevisiae*. The proteins generally expressed well in *E. coli*, but they only expressed well in *S. cerevisiae* when the gene was on a 2µ plasmid. The positive control had a pTDH3-mRuby2 cassette integrated into the genome.

I tested the PDHs both with and without the signal sequence, in *E. coli* and *S. cerevisiae*, measuring protein expression and protein activity in a variety of ways. First, I expressed the dehydrogenases tagged with the mRuby2 fluorescent protein in *E. coli* and *S. cerevisiae*, and I measured mRuby2 expression in culture (**Figure 3-7**). In *E. coli*, the G3DHs and PDHs were driven by the BBa_J23100 strong promoter, and I made an RBS library between the promoter and the gene. This library was transformed into *E. coli*, and the colonies were screened for the brightest red colony. The selected colonies were then grown in LB, and the OD and mRuby2 fluorescence was measured after 42 h. The dehydrogenases expressed moderately well. Surprisingly, the *A. meleagris* PDH seemed to express almost as well as the bacterial G3DHs. In *S. cerevisiae*, I first tried expressing the dehydrogenases with the strong TDH3 promoter and integrating the genes into the genome, but the expression was extremely low, even after 42 hours of growth. When the genes were expressed on a 2μ plasmid, however, the proteins expressed much better.

After getting confirmation that the dehydrogenases could be expressed at least somewhat well in *E. coli* and *S. cerevisiae*, I looked for dehydrogenase activity. There are a number of ways to observe this activity, but they generally fell into two categories: I could indirectly look for evidence that an oxidation reaction was happening, or I could directly look for the appearance of the final product, isatan B.

The most common way to observe G3DH/PDH activity in the literature is to use a synthetic electron acceptor that changes color when reduced. Indicators such as 2,6-dichlorophenolindophenol (DCPIP)[88,90,93,98,100], 1,4-benzoquinone[104,106], and potassium ferricyanide[88,93,98] have been used as an electron acceptor, with DCPIP being the most common and reliable. DCPIP is a blue dye when in the oxidized form, but when it is reduced, it becomes colorless. This color change can be detected by a loss in absorbance at 600 nm. Although the DCPIP assay is typically done on using purified enzyme, a whole-cell version of the assay was developed by Peters et al[114]. I developed a protocol adapted from existing methods that would allow me to measure oxidase activity when the dehydrogenase, or a cell expressing the dehydrogenase, was fed indican. I tested the five bacterial G3DHs in *E. coli* (**Figure 3-8**), since the fungal PDHs were unlikely to work in *E. coli*. As a positive control, I used a uronate dehydrogenase (provided by labmate Ryan Protzko) that was known to express in *E. coli*, and whose oxidation of galacturonic acid was known to show a signal with the DCPIP assay. Sucrose was also used as a positive control for the dehydrogenases, since the ones I tested were known to act on sucrose *in vitro*. For negative controls, I used cells that did not express a heterologous dehydrogenase, or were not fed substrate.

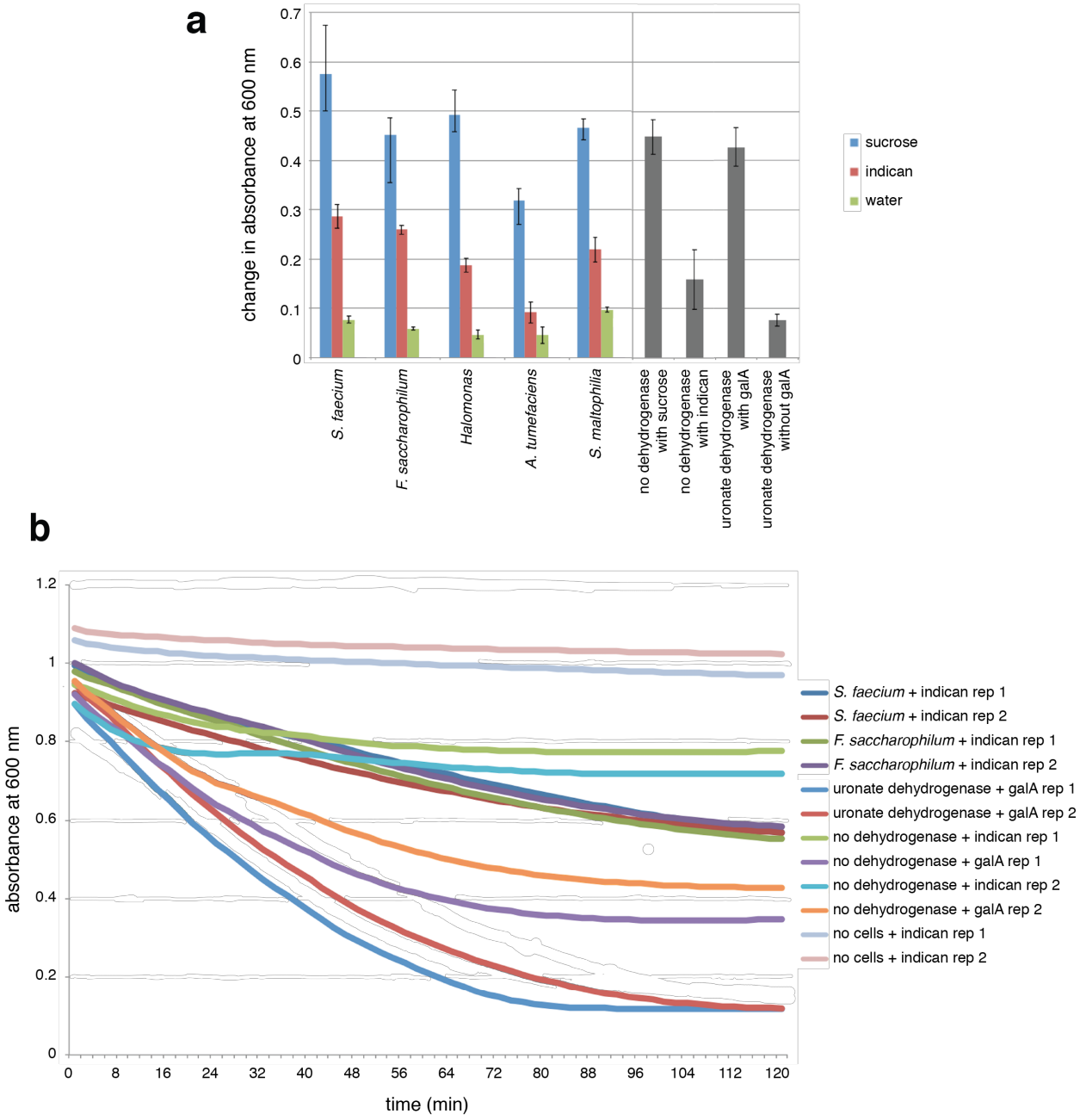

Figure 3-8. DCPIP assay on *E. coli* cells expressing various dehydrogenases. a) change in DCPIP signal, measured in the change of absorbance at 600 nm from the beginning of the assay (t = 0 min) to the end of the assay (t = 60 min). A positive change means that DCPIP signal decreased. *E. coli* cells expressing the various dehydrogenases are incubated with sucrose, indican, or water as a substrate. b) DCPIP signal over 2 hours, for the two most promising dehydrogenases and a few controls.

Out of the dehydrogenases tested, it seemed that most controls looked about as expected, except that cells not expressing heterologous dehydrogenase did reduce DCPIP in the presence of sucrose (**Figure 3-8a**). This could be because some other native *E. coli* enzyme could oxidize sucrose, or perhaps there was some small amount of residual media that

was not fully washed away, which could lead to higher *E. coli* metabolism and background oxidase activity. However, the other controls looked reasonable (cells expressing uronate dehydrogenase fed galacturonic acid showed a large decrease in DCPIP signal, while cells that were not fed substrate generally showed very small decreases in signal. For the dehydrogenases tested, the ones from *S. faecium* and *F. saccharophilum* seemed the most promising, especially on indican. Looking closer at the rates of DCPIP reduction (**Figure 3-8b**), these two dehydrogenases do show some activity over background, but not as much as uronate dehydrogenase on galacturonic acid.

It has also been reported[100] that ketosugars produce a characteristic peak at 340 nm when treated with base. This may be absorbance from the hydrolysis product of the ketoglycoside group. I looked for such a peak upon alkaline treatment, but though the base did cause an increase in absorbance at 340 nm, there was not much of a peak over background controls.

I also tried the DCPIP assay on *S. cerevisiae* cells expressing G3DH or PDH, with the *Neurospora crassa* cellodextrin transporter CDT1 expressed to import the sugar substrate. However, the yeast controls for the assay did not look promising, so I was not confident in the assay's ability to detect dehydrogenase activity in yeast.

The other way to find out if these dehydrogenases are working in *E. coli* is to try to see isatan B directly, via LC-MS or by visually looking for an increase in indigo color when treated with base. Unfortunately, an isatan B peak was never detected even on the sensitive QTOF, nor was blue color ever visible upon alkaline treatment.

Finally, I tried purifying the dehydrogenases (**Figure 3-9**) and running enzyme assays on the purified proteins. Though the dehydrogenases were purified to some degree, they did not show a DCPIP signal over background, nor was an isatan B peak visible on the QTOF.

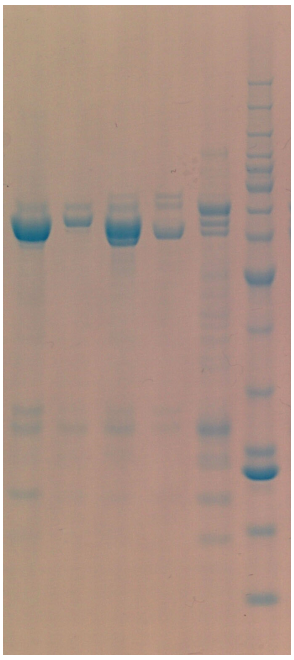

Figure 3-9. SDS-PAGE of G3DHs. G3DHs from *Halomonas*, *A. tumefaciens*, *S. maltophilia*, *F. saccharophilum* and *S. faecium* were purified using nickel beads. The proteins were then run on a polyacrylamide gel, in that order, with BenchMark Protein Ladder in the last lane. The expected protein sizes were, from left to right: 63.3 kDa, 63.5 kDa, 62.7 kDa, 64.6 kDa, and 63.1 kDa. Most of the proteins were moderately pure in the soluble fraction, with the exception of the *F. saccharophilum* G3DH, which was mostly found in the insoluble fraction.

I then tried evolving the two most promising G3DHs from the whole-cell DCPIP assay (those of *F. saccharophilum* and *S. faecium*) using error-prone PCR. The plasmids with the mutated G3DH gene was transformed into *E. coli* and plated on a library plate. There was no good way to screen the colonies for increased activity on indican at this point, so 32 colonies from each plate were tested at random. Each colony was grown in culture, and a small volume of the cultures were used in the DCPIP assay. The results were compared against those of the non-mutated, wild type G3DH (**Figure 3-10**). Most mutants showed deleterious effects, as expected, but interestingly, a few mutants showed higher activity on indican than their wild type parent. They did not show isatan B production as measured on the QTOF, however.

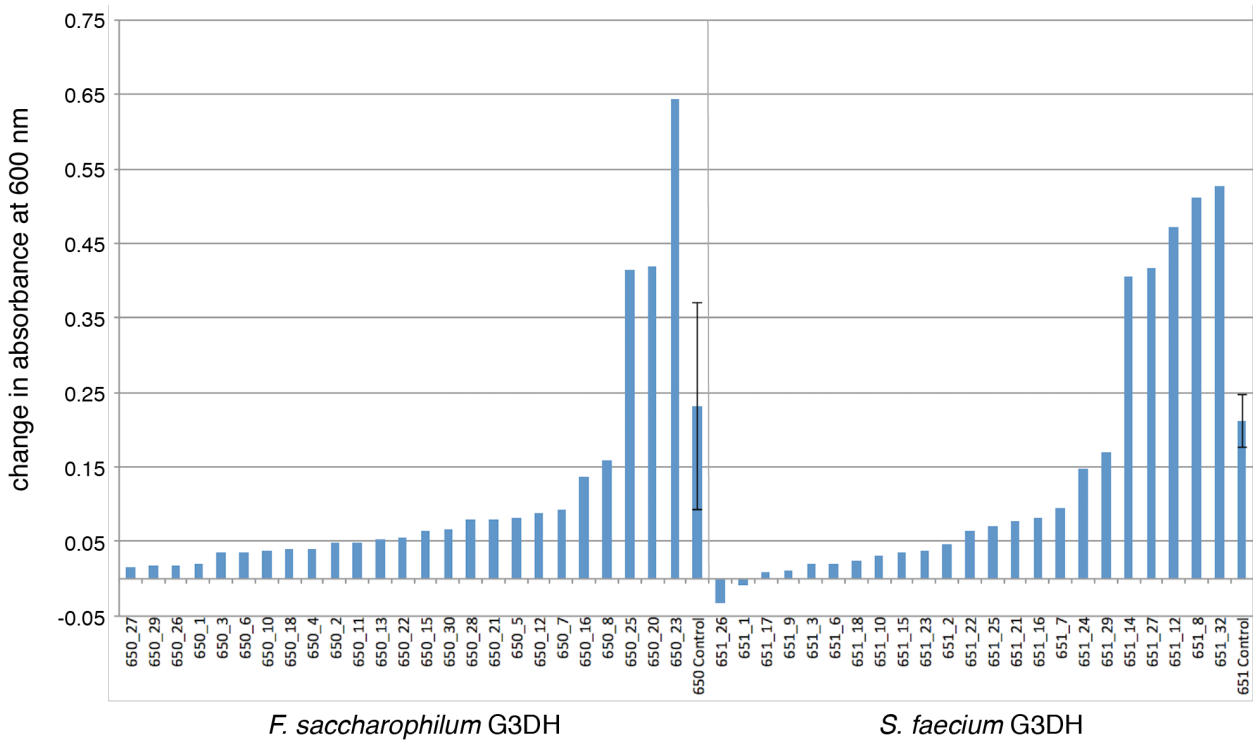

Figure 3-10. Error-prone PCR on *F. saccharophilum* and *S. faecium* G3DHs. The *F. saccharophilum* and *S. faecium* G3DHs were mutated using error prone PCR, and 32 mutants were chosen at random. The mutants were then used in a DCPIP assay, and the results were sorted by activity on indican. These mutants were then compared to the non-mutated control G3DH parent.

I was also interested in the possibility of using the DCPIP assay on colonies on a plate, so in the future there could be a better way to screen for improved activity when evolving a promising dehydrogenase. This has been tried successfully before to identify a gluconate dehydrogenase in *Erwinia cypripedii*: *E. coli* colonies expressing fragments of *E. cypripedii* genome were grown on a filter paper placed on agar plates, and the filters were moved to a thick paper soaked with DCPIP assay mixture[115]. The colony expressing the dehydrogenase gene was identified by a yellow clearing around it. Preliminary tests were run on this plate-based assay and a positive visual DCPIP signal was detected, but more work would be needed to tune the sensitivity of the assay, to show better signal over background.

3.2.4. Finding a glucoside 3-dehydrogenase from *I. tinctoria*

In addition to finding a known glucoside dehydrogenase that can oxidize indican to isatan B, I also looked for the dehydrogenase in *I. tinctoria* that natively performs the activity. Such an enzyme should exist, since the plant does make isatan B, and I hypothesize that the biosynthesis pathway involves an enzymatic oxidation of indican to isatan B. However, this enzyme has not been identified in *I. tinctoria* or any other isatan

B-producing plant. Fortunately, there was a publicly available transcriptome sequence for *I. tinctoria*, so this could be mined for a promising enzyme sequence. I looked for transcripts that had some sequence similarity to the known bacterial G3DHs and fungal PDHs, because although their activity on indican was questionable, they were known to carry out the desired reaction on similar molecules.

The bioinformatic program HMMER[116] was used to search the *I. tinctoria* transcriptome for possible hits. First, the Pfam database[117] searched to find the protein domains that the known dehydrogenases have in common. It turns out that all of them include a GMC_oxred_N domain and a GMC_oxred_C domain, and through Pfam, a multiple sequence alignment was made of each domain. Then, the transcriptome file was then translated into amino acids across all six frames, and hmmsearch was used to search the transcriptome for proteins that contained the GMC_oxred_N domain or the GMC_oxred_C domain. Of the hits, a handful had E-values very close to 0, and eight of those sequences were hits for both GMC_oxred_N and GMC_oxred_C. These were named wGMC1-8.

When the sequences for the candidate dehydrogenases were identified, PCR was used to amplify the sequences from frozen, crushed *I. tinctoria* leaves. For wGMC1, wGMC2, wGMC6, and wGMC8, one version of each dehydrogenase was amplified, though the obtained sequence had a few point mutations from the sequence from the published transcriptome. In the case of wGMC2, a premature stop codon was introduced near the 3′ end of the gene. For wGMC3, wGMC4, and wGMC7, multiple versions of the dehydrogenase were amplified, each with a few point mutations that differentiated them from each other and the sequence from the published transcriptome. wGMC3 and wGMC4 actually turned out to be highly homologous sequences, and in all cases wGMC4 had a premature stop codon that rendered it the same length as wGMC3. wGMC5 could not be amplified at all, even after multiple attempts at PCR.

These candidate wGMC genes were then expressed in *E. coli* under the P_{BAD} promoter and tagged with mRuby2, and their expression levels were quantified by measuring mRuby2 fluorescence. Since wGMC2 and wGMC4 had premature stop codons, their expression could not be measured this way. For the remaining wGMC dehydrogenases, mRuby2 fluorescence was measured (normalized by cell density) and compared against fluorescence from mRuby2 directly controlled by pBAD. wGMC1 and wGMC8 had relatively high expression, while wGMC3 and wGMC6 had very low expression. Unexpectedly, wGMC7-1 and wGMC7-2 had very different expression levels in *E. coli*, even though their sequences only differ by five amino acids.

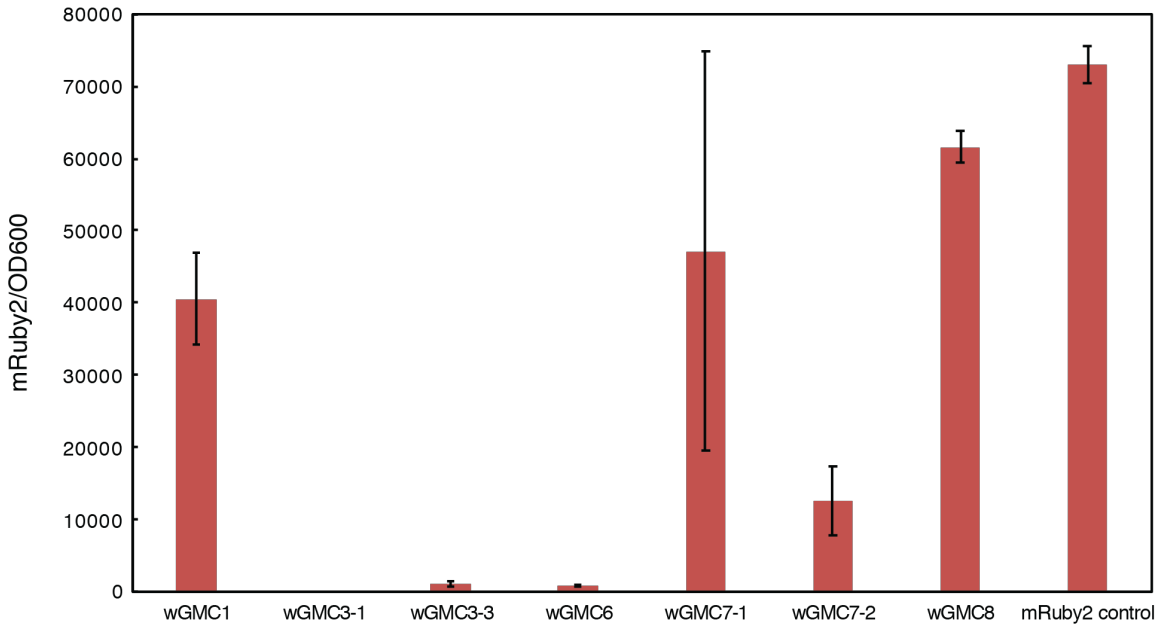

Figure 3-11. Expression of wGMC candidate dehydrogenases in *E. coli*. wGMC candidates from *I. tinctoria* were tagged with mRuby2 and expressed in *E. coli*. Protein expression is relatively high for wGMC1 and wGMC8, and expression is low for wGMC3 and wGMC6. Strangely, expression of the two wGMC7 variants are split.

The DCPIP assay was then used to test which dehydrogenases could be acting on indican. Each candidate was expressed under the pBAD promoter and grown in culture, and the cultures were incubated with DCPIP and indican. DCPIP reduction was measured by a decrease in absorbance at 600 nm over three hours (**Figure 3-12a**). Interestingly, several candidates showed strong, very similar signals, and they were all derived from wGMC3 or wGMC4. Since wGMC3 and wGMC4 are derived from the same base sequence, there was essentially one primary hit from this assay. This is unusual given that wGMC3 showed almost no fluorescence when tagged with mRuby2, implying that its expression is very low. To see if wGMC3 and wGMC4 just have high background oxidase activity on some other compound native to *E. coli*, I re-ran the assay, comparing cells fed indican with cells not fed indican (**Figure 3-12b**). Although the dehydrogenases did show some activity even without indican, the DCPIP reduction effect was much stronger when indican was present.

However, when these samples were run on the LC-MS, no isatan peak was seen when compared to woad-extracted isatan B positive control and a no-dehydrogenase negative control.

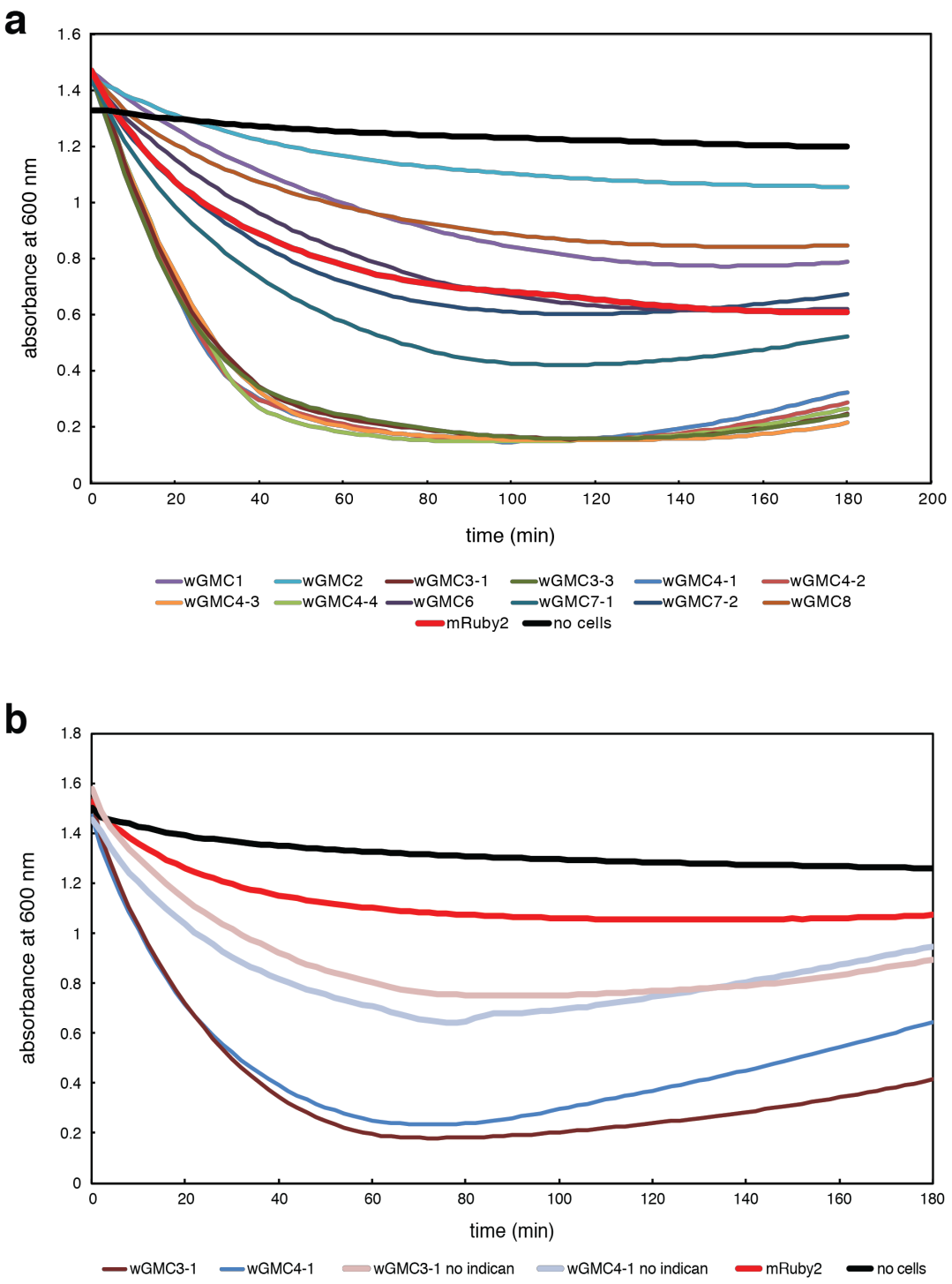

Figure 3-12. DCPIP assay on *E. coli* expressing wGMC candidate dehydrogenases. a) The wGMC candidate dehydrogenases were tested with the DCPIP assay, and variants of wGMC3 and wGMC4 (actually the same base sequence) showed a consistently strong signal. b) wGMC3-1 and wGMC4-1 were tested with and without indican substrate, and the signal was much stronger with indican than without it.

I also tried purifying wGMC3 (wGMC3-1 and wGMC3-3) and the next most likely candidate, wGMC7-1 (**Figure 3-13**). With purified proteins, I could be more certain that any effects observed are due to indican oxidation to isatan B. Unfortunately, as suggested by the mRuby2 tagging data in **Figure 3-11**, wGMC3-1 and wGMC3-3 did not express well at all. wGMC7-1 expressed moderately well, but it was only present in the pellet. These proteins were used in enzyme assays (with DCPIP present as an electron mediator) and the results were run on the LC-MS to look for isatan B production. However, no peak was observed with a retention time that corresponds to isatan B from woad. There was a visible peak nearby, but it was no larger than the no-enzyme negative control, and it did not disappear when the sample was treated with base, implying that it is not isatan B.

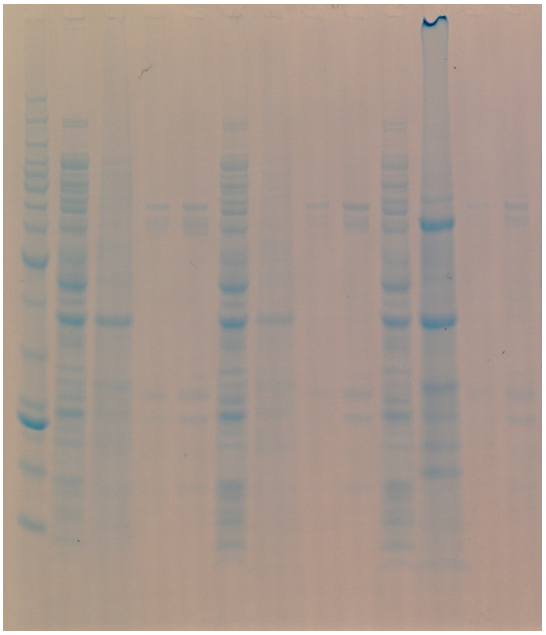

Figure 3-13. SDS-PAGE of wGMC candidate dehydrogenases. wGMC candidates were purified using nickel affinity chromatography. BenchMark protein ladder is in the first lane, and the lysate, pellet, first elution, and second elution were run for wGMC3-1, wGMC3-3, and wGMC7-1, respectively. wGMC3-1 and wGMC3-3 are expected to be 81.6 kDa, and wGMC7-1 is expected to be 61.2 kDa. wGMC3-1 and wGMC3-3 showed very little expression, and wGMC7-1 showed moderate expression in the pellet.

Finally, I attempted to verify that there is indeed a dehydrogenase in *I. tinctoria* that acts on indican to form isatan B. Crushed woad leaves were extracted with buffer to create a woad lysate, and the lysate was incubated with or without indican for the DCPIP assay. The presence of added indican did not seem to lead to a higher DCPIP signal; however, the assay was not that sensitive because there was high background signal even without added indican, and it is possible that small differences could not be observed.

3.3. Discussion

I have been quite successful in producing indican in *E. coli* and showing its use as a dye, but to develop an industrially scalable indigo dyeing process, it would be even better to use a dye precursor that was as close to a drop-in product as possible. Isatan B is a good candidate for such a dye precursor, because 1) it does not require an enzyme to hydrolyze, only the strong base already needed in the indigo dyeing process, and 2) the existing indican biosynthesis pathway could potentially just be extended by one step to make isatan B. Since isatan B is a natural product made in *I. tinctoria* plants, there must also be an existing enzymatic pathway that can produce it.

Although I was ultimately not successful in producing isatan B in *E. coli*, progress was made in a number of areas towards this goal. First, a protocol was developed to purify isatan B from *I. tinctoria* leaves, and though the purity of the indican sample could not be determined, I did verify via LC-MS that at least the similarly-structured indican was not present in the sample. Then, conditions for growing *E. coli* were found that would be conducive to producing isatan B and not readily degrading it. I also looked into glucoside dehydrogenases, from bacterial and fungal sources as well as putative glucoside dehydrogenases from *I. tinctoria*.

There are a few next steps that would help with producing isatan B. First, it would greatly help to have a dependable isatan B standard and LC-MS method. While it was possible to identify an isatan B peak with the right mass per charge and a reasonable retention time on the LC-MS, it was never a clean, symmetric peak. Further development of the LC-MS protocol could clean up the mass spectrometry trace. Moreover, the isatan B standard could be further purified to completion, so that a standard curve could be run on the LC-MS, and samples could be quantified absolutely rather than relatively. Having a robust isatan B quantification method is critical for any further *E. coli* engineering to produce the molecule.

The next step is engineering an *E. coli* strain to not degrade isatan B. A protective cellular environment is also essential for building the initial biosynthetic pathway in *E. coli*, because a dehydrogenase not well adapted to producing isatan B would not synthesize much of it, and if such a small amount of product was immediately degraded, the promising enzyme would never be identified. In fact, it is possible that one or more of the current candidates may be capable of catalyzing the reaction from indican to isatan B in the cell, but the product was not sufficiently stable in the cell to detect using LC-MS after 24 h of contact with the cellular environment (or the LC-MS method was not

sensitive enough to identify such small quantities of isatan B). The preliminary research for making a protective *E. coli* environment has been promising thus far, showing two strains of *E. coli* that seem to degrade isatan B to indican, and one strain that does not. Looking at a list of genes present in MG1655 and MDS42 but absent in DGF-298 could provide some clarity on what enzymes are degrading isatan B. From there, the candidate enzymes could be knocked out of the MG1655 *E. coli* genome, along with other similar enzymes.

Finally, an essential component of this plan is to find an enzyme to make isatan B from indican in this background *E. coli* strain. It is unfortunate that there is only one known family of enzymes that oxidize glucosides to 3-ketoglucosides, and the closest version of that enzyme in *I. tinctoria* does not seem to produce isatan B in *E. coli*. There are a few potential routes to an isatan B-producing enzyme. One option is that if there is one or more reductase native to *E. coli* that reduces isatan B to indican, homologous proteins could be identified in *I. tinctoria*. It is possible that if the enzyme(s) can convert isatan B to indican, they or their homologues could carry out the reverse reaction. Another option is to make a cDNA library of the *I. tinctoria* transcriptome, express it in *E. coli*, feed indican, and screen the library for isatan B-producing colonies. This requires the development a robust screen – either the DCPIP assay could be adapted for screening colonies on a plate, or a plate screen could be developed where colonies are fed indican through the agar, and the colonies are then transferred to a high-pH plate, where any isatan B produced by colonies would hydrolyze and form visible indigo. Finally, probably the most ambitious option for finding an indican oxidase is to shear the *I. tinctoria* genome and express the library in *Arabidopsis thaliana* plants. The leaves of each library member could then be treated with indican and screened for isatan B production.

I am hopeful that with future study, the biosynthesis of isatan B would be possible. This would be a great boon to this novel indigo dyeing strategy, and it may help shed light on yet-unanswered questions about the biosynthesis of indigo precursors in a key species of indigo plant.

3.4. Materials and Methods

Growing I. tinctoria

I. tinctoria seeds were purchased from Companion Plants, Inc. (Athens, OH) and planted in SuperSoil Potting Soil (Scotts Miracle-Gro, Marysville, OH) in a greenhouse environment. Since woad considered a weed and its cultivation is restricted in California,

a permit was obtained from the state to grow these plants in a controlled greenhouse setting. After seven weeks of growth, the plant leaves were harvested, flash frozen in liquid nitrogen, crushed finely, and stored at -80° C.

Extraction of isatan B from I. tinctoria

Hot water extraction

This hot water extraction protocol was adapted from a protocol found at http://www.woad.org.uk/html/extraction.html. Flash-frozen *I. tinctoria* leaves were ground finely. Water (5x the mass of the leaves) was heated to 80° C, and the frozen ground leaves were added to the water. The leaves were extracted for 10 min and then cooled on ice to less than 55° C within 5 min. The leaves were then removed by centrifugation.

Methanol extraction

Flash-frozen *I. tinctoria* leaves were ground finely and blended in a Waring blender with methanol, at a ratio of 1 g leaves to 5 mL methanol. The mixture was filtered through a 0.2 μm filter and concentrated close to dryness with a rotary evaporator. The concentrate was either resuspended in a small volume of water and filtered, if it was used as a standard or in an assay; or, it was filtered and directly applied to a column to further purify isatan B.

Chromatography of isatan B

Silica gel column

50 g silica beads were soaked in a mixture of 85:15 chloroform to methanol, and the slurry was packed into a column with nitrogen pressure. The concentrate from methanol extraction of woad was applied to the silica column and isocratically eluted with 85:15 chloroform/methanol. Fractions of about 8 mL each were collected. Each fraction was tested for the presence of isatan B by mixing 180 μL of the fraction with 20 μL of 5 M NaOH in a microtiter plate. Fractions with isatan B turned visibly blue within minutes. Those fractions were combined and dried under nitrogen.

Sephadex LH-20 column

20 g Sephadex LH-20 beads were swelled in methanol, and the slurry was packed into a column. The dried fractions from the silica gel column were resuspended in 400 μL methanol, applied to the LH-20 column, and isocratically eluted with methanol. Fractions of about 8 mL each were collected. Each fraction was tested for isatan B in the way

described in "Silica gel column." Fractions with isatan B were combined, dried under nitrogen, and stored at -80° C.

Mass spectrometry for detection of isatan B

Isatan B was detected by LC-MS using a 6520 Accurate-Mass Q-TOF LC-MS (Agilent). Samples that were not first column purified were combined with two volumes of ethyl acetate to extract the isatan B. The top (organic) layer was removed, dried under nitrogen, and resuspended in methanol. Five microliters of the sample were then injected onto a ZORBAX Eclipse Plus C18 4.6 × 100 mm 3.5 μm column (Agilent) using a flow rate of 0.5 mL/min. The solvents were water with 0.1% formic acid and acetonitrile with 0.1% formic acid. The column was flushed with 95% water/5% acetonitrile plus 0.05% ammonium hydroxide for 3 min, before an elution with a linear gradient to 2% water/98% acetonitrile plus 0.05% ammonium hydroxide over 11 min. Isatan B (m/z 292.0821 $[M-H]^-$, R_t 10.5 min) was ionized by electrospray ionization in negative mode using a fragmentor voltage of 100 V.

DCPIP assay for oxidase activity

Overnight cultures are grown of all desired samples. 100 μL of each culture was spun down in a V-bottom microtiter plate, and washed twice with 150 μL PBS. The final pellet was resuspended in 100 μL PBS. The DCPIP assay mixture was prepared: 100 mM potassium phosphate pH 7.0, 150 μM DCPIP, 100 μM phenazine methosulfate, 10mM indican. 10 μL cells in PBS were added, and absorbance at 600 nm was monitored at 2 min intervals for 3 h using an Infinite M1000 PRO microplate reader (Tecan).

Purification of dehydrogenases

G3DHs, PDHs, and wGMC candidates were purified as described in "Expression & purification of PtUGT1 for crystallization," in the **Materials and Methods** section of **Chapter 2**.

Woad protein extraction

1 g crushed frozen *I. tinctoria* leaves were extracted in 5 mL cold extraction buffer, comprised of 100mM MOPS pH 7.0, 1mM DTT, 0.25x cOmplete protease inhibitor (Roche). The lysate was spun down at 20,000 × g for 30 min at 4° C, to clear out solid matter.

Chapter 4. Conclusion

Indigo is the irreplaceable dye for the iconic denim fabric. Without the unique binding properties of indigo on cotton yarns, denim would not have the characteristic fading that makes jeans so beloved. In fact, indigo has been used as a textile dye for thousands of years, and before the 1900s, the dye molecule was extracted from plant leaves. However, with the discovery of a chemical synthesis for indigo at the end of the 19th century, it became increasingly difficult for plant-sourced indigo to compete, especially as denim exploded in popularity in the second half of the 20th century. Unfortunately, indigo synthesis is toxic for the environment, and the rising demand for chemically-synthesized indigo has exacerbated this ecological problem. Industrial indigo dyeing also requires a reducing agent to solubilize the indigo powder into leucoindigo, and this chemical is also harmful to the environment and to the dye workers.

This dissertation has described a novel method for producing and dyeing with indigo without the need for chemical synthesis or reducing agent, based on a process carried out by indigo plants. In *P. tinctorium* leaves, the indigo precursor indoxyl is biosynthesized and then immediately glucosylated to produce indican, thereby preventing indoxyl from spontaneously dimerizing to form indigo. Indican can be stably stored in the vacuole, but if the cell membranes are disrupted then a chloroplast-localized β-glucosidase will come into contact with indican. The β-glucosidase will hydrolyze indican, and the released indoxyl will dimerize to indigo. I adapted this plant-based process to produce indigo via *E. coli* biosynthesis. In this *E. coli* system, a flavin monooxygenase enzyme (FMO) from *M. aminisulfidivorans* was expressed to oxidize the native *E. coli* metabolite indole into indoxyl. Then, I identified the protein sequence of the glucosyltransferase from *P. tinctorium* that glucosylates indoxyl, by isolating the enzyme using chromatography and using tandem mass spectrometry to identify the sequence. This enzyme was also expressed in the *E. coli* host, and it produced indican just like in the plant cell. In another *E. coli* strain, I expressed a β-glucosidase active on indican; combining β-glucosidase and indican produces indoxyl, which quickly oxidizes to leucoindigo and then indigo. I showed that applying biosynthesized indican and β-glucosidase to a cotton textile will dye it indigo blue, because the leucoindigo intermediate can bind to cotton fibers whereas indigo cannot. By scaling up indican production in a bioreactor, I was able to dye a cotton scarf.

Though the strategy of using indican as a dye precursor was shown to be successful, I aimed to improve the process by instead using a related molecule, isatan B. The main benefit isatan B provided was that instead of requiring an enzyme for hydrolysis, isatan

B could be hydrolyzed by increasing the pH past 9. Since an indigo dye bath is typically around pH 11, the conditions for isatan B hydrolysis would have occurred anyways. Since there would be no need for β-glucosidase, production costs for the process would be far lower. Isatan B was also a promising next-generation dye precursor molecule because like indican, it is also produced by indigo plants, so its biosynthesis is possible. Although I was not ultimately able to produce isatan B in *E. coli*, I laid the groundwork for measuring isatan B, developing a strain that could tolerate isatan B, and identifying an enzyme that can convert indican to isatan B. First, I was able to isolate isatan B from *I. tinctoria* leaves and show that its properties matched those described in the literature. I also developed a strain of *E. coli* that had decreased ability to hydrolyze isatan B, so that in the future isatan B could be produced in the cell without fear that it would promptly be degraded. Finally, I adapted a spectrophotometric assay to detect enzymatic oxidation, so that indican oxidation to isatan B could be detected. This would enable a screen to be designed, to identify enzyme variants that can successfully produce isatan B.

Although this indican process is still at a lab scale, it has potential to succeed at an industrial scale. For that to happen, the yield and titer must be improved to cut feedstock and purification costs. This will require engineering the indican metabolic pathway to reduce side products, overexpressing genes, and engineering enzymes in the pathway to be more efficient. Equally important will be implementing ways to reduce the cost of the media, such as eliminating tryptophan by engineering the strain to biosynthesize its own tryptophan precursor, or eliminating IPTG by finding workarounds for inducing the T7 promoter driving the indican pathway. Then, the fermentation must be scaled to larger reactors with high productivity, so that each fermentation run will be cost-effective. The concentration of indican will also need to be increased through downstream processing, since indigo dye bath concentrations are typically upwards of 50 g/L (or 112 g/L indican) to achieve a dark shade. With these process improvements, and by partnering with sustainable denim producers in the supply chain, I believe we can help make the most sustainable jeans for the planet.

Chapter 5. References

1. Balfour-Paul, J. *Indigo*. (Firefly Books, 2011).
2. Paul, R. *Denim*. 1–612 (Elsevier Ltd., 2015).
3. 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).
4. 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).
5. Hsu, T. M. *et al.* Employing a biochemical protecting group for a sustainable indigo dyeing strategy. *Nat Chem Biol* **89,** 44 (2018).
6. Splitstoser, J. C., Dillehay, T. D., Wouters, J. & Claro, A. Early pre-Hispanic use of indigo blue in Peru. *Science Advances* **2,** e1501623–e1501623 (2016).
7. Wolf, L. K. What's That Stuff? Blue Jeans. *C&EN* **89,** 44 (2011).
8. Schimper, C. B., Ibanescu, C. & Bechtold, T. Surface activation of dyed fabric for cellulase treatment. *Biotechnology Journal* **6,** 1280–1285 (2011).
9. Pfleger, J. Process of making indoxyl derivatives. (1901).
10. Blackburn, R. S., Bechtold, T. & John, P. The development of indigo reduction methods and pre-reduced indigo products. *Coloration Technology* **125,** 193–207 (2009).
11. Ensley, B. D. *et al.* Expression of naphthalene oxidation genes in Escherichia coli results in the biosynthesis of indigo. *Science* **222,** 167–169 (1983).
12. 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 Biochemistry* **46,** 788–791 (2011).
13. Padden, A. N. *et al.* An indigo-reducing moderate thermophile from a woad vat, Clostridium isatidis sp. nov. *International Journal of Systematic Bacteriology* **49,** 1025–1031 (1999).
14. Yumoto, I. *et al.* Alkalibacterium psychrotolerans sp. nov., a psychrotolerant obligate alkaliphile that reduces an indigo dye. *International Journal of Systematic and Evolutionary MIicrobiology* **54,** 2379–2383 (2004).
15. Gäng, M., Krüger, R. & Miederer, P. Concentrated leucoindigo solutions. (2002).
16. Roessler, A., Crettenand, D., Dossenbach, O. & Rys, P. Electrochemical reduction of indigo in fixed and fluidized beds of graphite granules. *Journal of Applied Electrochemistry* **33,** 901–908 (2003).
17. 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 Physiology* **41,** 218–225 (2000).
18. Minami, Y. *et al.* β-Glucosidase in the indigo plant: intracellular localization and tissue specific expression in leaves. *Plant Cell Physiology* **38,** 1064–1074 (1997).
19. 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).
20. Chen, J. *et al.* Biosynthesis of the active compounds of Isatis indigotica based on transcriptome sequencing and metabolites profiling. *BMC Genomics* **14,** 857–13 (2013).
21. Tang, X. *et al.* High-throughput sequencing and de novo assembly of the Isatis indigotica transcriptome. *PLoS ONE* **9,** e102963–8 (2014).
22. Minami, Y., Sarangi, B. K. & Thul, S. T. Transcriptome analysis for identification of indigo biosynthesis pathway genes in Polygonum tinctorium. *Biologia* **70,** 1026–1032 (2015).
23. Bechtold, T. & Mussak, R. *Handbook of Natural Colorants*. 1–432 (John Wiley & Sons, 2009).
24. Gilbert, K. G. *et al.* Quantitative analysis of indigo and indigo precursors in leaves of Isatis spp. and Polygonum tinctorium. *Biotechnol. Prog.* **20,** 1289–1292 (2004).
25. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29,** 644–652 (2011).
26. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7,** 539 (2011).
27. de Beer, T. A. P., Berka, K., Thornton, J. M. & Laskowski, R. A. PDBsum additions. *Nucleic Acids Research* **42,** D292–6 (2014).
28. Mackenzie, P. I. *et al.* The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* **7,** 255–269 (1997).
29. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42,** D490–D495 (2014).
30. Wang, X. Structure, mechanism and engineering of plant natural product glycosyltransferases. *FEBS Letters* **583,** 3303–3309 (2009).
31. 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 USA* **104,** 20238–20243 (2007).
32. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).

33. Osmani, S. A., Bak, S. & Møller, B. L. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry* **70,** 325–347 (2009).
34. Loutre, C. *et al.* Isolation of a glucosyltransferase from Arabidopsis thaliana active in the metabolism of the persistent pollutant 3,4-dichloroaniline. *The Plant Journal* **34,** 485–493 (2003).
35. Nakamura, C. E. & Whited, G. M. Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* **14,** 454–459 (2003).
36. Yim, H. *et al.* Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* **7,** 445–452 (2011).
37. Patnaik, R. & Liao, J. C. Engineering of Escherichia coli central metabolism for aromatic metabolite production with near theoretical yield. *Applied and Environmental Microbiology* **60,** 3903–3908 (1994).
38. Malla, S., Pandey, R. P., Kim, B.-G. & Sohng, J. K. Regiospecific modifications of naringenin for astragalin production in Escherichia coli. *Biotechnol. Bioeng.* **110,** 2525–2535 (2013).
39. Lim, C. G. *et al.* Development of a recombinant Escherichia coli strain for overproduction of the plant pigment anthocyanin. *Applied and Environmental Microbiology* **81,** 6276–6284 (2015).
40. Choi, H. S. *et al.* A novel flavin-containing monooxygenase from Methylophaga sp. strain SK1 and its indigo synthesis in Escherichia coli. *Biochemical and Biophysical Research Communications* **306,** 930–936 (2003).
41. Anderson, J. C. *et al.* BglBricks: A flexible standard for biological part assembly. *Journal of Biological Engineering* **4,** 1 (2010).
42. Li, G. & Young, K. D. A cAMP-independent carbohydrate-driven mechanism inhibits tnaA expression and TnaA enzyme activity in Escherichia coli. *Microbiology* **160,** 2079–2088 (2014).
43. Botsford, J. L. & DeMoss, R. D. Catabolite Repression of Tryptophanase in Escherichia coli. *Journal of Bacteriology* **105,** 303–312 (1971).
44. Minami, Y., Kanafuji, T. & Miura, K. Purification and characterization of a β-glucosidase from Polygonum tinctorium, which catalyzes preferentially the hydrolysis of indican. *Bioscience, Biotechnology, and Biochemistry* **60,** 147–149 (1996).
45. Song, J., Imanaka, H., Imamura, K., Kajitani, K. & Nakanishi, K. Development of a highly efficient indigo dyeing method using indican with an immobilized β-glucosidase from Aspergillus niger. *Journal of Bioscience and Bioengineering* **110,** 281–287 (2010).

46. Kim, J.-Y., Lee, J.-Y., Shin, Y.-S. & Kim, G.-J. Characterization of an indican-hydrolyzing enzyme from Sinorhizobium meliloti. *Process Biochemistry* **45,** 892–896 (2010).
47. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2,** 473–11 (2006).
48. Gehauf, B. & Goldenson, J. Detection and estimation of nerve gases by fluorescence reaction. *Anal. Chem.* **29,** 276–278 (1957).
49. Paavilainen, S., Hellman, J. & Korpela, T. Purification, characterization, gene cloning, and sequencing of a new β-glucosidase from Bacillus circulans subsp. alkalophilus. *Applied and Environmental Microbiology* **59,** 927–932 (1993).
50. Hansen, E. H. *et al.* De novo biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (Saccharomyces cerevisiae). *Applied and Environmental Microbiology* **75,** 2765–2774 (2009).
51. Moehs, C. P., Allen, P. V., Friedman, M. & Belknap, W. R. Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. *The Plant Journal* **11,** 227–236 (1997).
52. Etters, J. N. Advances in indigo dyeing: Implications for the dyer, apparel manufacturer and environment. *Textile Chemist Colorist* **27,** 17–22 (1995).
53. Sternberg, D., Vijayakumar, P. & Reese, E. T. beta-Glucosidase: microbial production and effect on enzymatic hydrolysis of cellulose. *Can. J. Microbiol.* **23,** 139–147 (1977).
54. Jäger, S., Brumbauer, A., Fehér, E., Réczey, K. & Kiss, L. Production and characterization of β-glucosidases from different Aspergillus strains. *World Journal of Microbiology and Biotechnology* **17,** 455–461 (2001).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
56. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
57. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research* **4,** (2015).
58. 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).
59. 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).
60. Towns, J. *et al.* XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **16,** 62–74 (2014).

61. Nystrom, N., Welling, J., Blood, P. D. & Goh, E. L. in *Contemporary High Performance Computing* (ed. Vetter, J. S.) 421–440 (CRC Press, 2013).
62. 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. *Journal of the American Society for Mass Spectrometry* **5,** 976–989 (1994).
63. Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of Proteome Research* **1,** 21–26 (2002).
64. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4,** 975–986 (2015).
65. Bond-Watts, B. B., Bellerose, R. J. & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat Chem Biol* 1–6 (2011). doi:10.1038/nchembio.537
66. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97,** 6640–6645 (2000).
67. Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J. & Schweizer, H. P. A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene* **212,** 77–86 (1998).
68. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Meth* **10,** 659–664 (2013).
69. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43,** 186–190 (2010).
70. Kabsch, W. XDS. *Acta Crystallographica Section D Biological Crystallography* **66,** 125–132 (2010).
71. McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of Applied Crystallography* **40,** 658–674 (2007).
72. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66,** 213–221 (2010).
73. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography* **68,** 352–367 (2012).
74. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66,** 486–501 (2010).

75. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66,** 12–21 (2010).
76. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Research* **43,** W443–7 (2015).
77. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38,** W545–W549 (2010).
78. Itaya, K. & Ui, M. A new micromethod for the colorimetric determination of inorganic phosphate. *Clinica Chimica Acta* **14,** 361–366 (1966).
79. McKee, J. R. & Zanger, M. A microscale synthesis of indigo: Vat dyeing. *Journal of Chemical Education* **68,** A242 (1991).
80. Oberthür, C., Schneider, B., Graf, H. & Hamburger, M. The elusive indigo precursors in woad (Isatis tinctoria L.) – identification of the major indigo precursor, isatan A, and a structure revision of isatan B. *C&B* **1,** 174–182 (2004).
81. 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).
82. Epstein, E., Nabors, M. W. & Stowe, B. B. Origin of indigo of woad. *Nature* **216,** (1967).
83. Kokubun, T., Edmonds, J. & John, P. Indoxyl derivatives in woad in relation to medieval indigo production. *Phytochemistry* **49,** 79–87 (1998).
84. Hirokawa, Y. *et al.* Genetic manipulations restored the growth fitness of reduced-genome Escherichia coli. *Journal of Bioscience and Bioengineering* **116,** 52–58 (2013).
85. Pósfai, G. *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312,** 1044–1046 (2006).
86. Yu, T.-W. *et al.* Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in Amycolatopsis mediterranei S699. *Journal of Biological Chemistry* **276,** 12546–12555 (2001).
87. Hayano, K. & Fukui, S. Purification and properties of 3-ketosucrose-forming enzyme from the cells of Agrobacterium tumefaciens. *Journal of Biological Chemistry* **242,** (1967).
88. van Beeumen, J. & de Ley, J. Hexopyranoside: cytochrome c oxidoreductase from Agrobacterium tumefaciens. *European Journal of Biochemistry* **6,** 331–343 (1968).
89. Fukui, S. Conversion of glucose-1-phosphate to 3-keto-glucose-1-phosphate by cells of Agrobacterium tumefaciens. *Journal of Bacteriology* **97,** 793–798 (1969).

90. Klekner, V., Löbl, V., Šímová, E. & Novák, M. Conversion of disaccharides to 3-ketodisaccharides by nongrowing and immobilized cells of Agrobacterium tumefaciens. *Folia Microbiol* **34,** 286–293 (1989).
91. Schuerman, P. L., Liu, J. S., Mou, H. & Dandekar, A. M. 3-Ketoglycoside-mediated metabolism of sucrose in E. coli as conferred by genes from Agrobacterium tumefaciens. *Appl Microbiol Biotechnol* **47,** 560–565 (1997).
92. Maeda, A., Kataoka, H., Adachi, S. & Matsuno, R. Transformation of cellobiose to 3-ketocellobiose by the EDTA-treated Agrobacterium tumefaciens cells. *Journal of Bioscience and Bioengineering* **95,** 608–611 (2003).
93. Zhang, J.-F., Zheng, Y.-G., Xue, Y.-P. & Shen, Y.-C. Purification and characterization of the glucoside 3-dehydrogenase produced by a newly isolated Stenotrophomonas maltrophilia CCTCC M 204024. *Appl Microbiol Biotechnol* **71,** 638–645 (2005).
94. Zhang, J.-F., Zheng, Y.-G. & Shen, Y.-C. Study on optimal production of 3-ketovalidoxylamine A C-N lyase and glucoside 3-dehydrogenase by a newly isolated Stenotrophomonas maltrophilia. *African Journal of Biotechnology* **8,** 5493–5499 (2009).
95. Kojima, K., Tsugawa, W., Hamahuji, T., Watazu, Y. & Sode, K. Effect of growth substrates on production of new soluble glucose 3-dehydrogenase in Halomonas (Deleya) sp. α-15. *Applied Biochemistry and Biotechnology* **79,** 827–834 (1999).
96. Kojima, K., Tsugawa, W. & Sode, K. Cloning and expression of glucose 3-dehydrogenase from Halomonas sp. α-15 in Escherichia coli. *Biochemical and Biophysical Research Communications* **282,** 21–27 (2001).
97. Asano, N., Takeuchi, M., Ninomiya, K., Kameda, Y. & Matsui, K. Microbial degradation of validamycin A by *Flavobacterium saccharophilum*. *J. Antibiot.* **37,** 859–867 (1984).
98. Takeuchi, M. *et al.* Purification and properties of glucoside 3-dehydrogenase from Flavobacterium saccharophilum. *The Journal of Biochemistry* **100,** 1049–1055 (1986).
99. Takeuchi, M., Asano, N., Kameda, Y. & Matsui, K. Purification and properties of soluble D-glucoside 3-dehydrogenase from Flavobacterium saccharophilum. *Agricultural and Biological Chemistry* **52,** 1905–1912 (1988).
100. Zhang, J.-F., Chen, W., Ke, W. & Chen, H. Screening of a glucoside 3-dehydrogenase-producing strain, Sphingobacterium faecium, based on a high-throughput screening method and optimization of the culture conditions for enzyme production. *Applied Biochemistry and Biotechnology* **172,** 3448–3460 (2014).
101. Zhang, J.-F., Yang, B., Chen, W. & Chen, J. Purification and characterization of the glucoside 3-dehydrogenase produced by a newly isolated Sphingobacterium

faecium ZJF-D6 CCTCC M 2013251. *Applied Biochemistry and Biotechnology* **172,** 3913–3925 (2014).
102. Zhang, J.-F. Gene cloning and expression of a glucoside 3-dehydrogenase from Sphingobacterium faecium ZJF-D6, and used it to produce N-p-nitrophenyl-3-ketovalidamine. *World Journal of Microbiology and Biotechnology* **33,** 1–9 (2017).
103. Volc, J., Kubátová, E., Wood, D. A. & Daniel, G. Pyranose 2-dehydrogenase, a novel sugar oxidoreductase from the basidiomycete fungus Agaricus bisporus. *Archives of Microbiology* **167,** 119–125 (1997).
104. Volc, J., Sedmera, P., Halada, P., Přikrylova, V. & Daniel, G. C-2 and C-3 oxidation of D-Glc, and C-2 oxidation of D-Gal by pyranose dehydrogenase from Agaricus bisporus. *Carbohydrate Research* **310,** 151–156 (1998).
105. Morrison, S. C., Wood, D. A. & Wood, P. M. Characterization of a glucose 3-dehydrogenase from the cultivated mushroom (Agaricus bisporus). *Appl Microbiol Biotechnol* **51,** 58–64 (1999).
106. Gonaus, C., Kittl, R., Sygmund, C., Haltrich, D. & Peterbauer, C. K. Transcription analysis of pyranose dehydrogenase from the basidiomycete Agaricus bisporus and characterization of the recombinantly expressed enzyme. *Protein Expression and Purification* **119,** 36–44 (2016).
107. Kittl, R. *et al.* Molecular cloning of three pyranose dehydrogenase-encoding genes from *Agaricus meleagris* and analysis of their expression by real-time RT-PCR. *Current Genetics* **53,** 117–127 (2008).
108. Sygmund, C. *et al.* Characterization of pyranose dehydrogenase from Agaricus meleagris and its application in the C-2 specific conversion of d-galactose. *Journal of Biotechnology* **133,** 334–342 (2008).
109. Tan, T. C. *et al.* The 1.6 Å crystal structure of pyranose dehydrogenase from Agaricus meleagris rationalizes substrate specificity and reveals a flavin intermediate. *PLoS ONE* **8,** e53567–14 (2013).
110. Jin, L.-Q. & Zheng, Y.-G. Properties of glucoside 3-dehydrogenase and its potential applications. *African Journal of Biotechnology* **7,** 4843–4849 (2008).
111. Peterbauer, C. K. & Volc, J. Pyranose dehydrogenases: biochemical features and perspectives of technological applications. *Appl Microbiol Biotechnol* **85,** 837–848 (2009).
112. Sedmera, P. *et al.* New biotransformations of some reducing sugars to the corresponding (di)dehydro(glycosyl) aldoses or aldonic acids using fungal pyranose dehydrogenase. *Journal of Molecular Catalysis B: Enzymatic* **41,** 32–42 (2006).
113. Sygmund, C. *et al.* Simple and efficient expression of Agaricus meleagris pyranose dehydrogenase in Pichia pastoris. *Appl Microbiol Biotechnol* **94,** 695–704 (2011).

114. Peters, B. *et al.* Characterization of membrane-bound dehydrogenases from Gluconobacter oxydans 621H via whole-cell activity assays using multideletion strains. *Appl Microbiol Biotechnol* **97,** 6397–6412 (2013).
115. Yum, D.-Y., Lee, Y.-P. & Pan, J.-G. Cloning and expression of a gene cluster encoding three subunits of membrane-bound gluconate dehydrogenase from Erwinia cypripedii ATCC 29267 in Escherichia coli. *Journal of Bacteriology* **179,** 6566–6572 (1997).
116. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. *Genome Informatics* 205–211 (2009). doi:10.1142/9781848165632_0019
117. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44,** D279–85 (2016).