Metabolic Engineering using Computer Simulation Predictions and Protein Expression Studies towards Anthocyanin Production in *Escherichia coli*

by

Hila Dvora
August 11, 2011

A thesis submitted to the
Faculty of the Graduate School of
the University at Buffalo, State University of New York
in partial fulfillment of the requirements for the
degree of

Master of Science

Department of Chemical and Biological Engineering

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Anthocyanin is a water-soluble plant pigment in the flavonoid family that has recently drawn significant interest for its potential health benefits. This high-valued compound can be produced in *Escherichia coli* by heterologous expression of the plant proteins anthocyanidin synthase (ANS) and UDP-glucose:flavonoid 3-*O*-glucosyltransferase (3GT), when supplemented with the substrate (+)-catechin. Metabolic engineering strategies are implemented to improve anthocyanin production in *E. coli*.

Two computer models were used to predict metabolic interventions (gene overexpressions and knockouts) that may direct carbon flux away from non-essential pathways and towards anthocyanin production. The cipher for evolutionary design (CiED) model predicted gene knockouts that primarily aimed at increased bioavailability of 2-oxoglutarate, a substrate of ANS. The OptForce model predicted overexpressions and knockouts that primarily aimed to improve the bioavailability of UDP-glucose, used by 3GT to glycosylate the anthocyanin. Expression studies were performed to investigate ANS and 3GT expression in *E. coli*.

Anthocyanin production fermentations using the single-deletion strains predicted by the two computer models did not significantly improve anthocyanin production when compared to the control strain. Expression studies were performed using SDS-PAGE and western blot analysis. When fermentation samples were separated into soluble and insoluble protein fractions, 3GT was found to be mostly insoluble. The insolubility leads to incorrect protein folding, which likely leads to reduction or loss of activity. The low levels of soluble protein expression may explain the fact that increased precursor bioavailability did not improve anthocyanin production yields.

Chapter 1: Introduction

1.1 Flavonoids and Anthocyanins

Natural products have long played a significant role in drug discovery and development [1-3]. More than a quarter of all small molecule drugs produced from 1981 to the middle of 2006 are natural products or derivatives thereof. Furthermore, almost half of all drugs that are produced completely synthetically are natural product mimics and/or contain a pharmacophore derived from a natural product [1]. Flavonoids comprise one class of natural products that has drawn significant attention in recent years for their potential as drug targets [4-6].

1.1.1 Background

Flavonoids are plant phytochemicals that play important physiological roles including UV protection, antimicrobial activity, and providing color that attracts pollinators [5, 7]. The core flavonoid structure is composed of two benzene rings linked together by a pyran or pyrone ring. The possible variation in chemical substitutions on the three-ring core, such as hydroxylation, methylation, and glycosylation, leads to a vast diversity in structure among flavonoids [6, 8]. Over 9000 different naturally occurring compounds have been discovered [9]. Flavonoids are classified into six major subgroups: chalcones, flavones, flavonols, leucoanthocyanidins, flavan-3-ols, and anthocyanins. Over the past two decades, these polyphenolic compounds have attracted some interest for their antioxidative and nutraceutical effects, but they recently have been more extensively investigated for their potential as treatments for various diseases such as cancer, diabetes, and cardiovascular diseases [6, 10-12].

Anthocyanins make up one of the major subclasses of flavonoids. These water-soluble plant pigments are characterized by a flavylium ion, to which their intense color is attributed [13], and at least one attached sugar group [14]. The aglycone form is called anthocyanidin. The value of anthocyanins to humans has mostly been for their use as neutraceuticals and natural food coloring. The latter is particularly significant with the increased awareness that artificial food coloring is linked to attention deficit hyperactivity disorder (ADHD) in children [15]. Recently, there has been increasing interest in anthocyanins due to their unique medicinal and nutraceutical properties among all flavonoids [16-19], as will be discussed later.

1.1.2 The flavonoid biosynthetic pathway

The elucidation of the flavonoid biosynthetic pathway has opened the door to metabolic engineering for flavonoid production [20, 21]. The pathway begins with a propanoic acid, which is activated by 4-coumarate:coenzyme A ligase (4CL). A coenzyme A (CoA) moiety is attached to the acid, giving a propanoyl-CoA ester. The addition of three malonyl-CoA molecules by chalcone synthase (CHS) commits the resulting chalcone to the flavonoid biosynthetic pathway [7]. Chalcone isomerase (CHI) isomerizes chalcones selectively to (2*S*)-flavanones [22], which are then hydroxylated by flavanone 3*β*-hydroxylase (FHT) in the 3-carbon position to give dihydroflavonols. These are reduced by dihydroflavonol 4-reductase (DFR) in the 4-carbon position, yielding the unstable leucoanthocyanidins. These, in turn, are reduced by leucoanthocyanidin reductase (LAR) to give flavan-3-ols. Both the leucoanthocyanidins and the flavan-3-ols are possible substrates for anthocyanidin synthase (ANS), which catalyzes the reaction to anthocyanidins. Finally, UDP-glucose:flavonoid 3-*O*-glucosyltransferase (3GT) catalyzes the glycosylation at the 3-carbon, yielding anthocyanins [9].

For the production of anthocyanins in microorganisms, the starting material supplied to the culture is (+)-catechin (also called flavan-3-ol), and the heterologous enzymes ANS and 3GT catalyze reactions 1-1 and 1-2, respectively, to yield anthocyanins (Figure 1-1). Most of the product is excreted such that it can be extracted directly from the media broth.

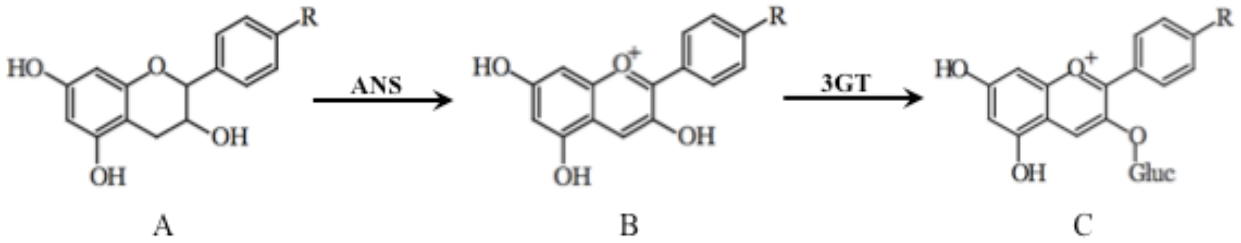
1-1 (+)-catechin + 2-oxoglutarate + O_2 → anthocyanidin + succinate + CO_2 + $2H_2O$

1-2 anthocyanidin + UDP-glucose → anthocyanin + UDP

1.1.3 Anthocyanins and human health

While all flavonoids are thought to carry important health benefits, anthocyanins in particular are being studied for their unique contribution to human health [16, 19, 23]. Some of the most widely studied properties include the role of anthocyanins in the prevention of cancer and cardiovascular disease, as well as their potential as a candidate for treatment of Type II diabetes.

The suggested preventative effect of anthocyanins against cancer is attributed, at least in part, to their antioxidative activity [24]. Anthocyanins have been tested *in vitro* in several systems including colon, liver, breast, and leukemic cells, and were found to yield a variety of anticarcinogenic effects [18]. Berry extracts rich in anthocyanins have also been used to test for chemopreventative activity *in vivo* using mice and rats [25, 26]. These, too, show effects against various biomarkers of cancer.



A B C

Figure 1-1. Anthocyanin pathway reactions from (+)-catechin. Substrate (+)-catechin (A) is supplied to the media broth of the *E. coli* culture. Plant proteins anthocyanidin synthase (ANS) and UDP-glucose:flavonoid 3-*O*-glucosyltransferase (3GT) are expressed heterologously to catalyze the conversion of (+)-catechin to anthocyanidin (B), which is glycosylated to yield anthocyanin (C).

Anthocyanins have also been credited with preventative effects against cardiovascular diseases. This is shown through *in vitro* studies by measurement of vascular cell model responses in studies of gene and protein expression [16]. For example, in a study by Kim et al., cardiovascular disease, characterized by ischemia and reperfusion of the heart, was induced in bovine aortic endothelial cells. Treatment with anthocyanin extract from black soybean was found to regulate the response through a family of transcription factors known to be involved in inflammatory response and in control of cell adhesion and growth [27].

The sugar group attached to the flavonoid core may make anthocyanins a promising candidate for treatment of Type II diabetes. This chronic condition is caused by an inability of cells to efficiently use insulin from the blood for glucose uptake. Patients therefore require increased levels of insulin, but over time their pancreas loses its ability to make it [28].

In a series of studies performed by Matsui et al., the potential of anthocyanins to act as anti-diabetic agents was investigated. These studies focused on the capacity of anthocyanins to inhibit rat intestinal α-glucosidase. In the first part of a two-part study reported in 2001, the group showed that plant extracts of anthocyanins did inhibit α-glucosidase activity against maltose, although not against sucrose [29]. The inhibition was enhanced during the immobilized α-glucosidase assay, which mimicked the membrane-bound state of the enzyme. The second part of the study followed up to confirm that the α-glucosidase inhibition was due to the anthocyanins and not other compounds in the extracts [30]. It was confirmed that the most active compounds were indeed acylated anthocyanins. The following year, the group sought to show the *in vivo* effects of anthocyanins on blood glucose levels [31]. They found that a single dose of

anthocyanin extract reduced the rate of increase of the blood glucose level in rats following carbohydrate consumption.

In 2005, McDougall et al. investigated whether polyphenol-rich extracts inhibited human salivary α-amylase and rat intestinal α-glucosidase [32]. Extracts were taken from several different berries and from red cabbage and green tea. After identifying which extracts were particularly effective inhibitors, the group worked to determine what the active compounds were. They fractionated the extracts and found that fractions containing high levels of tannins were eliciting the inhibitory effects on α-amylase, while the fractions rich in anthocyanins elicited a stronger inhibitory effect on α-glucosidase.

The findings of the research of health effects of anthocyanins merit further investigation of these compounds. However, it is necessary to obtain pure samples of the compounds in sufficient quantities for investigation and for production as a drug target by a cost-effective method.

1.1.4 Approaches to obtaining flavonoids

The medicinal effects of flavonoids are generally studied through the use of plant extracts. Using extracts can have some advantage in that synergistic activity between two or more compounds in the extract may elicit more pronounced effects than any one compound would on its own [33]. However, the use of extracts has several weaknesses. First, the chemical composition of plant extracts is not consistent. Varying environmental conditions such as location of growth or season of harvest can lead to substantial variation in extract composition [9]. Such variance can negatively impact the reproducibility of results of medicinal studies. Another drawback is that the natural availability of particular compounds is very limited. While simple compounds are

often ubiquitous, compounds with more complex chemical structures are the products of complicated biosynthetic pathways involving several steps and enzymes. Such compounds are typically only produced by the plant as a response to certain environmental stimuli. Therefore, compounds with complex chemistry could be missed when crude extracts are used since they are likely to be at low concentrations or absent altogether. Furthermore, even if such a compound were present and found to yield the desired effect, it would be extremely difficult and expensive to produce on a large scale if purified solely from extracts.

For these reasons, the use of pure compounds should complement the plant extract studies. In the event that compounds in an extract do elicit a synergistic effect, mixtures of pure compounds can be tested at various known concentrations and ratios in order to elucidate the source of the suggested synergy. This is especially relevant as technology for high-throughput assays and analyses becomes more widely available.

Several approaches may be used to obtain pure samples of compounds of interest. One method is chemical synthesis of the desired compound. The chemical synthesis of the flavonoid core structure has been extensively investigated [34, 35]. While the synthesis of several flavonoid compounds is straightforward on a small scale, it relies on expensive precursors and toxic material; as a result, its scale-up remains an engineering conundrum. Furthermore, the organic synthesis of more elaborate flavonoids, such as the glycosylated anthocyanins, requires several steps, which result in low overall yields. Finally, the possible structure modifications that can be generated are limited by the available methods. For example, specific hydroxylation and

glycosylation is either not possible at all or only feasible after several steps, again leading to prohibitively low yields.

Another approach is expression in plant cell culture. This has also been studied extensively; however, due to inherent problems, scalability of such production has not been realized [36]. One challenge is that when scale-up of plant cell culture has been attempted, the cells formed aggregates. Cells in aggregates were not properly supplied with nutrients and substrate from the media, and therefore did not produce the target compounds with good yields. Another challenge is posed by competing pathways drawing resources away from the heterologous flavonoid pathway, restricting production of the desired compounds [37].

Therefore, the most promising approach for obtaining significant quantities of pure flavonoid compounds is the use of recombinant microorganisms for production. With recent advances in genomics research and knowledge of the biosynthetic pathways, the use of recombinant organisms for flavonoid production can provide an attractive alternative to chemical synthesis, plant cell culture, or purification from extracts. Indeed, efficient flavonoid production in microorganisms has been demonstrated in several recent publications [9, 22, 38-44]. Once the heterologous proteins are expressed and active in the microorganism, production can be increased using engineering strategies, such as gene deletions and overexpressions, to direct carbon-flow away from non-essential pathways and towards increased precursor bioavailability [22, 41]. The efficiency can be further increased by using the best plant source for the genes that comprise the recombinant flavonoid biosynthetic pathways [9]. In addition, fermentation variables such as type of media, carbon source, induction levels, and oxygen availability may be

optimized for production of particular target compounds in *E. coli*, as has been done in the past [9, 38].

1.2 Metabolic Engineering to Improve Anthocyanin Production

One of the most successful strategies in metabolic engineering is directing carbon-flow towards production of a compound of interest. Once the heterologous enzymes are functionally expressed within the host organism, perturbations can be made to the homologous metabolic network in the form of gene knockouts (referred to synonymously in this work as gene deletions) and gene overexpressions.

The plasmid containing the heterologous genes was available for this study from previous work with anthocyanins [9]. This plasmid was the pCDFDuet-1 vector (Novagen) containing ANS from *Petunia hybrida* (PhANS) and 3GT from *Arabidopsis thaliana* (At3GT). Some rational overexpression targets that were previously found to improve anthocyanin production in *E. coli* include *pgm*, *galU*, *cmk*, *ndk* (Figure 1-2). Overexpression of *pgm* and *galU* promote flux from glucose to UDP-glucose. When the overexpressions *ndk* and *cmk* are included, along with supplementation of orotic acid, there is increased availability of UTP for use by *galU* to produce UDP-glucose. Information on these genes and plasmids can be found in Appendix A, along with the plasmids that carry them. These overexpressions drive the carbon-flow towards UDP-glucose, which is believed to be the limiting reactant in anthocyanin production [45].

A growing trend in metabolic engineering is the use of computer models of the genetic network of host microorganisms to improve production of a compound of interest. Such a model uses all

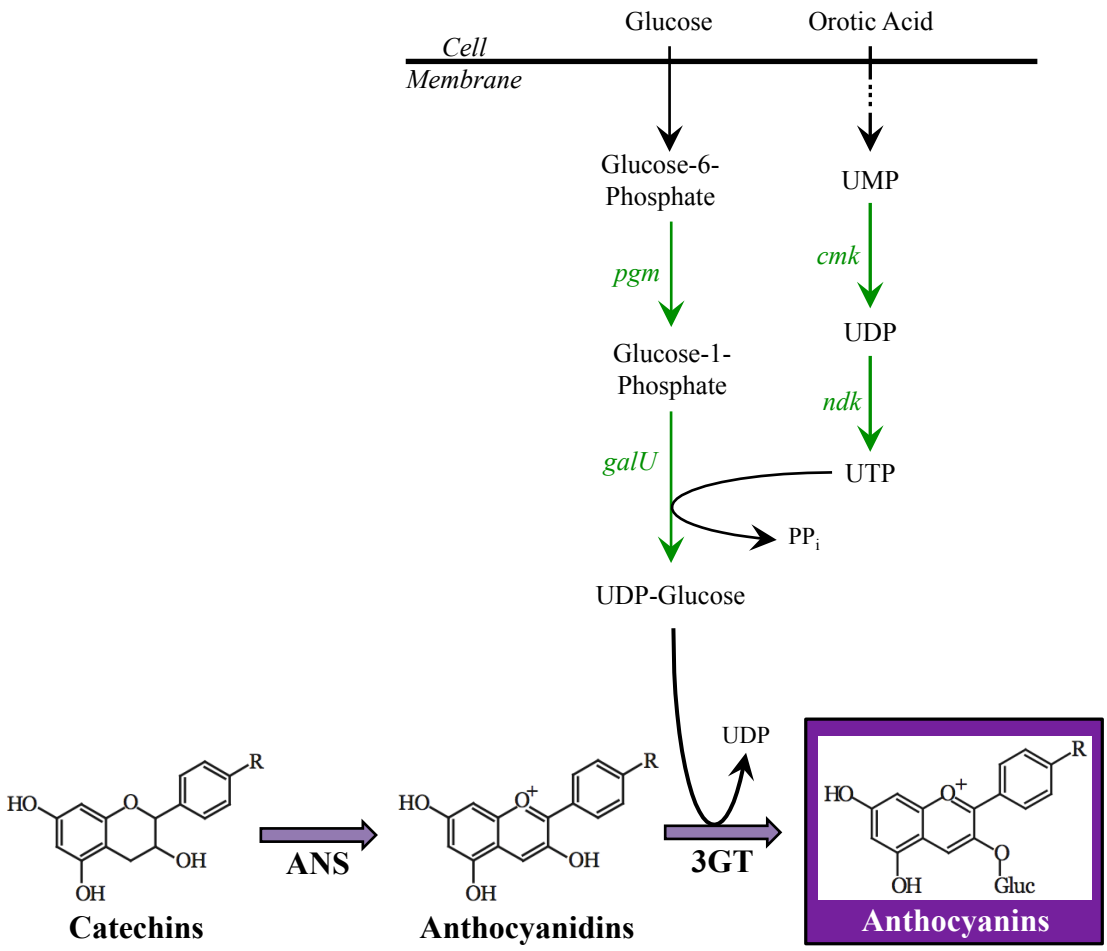


Figure 1-2. Rational gene overexpressions (green arrows) to improve UDP-glucose bioavailability for the production of anthocyanins. Overexpressing *pgm* and *galU* together directs carbon-flow from glucose-6-phosphate to UDP-glucose, a key precursor in anthocyanin production. Supplementation with orotic acid along with overexpression of *cmk* and *ndk* increases the pool of UTP available for UDP-glucose synthesis.

known information on the metabolic network of the producing organism to recommend metabolic perturbations that direct carbon-flow away from less essential processes and towards the production of the compounds of interest. Since *E. coli* in particular has been very well characterized, it is an excellent candidate for such work.

1.3 Project Summary

In this work two attempts to improve anthocyanin production in *E. coli* through metabolic perturbations recommended by computer models are presented. The first model, cipher for evolutionary design (CiED), recommended several gene knockouts that aimed primarily to increase the bioavailability of the ANS substrate 2-oxoglutarate. When the modified strains did not yield the expected improvement in anthocyanin production, predictions from a second model were implemented. This model, the OptForce procedure, predicted target gene knockouts and overexpressions that primarily aimed to increase bioavailability of the key precursor UDP-glucose. Two of the target deletions were aimed at decreased acetate production, which is toxic to the organism. The reduced acetate did contribute to improved anthocyanin production, but the increased UDP-glucose levels did not. This unexpected result led to the conclusion that precursor bioavailability may not be the limiting factor in anthocyanin production in *E. coli*.

Therefore, expression studies of the pathway enzymes ANS and 3GT were performed to investigate whether the heterologous expression of these proteins was successful. Initial studies showed that expression of ANS was far greater than expression of 3GT. Furthermore, the 3GT that was expressed was found to be mostly insoluble. Since insoluble protein is not folded correctly, it cannot properly catalyze the reaction. This may be the primary limitation in the

current anthocyanin production platform. Several variables were studied to test for improved solubility of heterologous 3GT, including host strain, inducer concentration, and expression temperature.

Chapter 2: Applying the Cipher for Evolutionary Design Model to Anthocyanin Production in *E. coli*

2.1 Introduction

With anthocyanin being such a high-value compound, it is desirable to produce it in significant quantities as a pure substance. *E. coli* containing the heterologous proteins anthocyanidin synthase from *Petunia hybrida* (PhANS) and UDP-glucose:flavonoid 3-*O*-glucosyltransferase from *Arabidopsis thaliana* (At3GT) was shown to be a successfully production platform. Some improvement to production had already been achieved through overexpressions of *E. coli* genes *pgm*, *galU*, *cmk*, and *ndk* (Figure 1-2, Appendix A), which increased the bioavailability of UDP-glucose, a key precursor for anthocyanin production.

The work presented in this chapter investigates whether anthocyanin production in *E. coli* can be further improved through gene deletions that may direct carbon-flow away from non-essential processes and towards anthocyanin biosynthesis. In order to find target gene deletions, the cipher for evolutionary design (CiED) was implemented. CiED was previously used to find target metabolic interventions that improved flavonoid production when applied towards the production of flavanones in *E. coli* [46]. Therefore, in this work we sought to similarly improve anthocyanin production by running the CiED model to find gene knockout targets, and constructing the recommended strains experimentally.

2.2 Theory

2.2.1 Cipher for evolutionary design

The CiED stoichiometric model was used in this work to obtain target gene knockouts for improved anthocyanin production in *E. coli*.

The model was based on an existing *E. coli* model [47] and has been expanded to include biochemical processes that were discovered since the basic model was published. For each application of CiED, the reaction fluxes and assumed metabolite fluxes of the heterologous pathway reactions are included. In this work, the target compound is anthocyanin, which is produced from (+)-catechin by the enzymes ANS and 3GT. Therefore, the fluxes of the reactions catalyzed by ANS and 3GT were included in the simulated metabolic network.

The central algorithm in this program (Figure 2-1) begins with an initial population of individual genotypes with different deleted genes. The flux profile and resulting fitness of each individual is determined. Individuals with zero cell growth are eliminated with a poor fitness score. Individuals with elevated growth and production fluxes are given good fitness scores. The individuals of the population are ranked by this fitness score and are used to generate the next population through crossovers and mutations. The five best strains of each generation are allowed to continue to the next generation unchanged. The cycle of selection, crossover, and mutations repeats until the generation limit of 1000 generations is reached, or until the best individual does not improve over a specified stall limit, set to a default of 50 generations. The best individuals are tested for production experimentally.

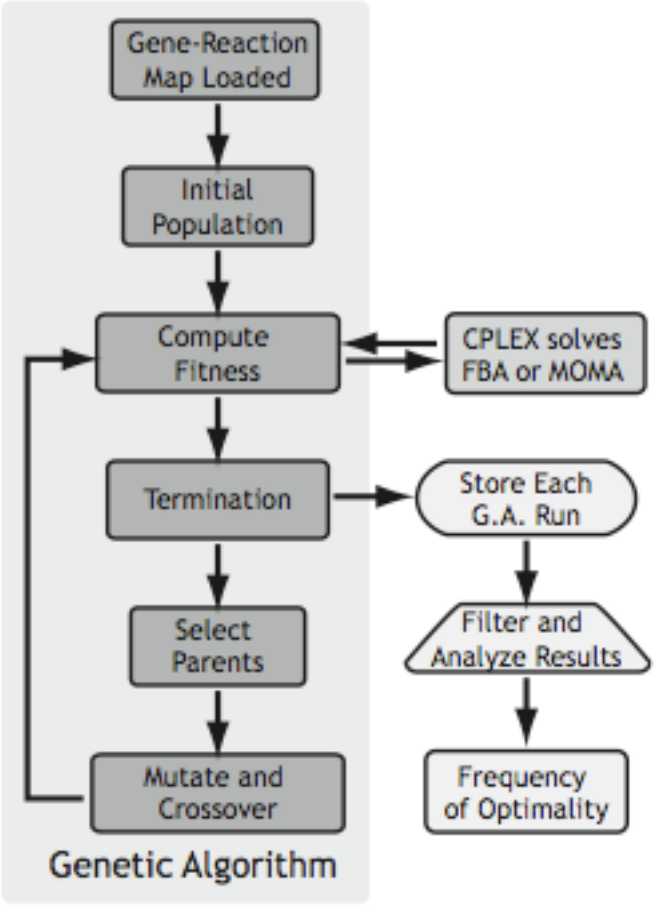


Figure 2-1. Schematic of CiED algorithm. The traditional genetic algorithm elements are in the shaded box.

Image reproduced with permission from the American Society for Microbiology (license number 2712040624538; DOI 10.1128/AEM.00270-09). Originally appeared in Applied and Environmental Microbiology, 2009. **75**(18): p. 5831-5839.

2.2.2 Gene knockouts in *E. coli* using λ–Red recombinase

To carry out the target gene deletions recommended by the CiED model, the λ–Red recombinase method of Datsenko and Wanner [48] was used. To knock out a gene, λ–Red recombinase is first used to replace the gene target with the kanamycin resistance gene (*kanR*) flanked by flippase recognition target (FRT) sites. The kanamycin marker is subsequently excised out by a flippase.

In this PCR-based method, the deletion cassette shown in Figure 2-2 is first constructed. The template plasmid pKD4 is used to amplify the kanamycin resistance gene flanked by FRT sites. The forward and reverse primers are designed such that each has a 20-basepair (bp) annealing section to pKD4, as well as a tail of 40-60 bps that are homologous to the ends of the gene targeted for deletion.

The host strain is made to be electrocompetent while expressing the λ–Red recombinase such that when the deletion cassette is transformed, the phage recognizes the 40-60 bp homology sequences on either side of the deletion cassette. The recombinase performs the recombination event, incorporating the transformed kanamycin cassette into the *E. coli* chromosome in place of the targeted gene. The temperature-sensitive helper plasmid can be expelled later by growth at temperatures above 30°C.

When a colony is identified as positive for the recombination, the temperature-sensitive plasmid pCP20 expressing the flippase is incorporated. This enzyme recognizes the FRT sites flanking the kanamycin resistance gene, now on the chromosome, and flips out the segment in between them such that the kanamycin resistance is ejected. Finally, pCP20 is expelled by growth in

Figure 2-2. Deletion cassette for gene deletion in *E. coli*. The cassette consists of the kanamycin resistance gene (*kanR*), flanked by flippase recognition target (FRT) sites. The pKD4 priming sites are used to amplify the cassette, and homology sites at the ends are homologous to the gene targeted for deletion. These are added to the ends by designing primers for pKD4 with homology tails.

increased temperature, and cultures are screened for the desired gene deletion, loss of kanamycin marker, and loss of pCP20.

2.3 Methods

2.3.1 Metabolic intervention prediction by CiED

The CiED model was run by its author, Dr. Zachary Fowler, using the following flux reactions for anthocyanin production from (+)-catechin in *E. coli*: (1) Transport of (+)-catechin to and from the cell, (2) conversion of (+)-catechin to anthocyanidin by ANS, (3) glycosylation of anthocyanidin by 3GT to yield anthocyanin, and (4) transport of anthocyanin into and out of the cell.

2.3.2 Constructing knockout strains

Single deletion strains were constructed by the Datsenko and Wanner method using λ–Red recombinase according to the predictions made by the CiED model. Electrocompetent cells expressing λ–Red recombinase were made by the following procedure. A single colony was used to start a 3 mL culture of the host strain containing the helper plasmid pKD46 in LB with 100 μg/mL ampicillin. The culture was grown overnight at 30°C, 300 rpm. The following day this culture was used to inoculate 25 mL of LB containing 100 μg/mL ampicillin to OD_{600} of 0.1. The culture was grown at 30°C, 300 rpm, for 1 hour (to OD_{600} ~ 0.2). At this point, λ–Red recombinase expression was induced by addition of arabinose to 100 mM, and the culture grown to OD_{600} ~ 0.4 – 0.6. The culture was transferred to a 50 mL centrifuge tube, incubated on ice for 10 minutes, washed twice with 20 mL ice cold sterile deionized water, resuspended in 5 mL of

ice cold DI water, and aliquoted to 5 microcentrifuge tubes. The cells were pelleted once again, the supernatant discarded, and the pellets were each resuspended in 40 µL of 10% glycerol.

Freshly made competent cells were used for each deletion cassette transformation. After electroporation and reconstitution in SOC media, the transformation reaction was spread on LB plates containing 50 µg/mL kanamycin and incubated overnight at 37°C. Several colonies from the plate were picked and screened for successful recombination by colony PCR as follows. Each colony was resuspended in 15 µL of water and the cell membranes were broken by three freeze/thaw cycles (5 minutes at -20°C, followed by 1 minute at 95°C). These samples were then used as templates for PCR using GoTaq Hot Start Polymerase (Promega). For each reaction, two sets of primers were used. One set was the check primer set, designed to amplify a region of known length within the gene that was targeted for deletion. If the recombination was not successful, this segment of the gene would be amplified and detected on the agarose gel. The second set of primers was the control set, which amplifies a separate region on the chromosome. This set prevents a false positive by ensuring that the PCR ran correctly. An additional control was necessary to ensure the integrity of check primers such that they would indeed detect a negative and amplify the gene segment if recombination had not occurred. Therefore, an additional PCR was run using the same two sets of primers, but containing the chromosome of the wildtype strain as a template. A positive was identified when the check primers did not amplify the gene in the check PCR, but did amplify the gene segment in the wildtype control reaction, and the control primers amplified the control segment in both the check PCR and the wildtype control reaction.

A positive colony was used to make electrocompetent cells as described above. No selective pressure was necessary since there were no plasmids at this point. Plasmid pCP20 was transformed by electroporation, the reaction was plated on LB with 100 μg/mL ampicillin, and the plate was incubated overnight at 30°C. The following day, five colonies were each grown in 3 mL LB at 42°C, 300 rpm for 6-8 hours to expel pCP20. Each culture was then streaked on three plates: LB, LB with 100 μg/mL ampicillin, LB with 50 μg/mL kanamycin. The plates were grown overnight. Positive cultures were identified by growth on LB only. Colonies from these plates were tested by colony PCR as described above.

This recombination method was used to delete each of the following genes individually as predicted by the single-deletion results of the CiED model: *sucB*, *sucC*, *aceA*, *gdhA*, *gabD*, and *fucO*. For deletion of *sucB*, supplementation with 15 mM succinate was required for all cultures throughout the procedure. The fragment lengths designed for the check primers are listed in Table 2-1. All the primers that were used can be found in Appendix B.

2.3.3 Anthocyanin production fermentations

Anthocyanin production fermentations were conducted using BL21Star (Invitrogen) wildtype or with various single deletions as predicted by the CiED model. To begin, each strain was transformed with pCDF-3A, pET-pgm-galU, pCOLA-ndk-cmk (see Appendix A for specific gene and plasmid descriptions). A single colony was used to inoculate 3 mL LB containing 70 μg/mL ampicillin, 40 μg/mL streptomycin, and 40 μg/mL kanamycin. The culture was grown overnight at 37°C, 300 rpm. The next day, the small culture was used to inoculate 50 mL of LB with the same antibiotic selection to OD_{600} = 0.1. The culture was grown to OD_{600} ~ 0.5 and

Table 2-1. Fragment lengths designed for check PCR of deletion targets

Gene[a]	Check fragment length (bp)
sucB	1250
sucC	1500
aceA	1000
gdhA	1500
gabD	901
fucO	750

[a]*Pgi* deletion was already available

induced with IPTG to a concentration of 1 mM. The induced culture was incubated at 30°C, 300 rpm for 3 hours, after which the cells were collected by centrifugation and resuspended in 10 mL M9 media (BD), adjusted to pH 5, containing antibiotic pressure as described above, as well as 1 mM IPTG, 0.1 mM 2-oxoglutarate, 2E-4% glutamate, 2.5 mM sodium ascorbate, 0.1 mM orotic acid, and 0.75 mM (+)-catechin for substrate. For the fermentation of the wildtype comparison with Δ*aceA,* Δ*gdhA*, Δ*sucC*, the procedure varied slightly in that the pH of the M9 media was adjusted to 4 instead of 5. Also, when changing the media, instead of resuspending the pellet in 10 mL of modified M9, the culture was resuspended in the necessary volume such that the OD_{600}~15, approximately twice the usual production OD. Each culture was then allowed to produce for up to 48 hours with samples collected after 17, 24, and/or 48 hours.

2.3.4 HPLC analysis of fermentation products

For each sample, 500 µL of culture were removed, and HCl was added to 1% by volume. The sample was vortexed to mix and cells were collected by centrifugation. The supernatant was analyzed by HPLC (Agilent 1100 series with diode-array detection) on a reverse phase ZORBAX SB-C18 column (4.6 mm x 150 mm) at 25°C. The method utilized a gradient of acetonitrile and water, each containing 0.1% formic acid, at a flow rate of 1 mL/min. The acetonitrile elution profile was 10-40% for 10 minutes, 40-60% for 5 minutes, and 60-10% for 2 minutes. Anthocyanin product was identified by retention time of 5 minutes detected at absorbance at 520 nm.

2.4 Results and Discussion

The CiED model was used to generate target gene knockouts to improve production of anthocyanins in *E. coli*. The target single knockouts are bolded in the CiED results (Table 2-2). The biomass flux represents the impact of the deletion on the ability of the strain to accumulate biomass. The optimality frequency represents the fraction of times the genotype appeared as the best individual over a number of evolutionary trials. The biomass-product-coupled-yield (BPCY) is the product of the flux through ANS and the growth rate of the strain, and is an indication of how well the strain is optimized for both production and biomass accumulation. The bolded genes are the ones that were selected for deletion. The others were either considered lethal (biomass flux <0.3) or did not appear in the higher order deletion strain predictions. Since the deletions are designed to build on each other, with the greatest production improvement in the higher-order deletions, it is not worthwhile to pursue gene targets that do not appear in the higher-order model results. *SucA/B* stands for *SucA* or *SucB*, which are part of the 2-oxoglutarate dehydrogenase complex. In order to knock out the reaction, it is sufficient to delete one of these.

All of the knockout targets, with the exception of *pgi*, aim to improve the bioavailability of the ANS substrate 2-oxoglutarate (Figure 2-3). Deletion of *sucA/B* prevents the conversion of 2-oxoglutarate to succinyl-CoA in the TCA cycle. Knocking out *gdhA* eliminates the conversion of 2-oxoglutarate to glutamate. Deletion of genes *sucC*, *aceA*, and *gabD*, and *fucO* do not immediately affect 2-oxoglutarate, but they use or produce other metabolites that would affect 2-oxoglutarate levels indirectly. Therefore, these deletions reduce the drain of intracellular 2-oxoglutarate. The only targeted deletion that seems to improve UDP-glucose bioavailability is *pgi*. This knockout prevents the conversion of glucose-6-phosphate to fructose-6-phosphate, thus

Table 2-2. CiED simulation results

Strain	Biomass Flux ($h r^{-1}$)	Optimality Frequency	BPCY (mmol•gDCW $•h r^{-2}$)
wiltype	*0.94*	*n/a*	*0.00*
SucA/B	0.83	95	0.12
GpmA/B	0.30	3	0.12
YibO	0.30	2	0.12
SucC	0.84	0	0.08
CyoA	0.10	0	0.06
AceA	0.85	0	0.04
GdhA	0.65	0	0.03
GlcB	0.85	0	0.03
AceB	0.85	0	0.03
Pgi	0.50	0	0.03
GabD	-	0	-
FucO	-	0	-
SucA/B AceA	0.72	72	0.21
SucA/B FucO	0.72	16	0.21
SucA/B GabD	0.72	8	0.21
SucA/B AceA FucO	0.69	48	0.28
SucA/B AceA GabD	0.69	16	0.28
SucA/B FucO GabD	0.69	8	0.28
SucA/B AceA GdhA	0.69	8	0.28
SucA/B Frd GabD	0.69	8	0.28
SucA/B GdhA Pgi	0.69	8	0.28

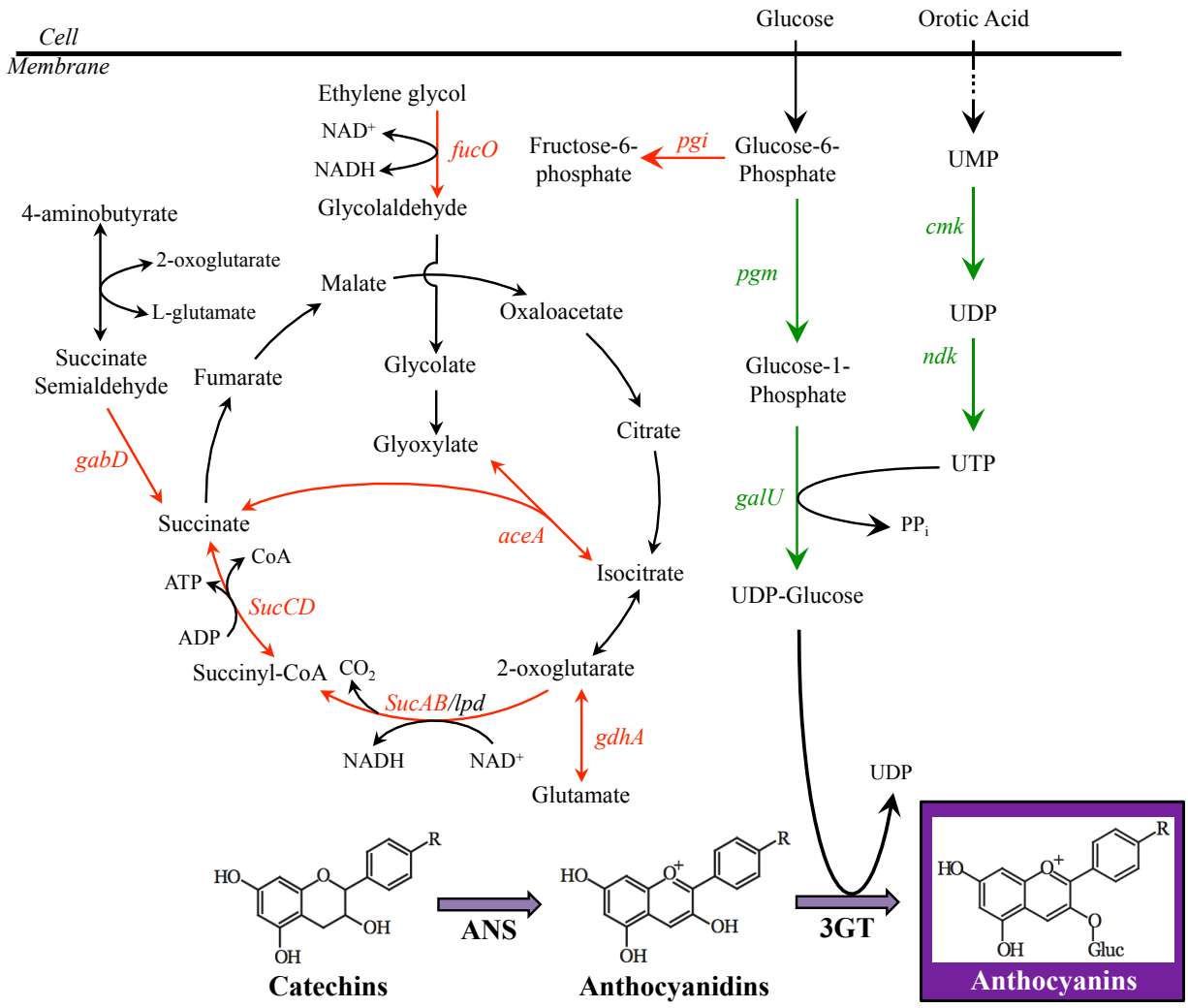


Figure 2-3. *E. coli* metabolic network affected by the CiED metabolic perturbations. Green arrows represent overexpression based on rational engineering. Red arrows represent deletion targets predicted by CiED.

allowing more conversion of glucose-6-phosphate to glucose-1-phosphate, which is subsequently converted to UDP-glucose.

The single knockout strains were constructed using the Datsenko and Wanner method. The agarose gels of the final check PCR for each deletion are shown in Figure 2-4. The *SucA/B* deletion was initially unsuccessful. This is likely due to a lethal reduction of succinate levels in the cell. Gene *sucB* is essential for the formation of the 2-oxoglutarate dehydrogenase complex, which catalyzes an irreversible conversion to succinyl-CoA, which is subsequently converted to succinate in the TCA cycle. The supposition was confirmed when the deletion was successfully constructed with supplementation of 15 mM succinate throughout the recombination procedure. The *pgi* knockout strain was available from previous work.

Once the single deletion strains were constructed, they were tested for anthocyanin production. Plasmid pCDF-3A was used for production. This plasmid contains a fusion protein of ANS and 3GT that was used previously in the lab [9]. For these fermentations, we also incorporated plasmids pET-pgm-galU and pCOLA-ndk-cmk. These overexpressions have shown significant improvement of anthocyanin production – likely due to increased carbon-flow toward the key precursor UDP-glucose (Figure 2-3). Appendix A lists further detail on these plasmids and genes.

Fermentation results are shown in Figure 2-5. Fermentations of cultures in Figure 5A were obtained using the protocol using M9 with pH=4 as described in Section 2.3.3. The results shown in Figure 5B were obtained using M9 with pH=5. Sampling was stopped after 24 hours of

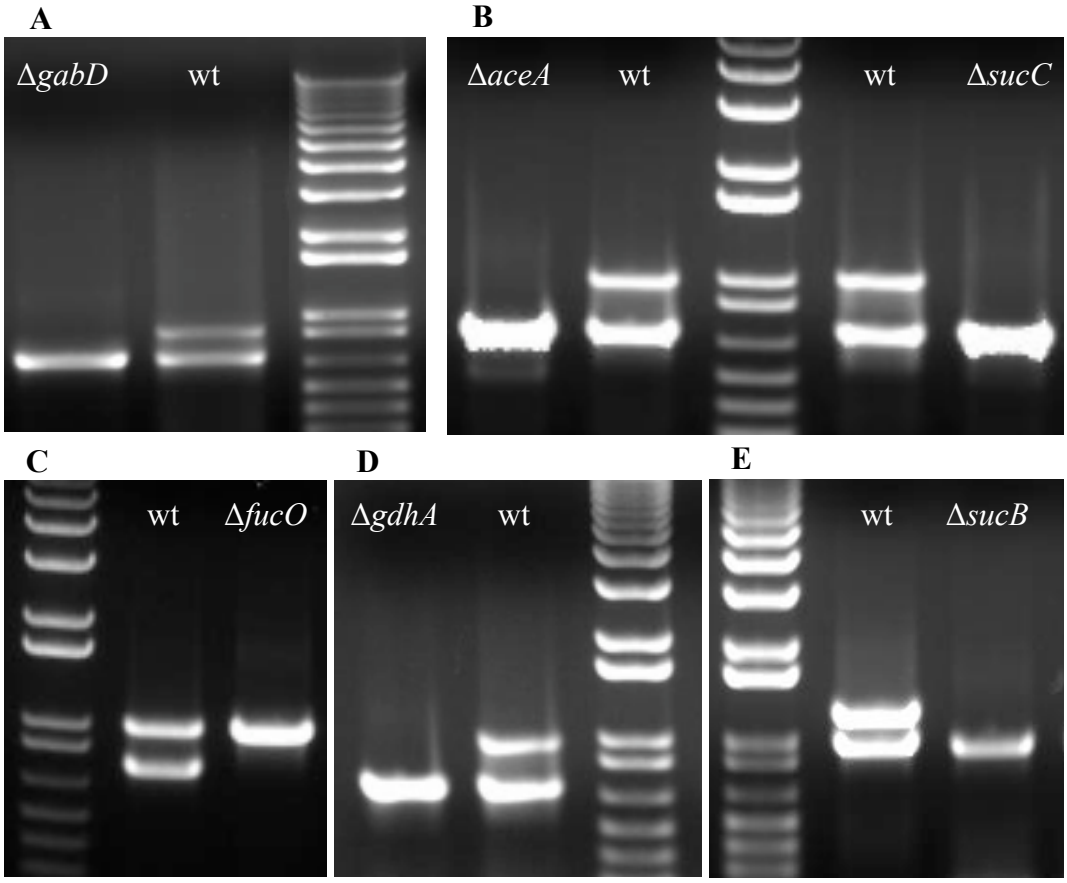
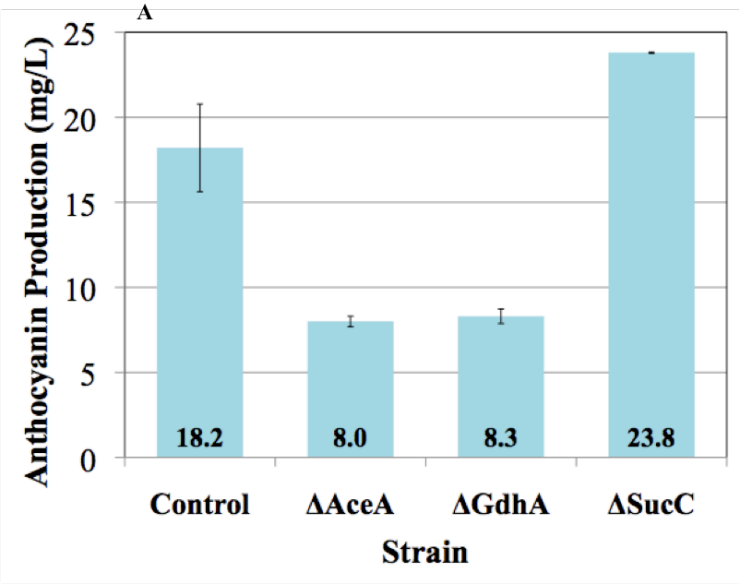


Figure 2-4. Agarose gel images showing results from the final check PCR of each CiED deletion. Both the control and deletion fragments amplified in the wildtype strains. The gene target lanes show that only the control fragment was amplified – indicating that the target gene was successfully knocked out. The ladder used was 1 kb Plus DNA Ladder (Invitrogen).



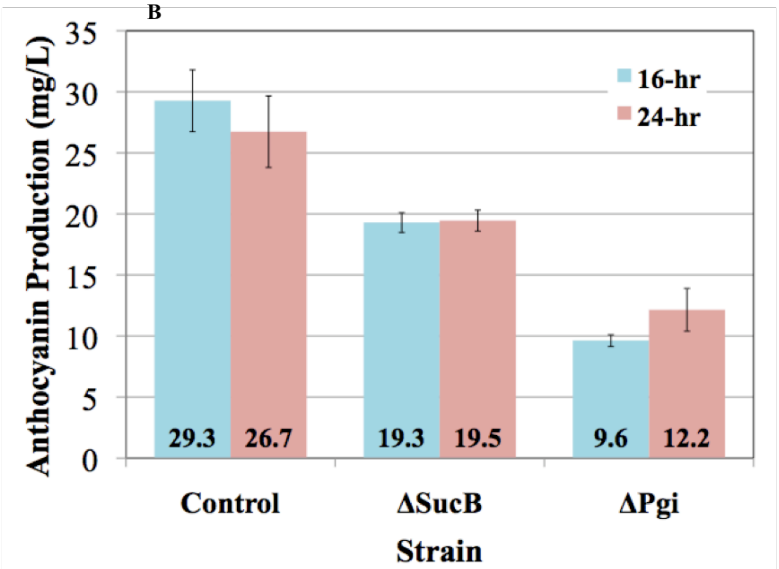


Figure 2-5. Anthocyanin production levels in BL21Star and CiED deletion strains. The control strains contain the overexpression plasmids such that only the effect of the deletion is depicted. Anthocyanin production for (A) production OD = 15, M9 pH = 4, samples collected after 48 hours; (B) biomass fermentation culture concentrated 5 times by volume, M9 pH=5, samples collected after 16 and 24 hours. Error bars represent the standard deviation of three replicates.

production since no increase was observed with increased fermentation time. Although the fermentation protocol varied slightly between the two sets of experiments, in both cases the deletion strains did not yield improved anthocyanin production when compared to the control strain. Furthermore, in most cases the production was lower than that yielded in the control.

Based on previous work using the CiED model for flavanone production [46], the single deletion strains were expected to yield the greatest improvement in production. In that work, the higher order deletion strains continued to raise production levels, but the greatest gain was observed between control strain and the single-deletion strain production levels, rather than between single and double deletion, double and triple deletion, etc. An alternate course was therefore pursued for this work.

2.5 Conclusion

The single deletion strains predicted by the CiED model failed to yield improved anthocyanin production over the control. This was unexpected because the model has predicted successful strain adjustments for other production targets. A second model was chosen to investigate whether it could produce better predictions.

Chapter 3: Applying the OptForce Model to Anthocyanin Production in *E. coli*

3.1 Introduction

In pursuit of better anthocyanin-producing *E. coli* strains, the cipher for evolutionary design (CiED) computer model was used to predict metabolic interventions that would direct carbon-flow towards anthocyanin production. Most of the deletions predicted by CiED aimed at improved bioavailability of 2-oxoglutarate, one of the substrates of ANS. However, when the single-deletion strains were constructed and tested for anthocyanin production levels, they did not yield any improvement in anthocyanin production.

The work presented in this chapter describes another attempt at this approach, using another computer model called OptForce [49]. This model predicts gene overexpressions in addition to knockouts. Therefore, in this work we aimed to find another set of metabolic intervention predictions that may predict more successful anthocyanin-producing strains.

3.2 Theory

3.2.1 OptForce procedure for strain redesign

The OptForce model was used in this work to identify metabolic interventions that may lead to improved anthocyanin production in the *E. coli* strain BL21Star.

OptForce first identifies the maximal range of flux variability in the wildtype strain by running iterations that maximize and minimize each flux within the constraints of reaction stoichiometry, uptake conditions, and any flux data available from metabolic flux analysis (MFA) for the

wildtype strain. Next, the model finds the ranges of bounds for the fluxes consistent with the production of the target compounds.

The algorithm identifies a MUST set of fluxes that must change to produce the target compound, from which a FORCE set is extracted, giving the minimal set of fluxes that must be changed through genetic manipulation in order to achieve the overproduction compound. To identify the MUST sets, the two sets of flux ranges are contrasted. The program identifies which fluxes require modification, and to what extent, for production of the desired target – in this case anthocyanins. Fluxes with perfect overlap require no modification. When comparing the wildtype with production strain flux analyses, the fluxes with no overlap must be modified to yield the target overproduction. Such modifications can be accomplished directly by affecting the particular reaction, or indirectly by adjustment of another reaction within the metabolic network of the flux in question. The difference between the non-overlapping ranges quantitatively dictates the required degree of flux modification.

Finally, the model finds the FORCE set, which is the minimal set of direct interventions needed to yield the required flux modifications. The interventions are achieved by a combination of gene knockouts, down-regulations, or up-regulations. OptForce runs an optimization formulation to minimize the number of modifications required to achieve the flux adjustments dictated by the MUST set. The results are presented as a Boolean diagram showing the possible minimal required sets of engineering modifications.

3.3 Methods

3.3.1 Metabolic intervention prediction by the OptForce procedure

The OptForce computer model was run by its author Sridhar Ranganathan from the Maranas lab at The Pennsylvania State University in order to identify metabolic interventions to yield anthocyanin overproduction in *E. coli* strain BL21Star. The non-native reactions for anthocyanin production were added to the wildtype metabolic network according to literature [41]. The biomass yield was maintained at 10% of its maximum biomass flux to support growth.

3.3.2 Deletion strains

Single gene knockouts of *pykA*, *pykF*, *gnd*, *dhaK*, *ackA*, and *pta* were constructed using the Datsenko and Wanner λ–Red recombinase method, as detailed in Section 2.2.2. The fragment lengths used for the check primers are listed in Table 3-1. The primer sequences are listed in Appendix B.

3.3.3 Anthocyanin production

Anthocyanin production fermentations were conducted as described in Section 2.3.3, with several modifications. First, the production plasmid used was pCDF-3AO, which has ANS and 3GT as an operon under one promoter rather than the fusion protein used previously. The production media M9 was set to pH=4.8, IPTG concentration was raised to 5 mM, and the (+)-catechin substrate concentration was raised to 2 mM. During production, the cultures were pulsed with glucose to 1% by volume and with (+)-catechin substrate to 1 mM at 17 hours. Samples were collected at 48 hours and analyzed by HPLC as described earlier.

Table 3-1. Fragment lengths designed for check PCR of deletion targets

Gene	Check fragment length (bp)
pykA	750
pykF	900
gnd	1000
dhaK	860
ackA	1100
pta	1250

3.4 Results and Discussion

The metabolic interventions recommended by OptForce are shown in Table 3-2. The impact of the deletions and overexpressions on the *E. coli* metabolic network are depicted in Figure 3-1. In general, the interventions lead to increased carbon-flow towards UDP-glucose which is mainly produced in the first two steps of glycolysis. The *gnd* knockout shuts down the competing pentose phosphate pathway. The *dhaK* and *pyk* deletions reduce carbon-flow from glycolysis to pyruvate metabolism. It should be noted that in order to knock out *pyk*, both isozymes *pykA* and *pykF* must be knocked out. Gene knockouts *ackA* and *pta* should lead to decreased formation of acetate, which is toxic to the cells.

The single-knockout strains were constructed using the Datsenko and Wanner method. The agarose gels of the final check PCR show the successful deletion of each target gene Figure 3-2. The two bands in the wildtype strain are amplified by the control primer set and the check primer set. The single band in the deletion strain lanes were amplified by the control primers used to determine whether the PCR was successful. The second strain of primers was also included in the reaction, but since they did not amplify the fragment within the target gene, it is clear that the gene deletion was successful. Each strain is also streaked on LB with kanamycin to ensure that the kanamycin marker was excised out as well.

Once the single deletions were constructed, anthocyanin production was measured for each strain. Two sets of fermentations were run. The anthocyanin production fermentations were run using two different sets of overexpression plasmids. First, the overexpressions recommended by the OptForce procedure were used (Figure 3-3A). These include pET-pgm-galU and pACYC-

Table 3-2. OptForce simulation results

Number of Metabolic Interventions	Reaction Knockouts	Reaction Up-Regulations	Minimum Guarenteed Flux for Anthocyanin ($mmol \bullet gDW^{-1}hr^{-1}$)
K=5	*gnd, pyk*	*galU, pgm, ycjU*	4.69 (39.4%)[a]
K=7	*gnd, pyk, dhaK, ackA* (or) *pta*	*galU, pgm, ycjU*	8.74 (73.4%)[a]

[a]Percent theoretical maximum

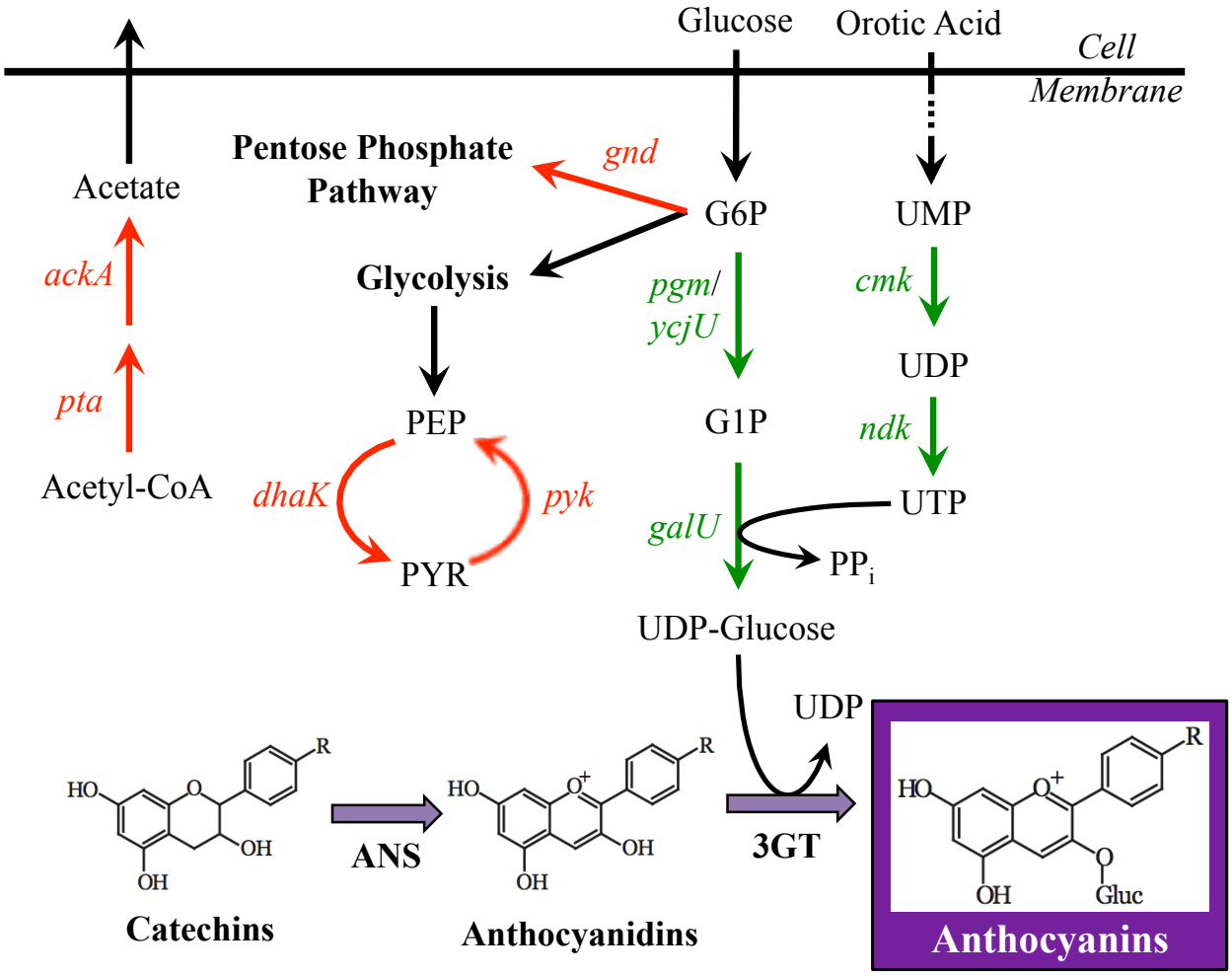


Figure 3-1. *E. coli* metabolic network affected by the OptForce metabolic perturbations. Green arrows represent overexpression based on rational engineering or OptForce. Red arrows represent deletion targets predicted by OptForce. Orotic acid is supplied in the media and goes through several modifications that ultimately result in UTP.

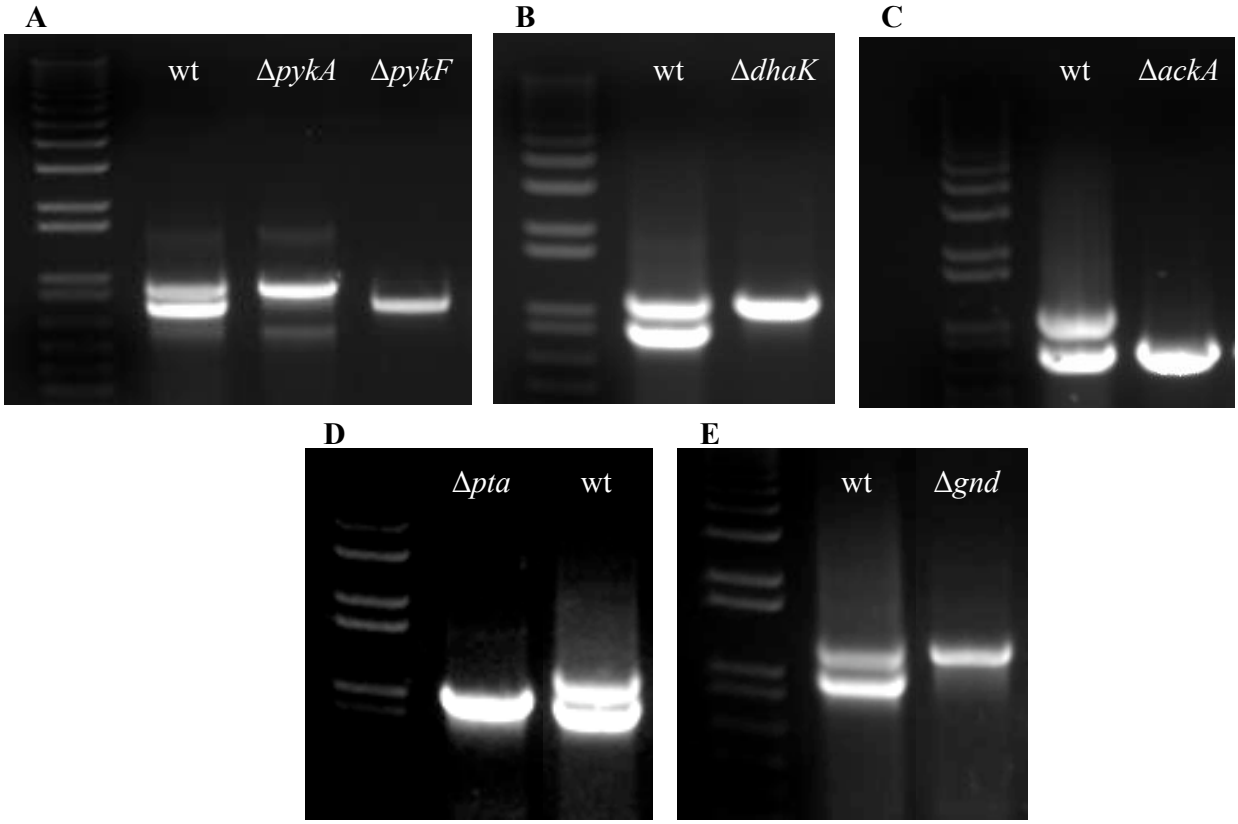


Figure 3-2. Agarose gel images showing OptForce gene deletions. Both the control and deletion fragments amplified in the wildtype strains. The gene target lanes show that only the control fragment was amplified, indicating that the target gene was successfully knocked out. Ladder used was 1 kb Plus DNA Ladder (Invitrogen).

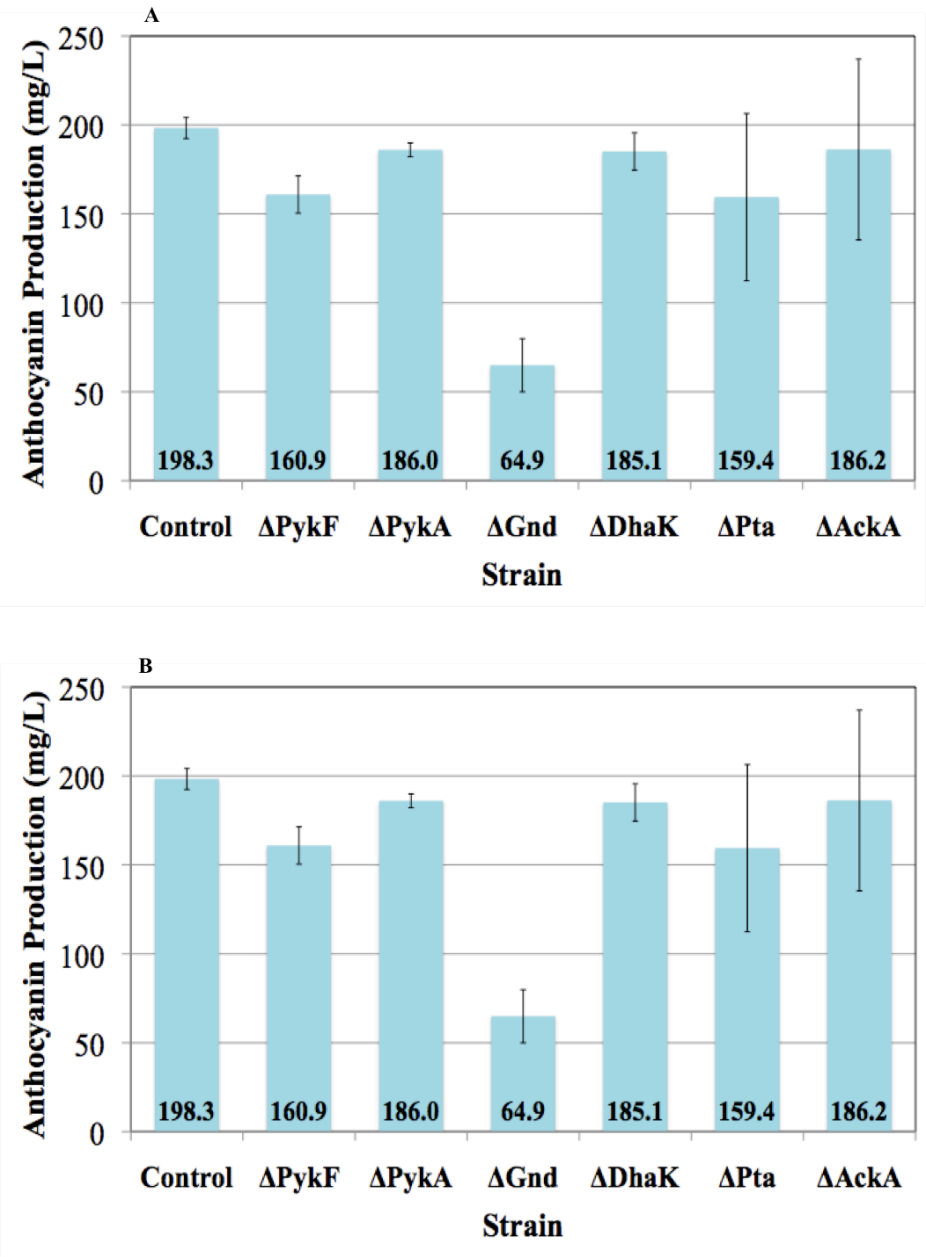


Figure 3-3. Anthocyanin production levels in BL21Star and OptForce deletion strains. The control strain contains the overexpression plasmids such that the effect of the deletion is isolated. The data shown in (A) was obtained using the overexpressions recommended by the OptForce model: *ycjU*, *pgm*, *galU*. The data shown in (B) was obtained using the overexpressions that were found to be best: *ycjU*, *galU*, *ndk*, *cmk*. Samples were collected after 48 hours of production. Error bars represent the standard deviation of three replicates.

ycjU. The production fermentations were also attempted using the combination of overexpressions that was found in other work (unpublished data) to give the best anthocyanin production (Figure 3-3B). These include pACYC-ycjU-galU and pCOLA-ndk-cmk. For both runs, plasmid pCDF-3AO was used for the heterologous expression of ANS and 3GT. Unlike the fusion protein used for the previous work with the CiED model, this plasmid contains the two proteins as an operon under a single promoter. This construct was used since it contributed to significant improvement in production.

The single-deletion strains with overexpressions predicted by the OptForce procedure did not significantly improve anthocyanin production over the control strain (Figure 3-3A). When the optimized overexpression set of plasmids was used, *pta* and *ackA* deletion strains each improved anthocyanin production significantly (Figure 3-3B). This is likely due to reduced levels of acetate, which may lead to increased biomass accumulation. None of the other single-deletion strains, aimed at improving UDP-glucose bioavailability, improved production.

3.5 Conclusion

The single-deletion strains aimed at directing carbon-flow towards UDP-glucose failed to show significant improvement in anthocyanin production over the control strain. However, an increase was observed by decreasing flux towards acetate in strains containing the optimal combination of overexpression plasmids. The increase is likely due to improved biomass maintenance with decreased levels of acetate. The results from the deletion strains that were expected to lead to an increase in UDP-glucose bioavailability were particularly surprising since UDP-glucose is thought to be a key precursor to anthocyanin production.

Since increased bioavailability of the ANS substrate 2-oxoglutarate (predicted by the CiED model) and of the 3GT substrate UDP-glucose (predicted by the OptForce model) did not improve anthocyanin production in *E. coli*, it was concluded that precursor bioavailability may not be the primary limitation. We therefore decided to investigate the expression of the heterologous proteins in *E. coli*.

Chapter 4: Investigation of Anthocyanin Pathway Protein Expression

4.1 Introduction

In an attempt to improve the anthocyanin production platform of *E. coli* carrying the heterologous genes for anthocyanidin synthase from *Petunia hybrida* (PhANS) and UDP-glucose:flavonoid 3-*O*-glucosyltransferase from *Arabidopsis thaliana* (At3GT), two computer models were used to predict metabolic interventions that may re-direct carbon-flux towards anthocyanin production. Gene knockout predictions from both CiED and OptForce computer simulations were utilized to construct single-deletion *E. coli* strains. However, the engineered single-deletion strains produced similar to or lower anthocyanin levels than the control strain.

Each of the two models aimed to improve the bioavailability of a different important precursor for anthocyanin production. CiED focused on improved 2-oxoglutarate levels, and OptForce predicted interventions that should have improved UDP-glucose levels. Since neither of these approaches was successful, there may be another factor limiting production other than precursor availability. The work presented in this chapter aims to investigate the levels and quality of heterologous protein expression of ANS and 3GT in *E. coli*, since poor expression may lead to poor activity, and perhaps the ultimate limitation to improved anthocyanin production.

4.2 Methods

The Champion pET SUMO Expression System (Invitrogen) was used for expression and purification of PhANS and At3GT. Briefly, this kit provides linearized pET SUMO vectors for TA cloning of the gene of interest for soluble expression in *E. coli*. Soluble expression is

achieved by creating an N-terminal fusion of the gene of interest with the small ubiquitin-like modifier (SUMO) protein. The vector includes a HIS-tag on the N-terminus of the SUMO protein for purification on a nickel column, and a cleavage site that is recognizable by the SUMO protease such that pure native protein can be obtained. After cleavage, the mixture is applied to the nickel column again such that the HIS-tagged SUMO protein binds and the pure native protein elutes in the flow-through fraction.

4.2.1 Cloning ANS and 3GT into pET-SUMO expression system

Primers for the amplification of PhANS and At3GT for TA cloning into the pET-SUMO vector were designed with the following considerations – as recommended by the SUMO Expression System manual: First, the ATG codon was included in the forward primer of each gene for later production of the native protein; second, since cleavage by SUMO protease is most efficient if the first amino acid is serine, the AGC codon for serine was included immediately preceding the start codon; finally, the stop codon of the gene was included in the reverse primer. The primer sequences are listed in Appendix B.

The genes were amplified by PCR with PlatinumTaq DNA Polymerase (Invitrogen) using as templates plasmids pMAL-PhANS and pMAL-At3GT – available from previous work. The polymerase was chosen for its high-fidelity activity and because it leaves A-overhangs, such that the PCR products can be used directly for TA cloning. The sizes of the PCR products were confirmed by agarose gel electrophoresis. The PCR gene product was used directly for the ligation reaction with linearized pET-SUMO vector using T4 DNA ligase (New England Biolabs). The reaction was incubated for 30 minutes at room temperature. To transform Mach1-

T1 chemically competent *E. coli*, 2 µL of each ligation reaction were added to 50 µL shots of chemically competent cells. The reaction was incubated on ice for 15 minutes, then heat-shocked for 30 seconds in a 42°C water bath, and finally reconstituted in 250 µL SOC media per reaction. The cultures were incubated for 45 minutes at 37°C and 225 rpm, and spread on LB plates containing 50 µg/mL kanamycin.

The resulting colonies were screened for proper cloning. Ten colonies were picked from each plate, resuspended in 3 mL LB with 50 µg/mL kanamycin, and grown overnight at 37°C, 300 rpm. The next day, plasmids were isolated from each culture using the Zyppy Plasmid Miniprep Kit (Zymo Research). Each plasmid sample was screened by restriction digest to ensure insertion and correct orientation. Any positive samples from this initial screen were sent for sequencing to ensure that the integrity of the gene sequence was preserved. The sequence of PhANS was confirmed immediately. The sample for SUMO-At3GT contained one mutation. This was corrected by point mutagenesis using the Strategene QuikChange Site Directed Mutagenesis Kit (Strategene). The primers used for the mutagenesis are listed in Appendix B.

Once the plasmids were constructed and shown to be correct, they were stored and maintained in Mach1-T1 cells (Invitrogen).

4.2.2 SUMO fusion protein expression

Expression of SUMO-PhANS and SUMO-At3GT was conducted in *E. coli* strain BL21(DE3) (Invitrogen) – as recommended by the SUMO Kit manual, except where otherwise noted. For each expression experiment, a fresh transformation of the SUMO fusion plasmid was performed.

For each transformation, 2 µL of the plasmid were transformed into the chemically competent strain, but instead of spreading the reconstituted culture on a selective plate, it was added to 10 mL of LB + 50 µg/mL kanamycin and grown overnight at 37°C, 300 rpm. The next day, this preinoculum was used to inoculate a larger culture (as noted for each specific experiment) of LB + 50 µg/mL kanamycin to an OD_{600} of 0.1. The culture was grown at 37°C, 300 rpm to OD_{600} ~ 0.4-0.6, then induced with 1 mM IPTG (except where otherwise noted), and grown at 30°C, 300 rpm. For all SUMO-At3GT cultures, the media was supplemented with 1% glucose to promote growth for increased protein expression. Samples were collected initially and at different stages of the experiment – as indicated for the individual experiments.

For each sample collected for the expression studies, 0.5-2 mL of the culture was transferred to a microcentrifuge tube on ice and the cells were collected by microfugation (1 minute, 14,000 rpm). The supernatant was discarded and the tube returned to ice. The pellet was then washed with ice cold 1X PBS and microfuged again. The supernatant was discarded and the final pellet was stored at -80°C for analysis.

4.2.3 SDS-PAGE and western blot

To prepare samples for SDS-PAGE, the proteins were extracted from the samples and separated into soluble and insoluble fractions. Each sample was thawed on ice and resuspended in 500 µL of ice cold SUMO lysis buffer (made according to kit manual). Lysozyme was added to a concentration of 10 µg/mL. The samples were then frozen at -80°C and thawed at 42°C for a total of 3 freeze/thaw cycles. Finally, the samples were microfuged at 14000 rpm, for 1 minute at 4°C to pellet the insoluble proteins. The supernatant containing the soluble proteins was

transferred to a clean tube. All samples were stored on ice. To prepare soluble samples for loading, the samples were each mixed 1:1 with 2X SDS-PAGE loading buffer and boiled for 5 minutes. The insoluble pellets were resuspended in 500 µL of 1X SDS-PAGE loading buffer and boiled for 5 minutes. Except where otherwise indicated, loading was normalized by the OD_{600} at the time that each sample was taken.

To run the western blots, proteins were transferred from the gel to a nitrocellulose membrane (Whatman) for 1 hour at 4°C. The membrane was blocked for 3 hours with 5% milk in TBST at room temperature, and then overnight with 2% milk in TBST containing 0.2 µg/mL goat anti-6-His antibody, HRP-conjugated (Bethyl Laboratories, Inc.) at 4°C on a platform shaker. Western blots were developed using LumiGlo chemiluminescence (Cell Signaling Technology) and Kodak intensifying screen, fixer solution, and developer solution.

4.3 Results and Discussion

The gels and western blot films of the expression experiment indicate that induced expression of SUMO-ANS and SUMO-3GT was successful (Figure 4-1). In the SDS-PAGE, the band of the SUMO fusion protein is absent in the uninduced samples, but becomes increasingly darker as time goes on in the induced culture. The western blots confirm the effectiveness of the HIS-tag and that the bands observed in the SDS-PAGE gels are the SUMO proteins. It is also clear from these images that SUMO-ANS is expressed at a much higher level than SUMO-3GT is. For this study, protein loading was not normalized, but rather each lane represents extract from a 1-mL culture sample. Even though the gel loading was not normalized, there is such a marked difference in the intensity of the bands on the SUMO-ANS western blot film versus the SUMO-

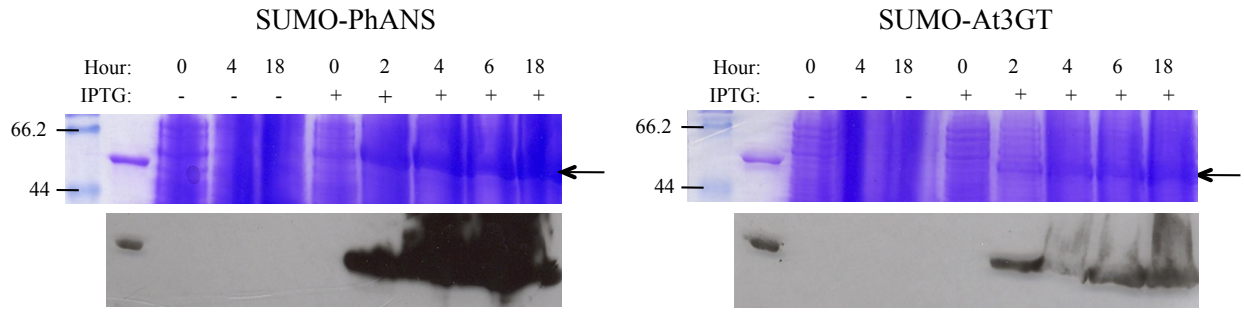


Figure 4-1. Comparison of SUMO fusion protein expression in BL21(DE3) from induced and uninduced cultures by SDS-PAGE and western blot. Samples were taken at 0, 2, 4, 6, and 18 hours. The arrows point to the SUMO fusion proteins. The ladder used was SDS-PAGE Molecular Weight Standards, Broad Range (Biorad). The lane directly to the right of the ladder contains a control HIS-tagged protein for the western blot analysis. The arrow on the right of each SDS-PAGE points to the SUMO fusion protein (61.7 kDa for SUMO-ANS and 63.6 for SUMO-3GT).

3GT film that it is evident that SUMO-3GT does not express nearly as well as SUMO-ANS. This does not seem to be due to differing exposure times for the western blot film, as indicated by the similar intensity of the control HIS-tagged protein (~65kDa) in the lane next to the ladder.

Separation of expression samples into soluble and insoluble fractions indicated that SUMO-ANS is expressed primarily in the soluble fraction, while SUMO-3GT is almost entirely in the insoluble fraction (Figure 4-2). This experiment was conducted in BL21Star, and the cultures were induced at OD_{600}~0.6. For these gels, the protein loading was normalized by OD_{600} at the time the sample was taken. The same experiment was conducted with inductions at OD_{600}~1 and OD_{600}~1.5, but expression was negligible (data not shown).

In order to test whether soluble expression of SUMO-3GT would be improved in a different host strain, the solubility experiment was repeated in three different strains: BL21(DE3), JM109, and XL1Blue. We also tested soluble expression levels with reduced concentrations of IPTG inducer since it is believed that slowing down protein folding may lead to higher levels of soluble protein. IPTG concentrations of 0.5 mM and 0.1 mM were tested, where previously 1 mM IPTG had been used. Sample loading was normalized by OD_{600} at the time the sample was taken. None of the strains or induction concentrations gave significant levels of soluble protein (Figure 4-3). BL21(DE3) expresses the best, but nearly all the protein is still in the insoluble fraction. The reduced IPTG concentration of 0.5 mM improved solubility.

Finally, we investigated expression of SUMO-3GT in strain BL21(DE3)pLysS, recommended by Invitrogen for reduced leaky expression. We compared expression when the induction was run at

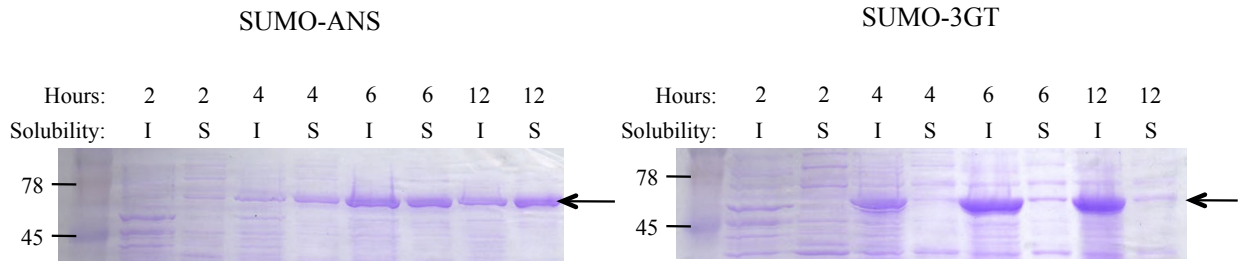


Figure 4-2. Soluble versus insoluble expression of SUMO fusion proteins in BL21Star(DE3). Samples were taken at 2, 4, 6, and 12 hours. Ladder used was Kaleidoscope Prestained Standards, broad range (BioRad). For each time point, the protein extract was divided into insoluble (I) and soluble (S) fraction. The arrows point to the SUMO fusion proteins.

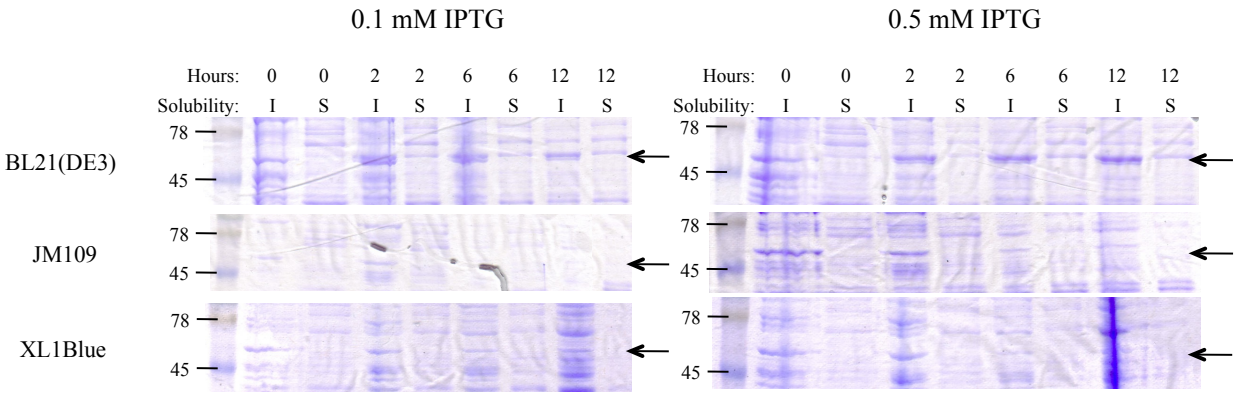


Figure 4-3. Soluble versus insoluble expression of SUMO-3GT in varying strains and induction levels. Expression was tested in strains BL21Star, JM109, and XL1Blue with induction at 0.1 mM and 0.5 mM IPTG. Samples were taken at 0, 2, 6, and 12 hours. Ladder used was Kaleidoscope Prestained Standards, broad range (BioRad). For each time point, the protein extract was divided into insoluble (I) and soluble (S) fractions. The arrows point to SUMO-3GT.

room temperature versus 16°C. Reducing the temperature increased the fraction of soluble protein, however, the insolubility persisted (Figure 4-4).

4.4 Conclusions and Future Work

We concluded from this study that the 3GT expressed in *E. coli* is mostly insoluble, except when expressed at the reduced temperature of 16°C. This may be the primary bottleneck for improved anthocyanin production.

While the solubility challenge can be manageable when the end application is protein purification (by simply increasing the culture volume), temperature reduction is not an implementable solution for anthocyanin production. Protein engineering of 3GT to improve solubility should be explored next. It is believed that improved solubility will improve activity, and therefore production. There are several methods that have worked well for other problematic proteins, but we recommend starting with cysteine scanning to alanine, as has elsewhere been shown to successfully improve protein solubility and ultimately production [52]. When a more soluble form of 3GT is engineered, it would be interesting to repeat the production studies using the knockout strains predicted by the computer models.

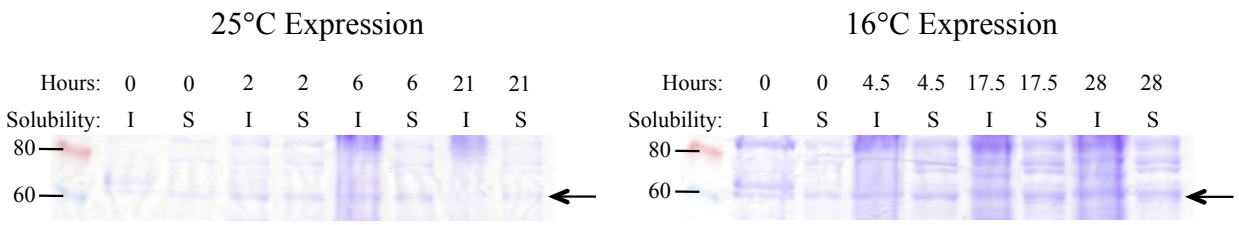


Figure 4-4. Soluble versus insoluble expression of SUMO-3GT in strains BL21Star pLysS with expression at room temperature (25°C) versus 16°C. Cultures were induced with 0.5 mM IPTG. Ladder used was ColorPlus Prestained Protein Ladder, Broad Range (NEB). For each time point, the protein extract was divided into insoluble (I) and soluble (S) fractions. The arrows point to SUMO-3GT.

APPENDIX A: Genes and Plasmids

Table A1. Genes used and mentioned in this work

Gene	Protein
pgm	Phosphoglucomutase
galU	Glucose 1-phosphate uridylyltransferase
cmk	Deoxycytidylate kinase
ndk	Nucleoside diphosphate kinase
sucB	2-oxoglutarate dehydrogenase complex
sucC	Succinyl-CoA synthetase
aceA	Isocitrate lyase
gdhA	Glutamate dehydrogenase
gabD	Succinate semialdehyde dehydrogenase
fucO	L-1,2-propanediol oxidoreductase
pgi	Phosphoglucose isomerase
pykA	Pyruvate kinase
pykF	Pyruvate kinase
gnd	6-phosphogluconate dehydrogenase
dhaK	Dihydroxyacetone kinase
ackA	Acetate kinase
pta	Phosphate acetyltransferase
ycjU	β-phosphoglucomutase

Table A2. Plasmids used and mentioned in this work

Plasmid	Relevant characteristics	Source
pETDuet-1	Double T7 promoters, ColE1 ori, AmpR	Novagen
pCOLADuet-1	Double T7 promoters, CloDF13 ori, SmR	Novagen
pCDFDuet-1	Double T7 promoters, ColA ori, KanR	Novagen
pACYCDuet-1	Double T7 promoters, P15A ori, CamR	Novagen
pCDF-3A	pCDFDuet-1 carrying Arabidopsis *3GT* fused to N-terminus of Petunia *ANS*	Yan et al. (2008)
pET-pgm-galU	pETDuet-1 carrying *E. coli galU* and *pgm*	Yan et al. (2008)
pCOLA-ndk-cmk	pCOLADuet-1 carrying *E. coli ndk* and *cmk*	Yan et al. (2008)
pCDF-3AO	pCDFDUET-1 carrying Arabidopsis *3GT* and Petunia *ANS* under a single promoter as an operon	This study
pACYC-ycjU	pACYCDuet-1 carrying *E. coli ycjU*	This study
pACYC-ycjU-galU	pACYCDuet-1 carrying *E. coli ycjU and galU*	This study
Champion pET SUMO	T7 promoter, pBR322 ori, KanR,	Invitrogen
pET-SUMO-PhANS	Champion pET SUMO carrying Petunia *ANS*	This study
pET-SUMO-At3GT	Champion pET SUMO carrying Arabidopsis *3GT*	This study

APPENDIX B: Primers

Table B1. Primers used for work in Chapter 2

Primer	Sequence
sucB homology primer - forward	GCTGAACGTCGAATAAATAAAGGATACACAATGAG TAGCGTAGATATTCTTGTAGGCTGGAGCTGCTTCG
sucB homology primer - reverse	CGGTCTACAGTGCAGGTGAAACTTAAACTACTACAC GTCCAGCAGCAGACATGGGAATTAGCCATGGTCC
sucC homology primer - forward	ATGAACTTACATGAATATCAGGCAAAACAACTTTTT GCCCGTGTAGGCTGGAGCTGCTTC
sucC homology primer - reverse	TTATTTCCCCTCCACTGCGGCAACAACCTGCTGAGC TGCAATGGGAATTAGCCATGGTCC
aceA homology primer - forward	ATGAAAACCCGTACACAACAAATTGAAGAATTACA GAAAGGTGTAGGCTGGAGCTGCTTC
aceA homology primer - reverse	TTAGAACTGCGATTCTTCAGTGGAGCCGGTCAGCGC GGTGATGGGAATTAGCCATGGTCC
gdhA homology primer - forward	ATGGATCAGACATATTCTCTGGAGTCATTCCTCAAC CATGGTGTAGGCTGGAGCTGCTTC
gdhA homology primer - reverse	TTAAATCACACCCTGCGCCAGCATCGCATCGGCAAC CTTCATGGGAATTAGCCATGGTCC
gabD homology primer - forward	ATGAAACTTAACGACAGTAACTTATTCCGCCAGCAG GCGTGTGTAGGCTGGAGCTGCTTC
gabD homology primer - reverse	TGTTCATTCGCATTCTCCAGTTAAAGACCGATGCAC ATATATGGGAATTAGCCATGGTCC
fucO homology primer - forward	ATGATGGCTAACAGAATGATTCTGAACGAAACGGC ATGGTGTGTAGGCTGGAGCTGCTTC
fucO homology primer - reverse	TTACCAGGCGGTATGGTAAAGCTCTACAATATCCTC AAGCATGGGAATTAGCCATGGTCC
sucB check primer - forward	GGATACACAATGAGTAGCGTAGATA
sucB check primer - reverse	CAGTGCAGGTGAAACTTAAACTAC
sucC check primer - forward	CCGCTTCAAAAATCGGTGCC
sucC check primer - reverse	GCTGCAATAATATTCAGGCCGCT
aceA check primer - forward	TTACGCGGTTCAGTCAATCCTG
aceA check primer - reverse	CGTTTGCCAGGTCAAACATGTT
gdhA check primer - forward	GTAATGACCACACTCTGGCCTTTTC
gdhA check primer - reverse	TGCTGGAACAGTTCAGTCGCTT
gabD check primer - forward	CACCATTCTGCGCAACTGGTTC
gabD check primer - reverse	CAGAATGGTCGGCTGGAAGAAGT
fucO check primer - forward	CAGGATACTTGTAAAGCGATTGGC
fucO check primer - reverse	AATGTCTTCCTTGCGTACACCA

Table B2. Primers used for work in Chapter 3

Primer	Sequence
pykA homology primer - forward	ACCTGGTTGTTTCAGTCAACGGAGTATTACATGTCC AGAAGGCTTCGCAGTGTAGGCTGGAGCTGCTTCG
pykA homology primer - reverse	GTGGCGTTTTCGCCGCATCCGGCAACGTACTTACTC TACCGTTAAAATACATGGGAATTAGCCATGGTCC
pykF homology primer - forward	ATCCTTCTCAACTTAAAGACTAAGACTGTCATGAAA AAGACCAAAATTGTGTGTAGGCTGGAGCTGCTTC
pykF homology primer - reverse	CGATATACAAATTAATTCACAAAAGCAATATTACAG GACGTGAACAGATGCATGGGAATTAGCCATGGTC
gnd homology primer - forward	CGCGGTGATCACACCTGACAGGAGTATGTAATGTCC AAGCAACAGATCGGGTGTAGGCTGGAGCTGCTTC
gnd homology primer - reverse	GGCCTCAATTTTATTGTTGGTTAAATCAGATTAATCC AGCCATTCGGTATATGGGAATTAGCCATGGTCC
dhaK homology primer - forward	ATGAAAAAATTGATCAATGATGTGCAAGACGTACTG GACGTGTAGGCTGGAGCTGCTTCG
dhaK homology primer - reverse	TTATTTACCCCAGTTAAGGGCCGGGGTGTGGACCGG GGCGCATGGGAATTAGCCATGGTC
ackA homology primer - forward	ATGTCGAGTAAGTTAGTACTGGTTCTGAACTGCGGT AGTTGTGTAGGCTGGAGCTGCTTC
ackA homology primer - reverse	TCAGGCAGTCAGGCGGCTCGCGTCTTGCGCGATAAC CAGTACATGGGAATTAGCCATGGT
pta homology primer - forward	ATGTCCCGTATTATTATGCTGATCCCTACCGGAACC AGCGTGTAGGCTGGAGCTGCTTCG
pta homology primer - reverse	TTACTGCTGCTGTGCAGACTGAATCGCAGTCAGCGC GATGACATGGGAATTAGCCATGGT
pykA check primer - forward	GGCCCAAAATCCGTGTATCCACCTT
pykA check primer - reverse	CGGCAGTTTCTGCAGACAGCATCAC
pykF check primer - forward	TGGTTGACGATGGTCTGATCGGTAT
pykF check primer - reverse	TGATCTCTTTAACAAGCTGCGGCAC
gnd check primer - forward	TGCAGGCACGGATGCTGCTATTGATT
gnd check primer - reverse	ATAAGCAACGACATCACGCAGCGCC
dhaK check primer - forward	CCTGTTGCAGGAAAAGTCGC
dhaK check primer - reverse	CCGGTCATATCCAGTGAGGT
ackA check primer - forward	GCCATCATCGATGCAGTAAATGG
ackA check primer - reverse	GTTGGGATAACCACCGCAGG
pta check primer - forward	TGAAGGTCTGGTCCCGACAC
pta check primer - reverse	GACGGATGGTGTTTGCGGTA

Table B3. Primers used for work in Chapter 4

Primer	Sequence
PhANS TA cloning to pET SUMO - forward	AGCATGGTGAATGCAGTAGTTACAACTCC
PhANS TA cloning to pET SUMO - reverse	CTATTTAGATTCTTCAGCAGCAACATCCTG
At3GT TA cloning to pET SUMO - forward	AGCATGACCAAACCCTCCGACCCAACCAGAGACT
At3GT TA cloning to pET SUMO - reverse	TCAAATAATGTTTACAACTGCATCCAACAATCCTCTGAAA
At3GT point mutation primer - forward	CACCAGAAGGAGTTGTGTTTGGGAATTTAGACTCTG
At3GT point mutation primer - reverse	CAGAGTCTAAATTCCCAAACACAACTCCTTCTGGTG

References

1. Newman, D.J. and G.M. Cragg, *Natural products as sources of new drugs over the last 25 years.* Journal of Natural Products, 2007. **70**(3): p. 461-477.

2. Butler, M.S., *The role of natural product chemistry in drug discovery.* Journal of Natural Products, 2004. **67**(12): p. 2141-2153.

3. Cowan, M.M., *Plant products as antimicrobial agents.* Clinical Microbiology Reviews, 1999. **12**(4): p. 564-+.

4. Peluso, M.R., *Flavonoids attenuate cardiovascular disease, inhibit phosphodiesterase, and modulate lipid homeostasis in adipose tissue and liver.* Experimental Biology and Medicine, 2006. **231**(8): p. 1287-1299.

5. Cushnie, T.P.T. and A.J. Lamb, *Antimicrobial activity of flavonoids.* International Journal of Antimicrobial Agents, 2005. **26**(5): p. 343-356.

6. Benavente-Garcia, O. and J. Castillo, *Update on uses and properties of Citrus flavonolds: New findings in anticancer, cardiovascular, and anti-inflammatory activity.* Journal of Agricultural and Food Chemistry, 2008. **56**(15): p. 6185-6205.

7. Winkel-Shirley, B., *Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology.* Plant Physiology, 2001. **126**(2): p. 485-493.

8. Crozier, A., I.B. Jaganath, and M.N. Clifford, *Dietary phenolics: chemistry, bioavailability and effects on health.* Natural Product Reports, 2009. **26**(8): p. 1001-1043.

9. Yan, Y.J., Z. Li, and M.A.G. Koffas, *High-yield anthocyanin biosynthesis in engineered Escherichia coli.* Biotechnology and Bioengineering, 2008. **100**(1): p. 126-140.

10. Manach, C. and J.L. Donovan, *Pharmacokinetics and metabolism of dietary flavonoids in humans.* Free Radical Research, 2004. **38**(8): p. 771-785.

11. Havsteen, B.H., *The biochemistry and medical significance of the flavonoids.* Pharmacology & Therapeutics, 2002. **96**(2-3): p. 67-202.

12. Yao, L.H., et al., *Flavonoids in food and their health benefits.* Plant Foods for Human Nutrition, 2004. **59**(3): p. 113-122.

13. Wrolstad, R.E., R.W. Durst, and J. Lee, *Tracking color and pigment changes in anthocyanin products.* Trends in Food Science & Technology, 2005. **16**(9): p. 423-428.

14. Castaneda-Ovando, A., et al., *Chemical studies of anthocyanins: A review.* Food Chemistry, 2009. **113**(4): p. 859-871.

15. Schab, D.W. and N.H.T. Trinh, *Do artificial food colors promote hyperactivity in children with hyperactive syndromes? A meta-analysis of double-blind placebo-controlled trials.* Journal of Developmental and Behavioral Pediatrics, 2004. **25**(6): p. 423-434.

16. Espin, J.C., M.T. Garcia-Conesa, and F.A. Tomas-Barberan, *Nutraceuticals: facts and fiction.* Phytochemistry, 2007. **68**(22-24): p. 2986-3008.

17. Kong, J.-M., et al., *Analysis and biological activities of anthocyanins.* Phytochemistry, 2003. **64**(5): p. 923-33.

18. Wang, L.-S. and G.D. Stoner, *Anthocyanins and their role in cancer prevention.* Cancer Lett, 2008. **269**(2): p. 281-90.

19. Wang, L.S. and G.D. Stoner, *Anthocyanins and their role in cancer prevention.* Cancer Letters, 2008. **269**(2): p. 281-290.

20. Dixon, R.A. and C.L. Steele, *Flavonoids and isoflavonoids - a gold mine for metabolic engineering.* Trends in Plant Science, 1999. **4**(10): p. 394-400.

21. Winkel-Shirley, B., *Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways.* Physiologia Plantarum, 1999. **107**(1): p. 142-149.

22. Leonard, E., et al., *Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli.* Applied and Environmental Microbiology, 2007. **73**(12): p. 3877-3886.

23. Kong, J.M., et al., *Analysis and biological activities of anthocyanins.* Phytochemistry, 2003. **64**(5): p. 923-933.

24. Prior, R.L. and X.L. Wu, *Anthocyanins: Structural characteristics that result in unique metabolic patterns and biological activities.* Free Radical Research, 2006. **40**(10): p. 1014-1028.

25. Cooke, D., et al., *Effect of cyanidin-3-glucoside and an anthocyanin mixture from bilberry on adenoma development in the Apc(Min) mouse model of intestinal carcinogenesis - Relationship with tissue anthocyanin levels.* International Journal of Cancer, 2006. **119**(9): p. 2213-2220.

26. Stoner, G.D., et al., *Protection against esophageal cancer in rodents with lyophilized berries: Potential mechanisms.* Nutrition and Cancer-an International Journal, 2006. **54**(1): p. 33-46.

27. Kim, H.J., et al., *Anthocyanins from soybean seed coat inhibit the expression of TNF-alpha-induced genes associated with ischemia/reperfusion in endothelial cell by NF-*

kappa B-dependent pathway and reduce rat myocardial damages incurred by ischemia and reperfusion in vivo. Febs Letters, 2006. **580**(5): p. 1391-1397.

28. National Institute of Diabetes and Digestive and Kidney Diseases. National Diabetes Statistics, 2007 fact sheet. Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, 2008.

29. Matsui, T., et al., *alpha-glucosidase inhibitory action of natural acylated anthocyanins. 1. Survey of natural pigments with potent inhibitory activity.* Journal of Agricultural and Food Chemistry, 2001. **49**(4): p. 1948-1951.

30. Matsui, T., et al., *alpha-glucosidase inhibitory action of natural acylated anthocyanins. 2. alpha-glucosidase inhibition by isolated acylated anthocyanins.* Journal of Agricultural and Food Chemistry, 2001. **49**(4): p. 1952-1956.

31. Matsiu, T., et al., *Anti-hyperglycemic effect of diacylated anthocyanin derived from Ipomoea batatas cultivar Ayamurasaki can be achieved through the alpha-glucosidase inhibitory action.* Journal of Agricultural and Food Chemistry, 2002. **50**(25): p. 7244-7248.

32. McDougall, G.J., et al., *Different polyphenolic components of soft fruits inhibit alpha-amylase and alpha-glucosidase.* Journal of Agricultural and Food Chemistry, 2005. **53**(7): p. 2760-2766.

33. Wagner, H. and G. Ulrich-Merzenich, *Synergy research: Approaching a new generation of phytopharmaceuticals.* Phytomedicine, 2009. **16**(2-3): p. 97-110.

34. Marais, J.P.J., D. Ferreira, and D. Slade, *Stereoselective synthesis of monomeric flavonoids.* Phytochemistry, 2005. **66**(18): p. 2145-2176.

35. Saladino, R., et al., *Advances and challenges in the synthesis of highly oxidised natural phenols with antiviral, antioxidant and cytotoxic activities.* Current Medicinal Chemistry, 2008. **15**(15): p. 1500-1519.

36. Deroles, S., *Anthocyanin Biosynthesis in Plant Cell Cultures: A Potential Source of Natural Colourants*, in *Anthocyanins: Biosynthesis, Functions, and Applications*, K. Gould, K. Davies, and C. Winefield, Editors. 2009, Springer Science+Business Media: New York.

37. Fowler, Z.L. and M.A.G. Koffas, *Biosynthesis and biotechnological production of flavanones: current state and perspectives.* Applied Microbiology and Biotechnology, 2009. **83**(5): p. 799-808.

38. Chemler, J.A., et al., *Standardized biosynthesis of flavan-3-ols with effects on pancreatic beta-cell insulin secretion.* Applied Microbiology and Biotechnology, 2007. **77**: p. 797-807.

39. Chemler, J.A., Y.J. Yan, and M.A.G. Koffas, *Biosynthesis of isoprenoids, polyunsaturated fatty acids and flavonoids in Saccharomyces cerevisiae.* Microbial Cell Factories, 2006. **5**.

40. Leonard, E. and M. Koffas, *Alteration of fatty acid metabolism in Escherichia coli for high-level production of plant-specific flavonoids.* Abstracts of the General Meeting of the American Society for Microbiology, 2006. **106**: p. 438.

41. Leonard, E., et al., *Strain improvement of recombinant Escherichia coli for efficient production of plant flavonoids.* Molecular Pharmaceutics, 2008. **5**(2): p. 257-265.

42. Leonard, E., Y.J. Yan, and M.A.G. Koffas, *Functional expression of a P450 flavonoid hydroxylase for the biosynthesis of plant-specific hydroxylated flavonols in Escherichia coli.* Metabolic Engineering, 2006. **8**(2): p. 172-181.

43. Leonard, E., et al., *Investigation of two distinct flavone synthases for plant-specific flavone biosynthesis in Saccharomyces cerevisiae.* Applied and Environmental Microbiology, 2005. **71**(12): p. 8241-8248.

44. Yan, Y.J., et al., *Metabolic engineering of anthocyanin biosynthesis in Escherichia coli.* Applied and Environmental Microbiology, 2005. **71**(7): p. 3617-3623.

45. Yan, Y., *Constructing microbial production platform for the biosynthesis of natural drug candidates-flavonoids*, in *Chemical and Biological Engineering*. 2008, State University of New York at Buffalo: Buffalo. p. 218.

46. Fowler, Z.L., W.W. Gikandi, and M.A.G. Koffas, *Increased Malonyl Coenzyme A Biosynthesis by Tuning the Escherichia coli Metabolic Network and Its Application to Flavanone Production.* Applied and Environmental Microbiology, 2009. **75**(18): p. 5831-5839.

47. Reed, J.L., et al., *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).* Genome Biology, 2003. **4**(9).

48. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

49. Ranganathan, S., P.F. Suthers, and C.D. Maranas, *OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions.* Plos Computational Biology, 2010. **6**(4).

50. Ford, C.M., P.K. Boss, and P.B. Hoj, *Cloning and characterization of Vitis vinifera UDP-glucose : flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo.* Journal of Biological Chemistry, 1998. **273**(15): p. 9224-9233.

51. Offen, W., et al., *Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification.* Embo Journal, 2006. **25**(6): p. 1396-1405.

52. Xie, X., et al., *Rational Improvement of Simvastatin Synthase Solubility in Escherichia coli Leads to Higher Whole-Cell Biocatalytic Activity.* Biotechnology and Bioengineering, 2009. **102**(1): p. 20-28.