5-2019

Preparation of a Flavonol Specific Glucosyltransferase found in Grapefruit and Site-Directed Mutants for Protein Crystallization

Aaron Birchfield
East Tennessee State University

Preparation of a Flavonol Specific Glucosyltransferase found in Grapefruit and Site-Directed Mutants for Protein Crystallization

---

A thesis

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biology

---

by

Aaron Birchfield

May 2019

Cecilia A. McIntosh, PhD, Chair

Ranjan Chakraborty, PhD

Aruna Kilaru, PhD

Keywords: Crystallization, Protein Tags, Structure, Function, Grapefruit, Flavonol, Glucosyltransferase

ABSTRACT

Preparation of a Flavonol Specific Glucosyltransferase found in Grapefruit and Site-Directed Mutants for Protein Crystallization

by

Aaron Birchfield

This research was designed to determine the conditions necessary to remove c-myc and 6x-His tags from a flavonol specific glucosyltransferase found in grapefruit (CP3GT) using thrombin in preparation for crystallization. X-ray crystallography of CP3GT crystals may elucidate structural features that account for flavonol specificity in some glucosyltransferase enzymes. A thrombin cleavage site was inserted into WT CP3GT and one mutant. Recombinant CP3GT was expressed in yeast and purified. Optimal conditions for thrombin digestion were explored. Digestion with 100U of thrombin for 2 hours at 4° C was optimal for removing tags from CP3GT. Storage at 4° C for 2 hours resulted in approximately 70% retention of activity. The effect of thrombin treatment on CP3GT activity was tested. Purified CP3GT protein with and without tags was tested for activity with the flavonol quercetin. Data showed no significant difference in overall activity between tagged and native protein.

ACKNOWLEDGEMENTS

I am thankful first to God for blessing me with the strength and abilities necessary to undertake this project. I thank my wife, Danielle, and all my family who have supported and encouraged me through this journey. I thank my committee chair, Dr. McIntosh, for allowing me the opportunity to work in her lab. Her unwavering patience and focused attention towards my development as a scientist was instrumental to my success. Her experience and wisdom motivated me to continue working through every obstacle and I could not have finished this work without her instruction.

I thank my advisor, Dr. Kilaru, for encouraging me to undertake this program after completing my undergraduate work. I thank my advisor, Dr. Chakraborty, for his wisdom and insight regarding our efforts to crystallize. I would like to thank the ETSU School of Graduate Studies and the Biological Sciences Department for admitting me to this program and giving me the opportunity to mature as a biologist. To all the professors, students, and friends who helped make this project an enjoyable experience, thank you. Lastly, I thank Sangam Kandel and Nathan Spaulding, who first welcomed me into Dr. McIntosh's lab and gave me hands on instruction on basic lab techniques.

DEDICATION

For Danielle: “Don’t you have a thesis to write?”

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

This research is designed to explore the conditions necessary to prepare a glucosyltransferase (GT) found in grapefruit for crystallization to ultimately determine the structural basis of its unique specificity. Determining structure function relationships of GT's is a multi-faceted process that encompasses many different research methods and has numerous applications in health, agriculture, and biotechnology.

For this research to be successful, it is necessary to first have a thorough understanding of the relevant background information. What are plant metabolites and what are their purposes and applications both in plant and mammals? How are they synthesized and what is their structure? How does their structure impact their function? What plants make the best model for studying plant metabolites and which metabolites should we study? Answering these questions allows one to evaluate what metabolites are best suited for scientific exploration through research. Additionally, the enzymes responsible for the synthesis of plant metabolites are of interest to researchers.

Plant Secondary Metabolites

Plant secondary metabolites are compounds that allow plants to survive and adapt to their environment (Bourgaud et al. 2001). These compounds serve many purposes in the plant and are required for their survival as well as contributing to their reproductive fitness. They protect the plant against herbivores and microorganisms such as bacteria and fungi (Bennet and Wallsgrove 1994). They also mimic pesticides, assist with the dispersal of seeds, and give rise to pigments and scents that attract pollinators (Newman and Gordon 2016 and ref. therein). Plant secondary metabolites are classified by

their chemical structure and are divided into many groups. The three major groups of plant secondary metabolites are phenolics, terpenoids, and alkaloids (Croteau et al. 2000 and ref. therein).

Terpenoids

Terpenoids, also called isoprenoids, are the largest class of naturally occurring compounds synthesized by plants (McGarvey and Croteau 1995). Their physiological and metabolic significance in plants includes photosynthetic pigments, structural support in cell membranes, and electron carriers (Harborne 1991). Their basic structure is composed of five-carbon isoprene units that are modified in a multitude of ways. Terpenoids are synthesized through one of two metabolic pathways: the mevalonic acid (MVA) pathway or the methylerythritol phosphate (MEP/DOXP) pathway (Tholl 2015). Terpenoids are classified by their chemistry and the number of isoprenoid units in their structure (Eisenreich et al. 2001 and ref. therein). The mono- and diterpenoids, such as geranyl pyrophosphate or retinol respectively, are synthesized in the plastids of plants through the MEP pathway (Jux et al. 2001). Sesquiterpenoids, such as farnesol, are synthesized in the cytosol of plants through the MVA pathway (Jux et al. 2001). The triterpenoid dilactone, limonin, is found in navel oranges, mandarins, grapefruits, and lemons, and accounts for an intense bitter sensation when consumed (Maier et al. 1977 and ref. therein; McIntosh 2000 and ref. therein; Fayoux et al. 2007).

Alkaloids

Plant alkaloids comprise a diverse class of secondary metabolites with significant medicinal and pharmacological importance (Qiu et al. 2014 and ref. therein). These compounds typically contain a heterocyclic nitrogen that originated from an amino acid. Over 12,000 nitrogen containing alkaloids have been identified to date and include such compounds as morphine, caffeine, and nicotine (Schlager and Drager 2016). Some plants produce alkaloids to act as feeding deterrents against herbivores, such as the yellow lupin (*Lupinus luteus*) (Wang et al. 2000; Vilarino and Ravetta 2008). Other plants synthesize alkaloids with stimulant properties in their nectar to attract pollinators (Irwin et al. 2014). Through

various roles, plant alkaloids increase the overall survivability of plants and lead to greater reproductive rates (Matsuura and Fett-Neto 2015).

Phenolics

Phenolic compounds contain at least one or more aromatic hydrocarbon rings to which are attached at least one or more hydroxyl groups (Khoddami et al. 2013). This structural arrangement makes phenolic compounds slightly acidic and may partially explain their antioxidant potential (Nimse and Pal 2015). Phenolic compounds are synthesized ubiquitously by plants. Sometimes they are produced in response to environmental pressures and to negate harmful byproducts of normal plant metabolism (Klepacka et al. 2011). Two pathways are known to be involved in the biosynthesis of plant phenolic compounds: the shikimate/chorizmate or phenylpropanoid pathway and the polyketide pathway (Knaggs 2003 and ref. therein). Common examples of naturally occurring plant phenolics include tannins, coumarins, and flavonoids (Huang et al. 2010).

Flavonoids

Flavonoids are polyphenolic secondary metabolites found in all plants, with over 10,000 naturally occurring compounds identified to date (Jia 2003 and ref. therein). They are known to have antimicrobial, antifungal, and anti-cancer properties, as well as having many other pharmaceutical applications (Cushnie and Lamb 2005; Benavente-Garcia and Castillo 2008 and ref. therein; Panche et al. 2016). A typical flavonoid consists of two aromatic rings, designated A and B, that are joined to a third ring between the A and B rings (Fig. 1.1), designated as the C ring (Miller et al. 2012). This ring contains an oxygen atom and can have different oxidation states. One or more hydroxyl groups is usually present that are important sites for nucleophilic substitution (Cody 1988; Wang et al. 2018). The selectivity and activity of different flavonoids is determined by their structure and by the location of their respective hydroxyl groups (Cody 1988; Wang et al. 2018). For example, the flavanone hesperetin has hydroxyl groups at the C5 and C7

position and forms 7-O glucosides. The flavonol quercetin has a hydroxyl group at the C3, C5, and C7 position but usually forms 3-O glucosides (Yi et al 2017 and references therein).

Extensive work has been done on citrus flavonoids, with over 60 types identified over the past 20 years (Berhow et al. 1998; Lv et al. 2015)) These compounds are found in all parts of the citrus plant although at different concentrations (McIntosh and Mansell 1997; Yi et al. 2017 and ref. therein)

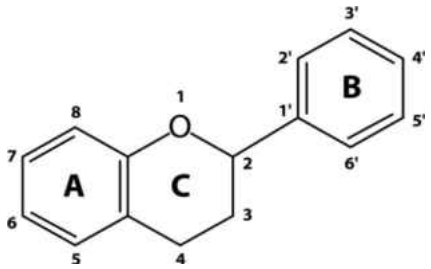


Figure 1.1: Basic flavonoid structure (Pal and Saha 2014 used with permission)

Citrus flavonoids have been shown in multiple studies to have beneficial health effects both *in vitro* and *in vivo (Chandrika et al. 2016)*. The antioxidant activity of citrus flavonoids is one of the most important markers for their bioactive potential and is usually associated with anti-cancer and anti-inflammatory properties (Gulcin 2012). Multiple studies have demonstrated the antioxidant potential of flavonoids using DPPH and ABTS radical scavenging assays (Zou et al. 2016). In citrus, flavonoids accumulate in seeds, pulp, and peels and can influence taste and smell (Yi et al. 2017). For example, the flavanone naringin is the primary bitter compound found in grapefruits (Owens and McIntosh, 2011 and ref. therein) Additionally, flavonoids in citrus protect the plant from U.V. light damage and help defend against herbivores (McIntosh and Owens, 2016 and ref. therein).

Pathways of Flavonoid Biosynthesis

The biosynthesis and production of flavonoids can be triggered by different factors such as environmental, physiological, and geographical variations of the plant under study (Figueiredo et al. 2008). Principally, the synthesis of flavonoids begins with the phenylpropanoid pathway, whereby

phenylalanine is transformed into 4-coumaroyl-CoA, which is then converted to chalcone by chalcone synthase (Heller and Hahlbrock 1980). This structure is then converted to a flavanone by chalcone isomerase (Owens and McIntosh, 2011 and ref. therein). Other classes of flavonoids such as flavonols, isoflavonoids, flavones, and anthocyanidins are then derived from this flavanone precursor (Stafford 1991; Owens and McIntosh 2011) There exist nine primary subclasses of flavonoids. The five classes found in citrus can be seen in Figure 1.2. It should be noted that anthocyanins are only found in some varieties of citrus, such as the blood orange (Kirca and Cemeroglu 2003). It should also be noted that naturally occurring flavonoid compounds are almost always found in their glycosylated form (McIntosh and Owens, 2016 and ref. therein).

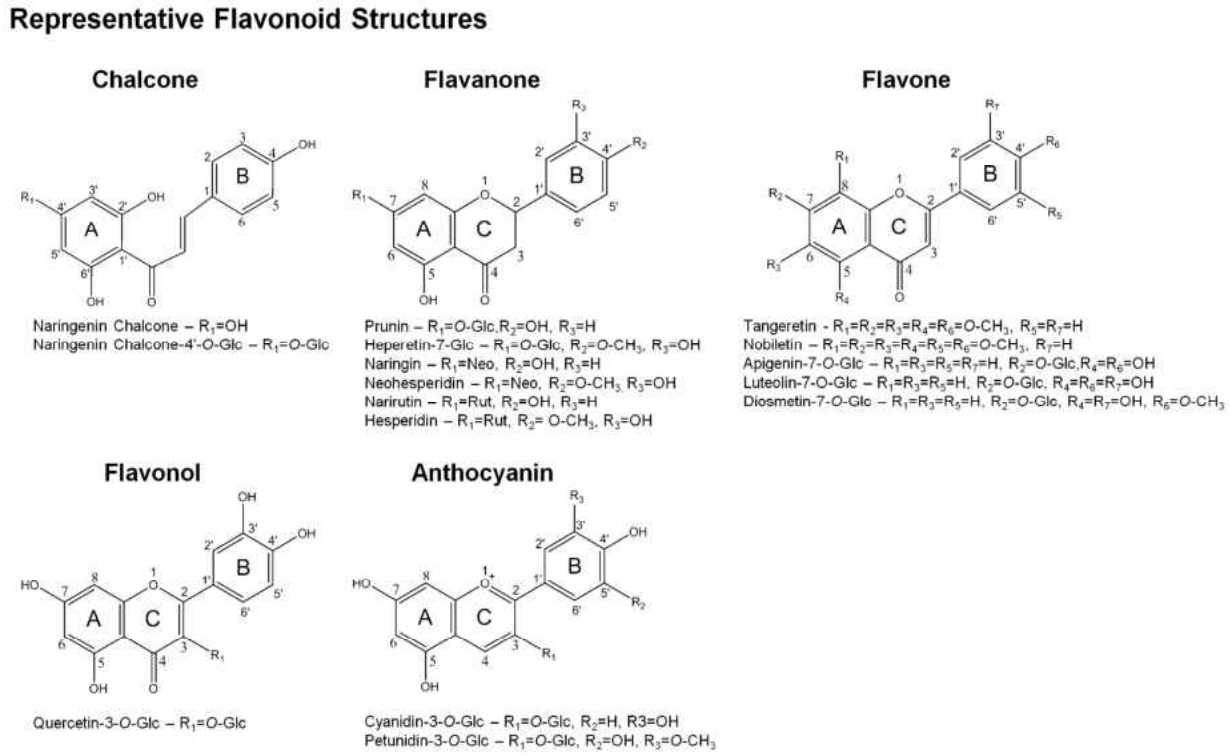


Figure 1.2: Representative Flavonoid Structures (McIntosh and Owens, 2016 used with permission)

Role of Flavonoids in Plants

The chemical properties of flavonoids give them significant functional diversity (Kelly et al 2002). Flavonoids have multiple roles in plant metabolism that allow the plant in which they are synthesized to adapt and survive. Some flavonoids, such as anthocyanins, are found in the vacuoles of mesophyll and epidermal cells and in chloroplasts, suggesting they play a role in protecting the plant from toxic substances (Alfenito et al. 1998; Agati et al. 2007). Additionally, flavonoids have been shown to protect plants from UV damage (Li et al. 1993; Agati et al. 2009). Flavonoids can also protect against microbial infection, influence taste and pigmentation, and modulate auxin distribution in the nodulation of legumes (Taylor and Grotewold 2005; Subramanian et al. 2007 and ref. therein). In citrus, flavonoids have been shown to attract and encourage egg laying by two species of butterfly, *Papilia xuthus* and *Papilio protentor demetrius* (reviewed in Owens and McIntosh, 2016). The taste of various citrus species is determined by their flavonoid profile, which is an important factor for consumption and seed dispersal (reviewed in Owens and McIntosh, 2016).

Role of Flavonoids with Respect to Mammals

Flavonoids are routinely ingested by animal species and have been shown to have a beneficial effect on health. Flavonoids act as antioxidants by scavenging free radicals and reducing inflammation (Rodriguez-Mateos et al. 2014). They have also been shown to reduce the risk of chronic heart disease when consumed as part of a balanced diet (Hooper et al. 2008; Dower et al. 2015). Flavonoids also have great potential in the development and usage of pharmaceuticals and nutriceuticals. For example, fisetin, a flavonol found in strawberries and apples, was recently shown to have a cytotoxic effect on triple-negative breast cancer cells without affecting non-malignant cell growth, as well as having a renoprotective effect against the drug cisplatin, a commonly used chemotherapeutic agent known to cause damage to the kidneys as a side effect (Sahu et al. 2014; Smith, et. al. 2016). Citrus flavonoids are also known to have many health benefits. Two of the most common flavonoids found in citrus, the flavanones

naringenin and hesperetin, were shown to increase antioxidant enzyme activity in the livers of old-age rats (Miler et al. 2016)

Properties of Glycosylation

The vast majority of flavonoids are found in glycosylated form in plants, thus there has been significant interest in exploring the mechanisms and enzymes responsible for catalyzing the reaction between a flavonoid and a nucleotide sugar. A brief background is given here that highlights recent developments in glycosyltransferase research and the potential applications of this research in biotechnology and elsewhere. Glycosylation is an important regulator of phenylpropanoid availability and localization (Le Roy et al. 2016). Glycosylation can alter the solubility of metabolites produced from this pathway and can influence compartmentalization and biological activity therein (Le Roy et al. 2016) The flavonoids found in citrus tissue are most often observed in glycosylated form (McIntosh and Owens, 2016). This modification includes the addition of a sugar group attached at one of the hydroxyl functional groups on the flavonoid acceptor compound. Flavonoid glycosides can be formed at a hydroxyl, carbon, or nitrogen group, referred to as O-glycosides, C-glycosides, and N-glycosides, respectively (Markham et al. 1978). In citrus, nearly all flavonoids are glucosylated at an -OH position, forming *O*-glycosides (McIntosh and Owens, 2016). In their glycosylated form, flavonoids are more stable (Xiao et al. 2014). Additionally, one study showed that a glucosylated flavonoid had a higher antioxidant capacity than a non-glucosylated biflavone (Yang et al. 2015).

Plant Secondary Product Glucosyltransferase

Glucosyltransferases are a subclass of glycosyltransferases and are specific for the addition of a glucose moiety to an aglycone (acceptor) molecule. Glucosyltransferases serve a host of functions in plants such as the synthesis of flavonoid glucosides. The type of flavonoid that can be glucosylated (substrate specificity) and where the sugar can be added (regio- and stereospecificity) depends on the enzyme. Some glucosyltransferases have broad regio- and stereospecificity, while others are specific to

certain compounds and/or sites for transfer (McIntosh and Mansell, 1990; McIntosh et al 1990; Tiwari et al. 2016; McIntosh and Owens, 2016 and ref. therein). In *Citrus paradisi*, a glucosyltransferase (CP3GT) was shown to exclusively glucosylate the flavonol class of flavonoids at the 3-OH position (Owens and McIntosh, 2009). A GT found in *Vitis vinifera* (VvGT1) that shares 57% sequence identity (83% similarity) with CP3GT glucosylates both flavonols and anthocyanidins (Offen et al. 2006). Despite structural similarities between CP3GT and VvGT1, the former is highly specific for flavonols, while the latter is less specific and glucosylates flavonols and anthocyanidins. A glucosyltransferase found in *Clitorea ternatea* (UGT78K6) was shown to exclusively glucosylate anthocyanidins (Hiromoto et al. 2015). Furthermore, a glucosyltransferase found in *Gentiana trifloral* (UGT73A14) is highly specific for delphinidin 3, 5-diglucoside at the 3' hydroxyl group but shows no activity with delphinidin or its 3-glucoside (Fukuchi-Mizutani et al. 2003). To contrast this high degree of specificity, a glucosyltransferase found in *Allium cepa* (UGT73G1) and one found in *Medicago truncatuLa* (UGT78G1) glucosylate various (iso)flavonoids and flavonols at different positions (Kramer et al. 2003; Modolo et al. 2009).

Structural Classification of Glycosyltransferases

Glycosyltransferases are not only found in plants, rather, they have been identified throughout animal, plant, fungi, and eubacteria kingdoms (Gachon et al. 2005). Glycosyltransferases are divided into families based on their fold type, the stereochemical outcome of the glycoside product with respect to donor substrate, sequence similarity, and function (Coutinho et al. 2003). To date there are 105 GT families (Lairson et al. 2008; Laboratory Architecture and Function of Biological Macromolecules (AFMB) 2018). GT1 family enzymes are best described as enzymatic proteins that use a nucleoside diphosphate sugar to catalyze the delivery of a sugar group to a secondary metabolite (Lairson et al. 2008). In the case of CP3GT, UDP-glucose is used to form quercetin glucoside (Owens and McIntosh 2009).

Structurally speaking, GT1 glycosyltransferases are composed of two Rossman-like domains in one of two architectural configurations known as GT-A and GT-B (Lairson et al. 2008). The main

difference between GT-A and GT-B architecture lies in how tight of an association the two Rossman domains have and where the active site is positioned. The Rossman domains of the GT-B fold are more loosely associated than GT-A folds and the site of catalysis is within a cleft as a result (Lairson et al. 2008). These enzymes that bind to a sugar donor attached to a nucleotide are called Leloir enzymes, while those that do not are called non-Leloir (Lairson et al. 2008). Most GT1 family glucosyltransferases adhere to the GT-A/GT-B architecture, yet more recently a GT-C fold was identified (Figure 1.3a) (Liang et al. 2015). Glucosyltransferases that exhibit GT-A and GT-B fold structures can produce a product with one of two stereochemical outcomes: inverting or retaining, while enzymes that have GT-C folds are non-Leloir and exhibit inverting stereoselectivity (Figure 1.4a) (Lairson et al. 2008). This means that for a retaining enzyme, the glycoside product will retain its anomeric configuration. Whatever stereochemical orientation the sugar donor has (UDP-glucose), the glycoside product will retain that stereochemical orientation (quercetin glucoside). With an inverting enzyme, the glycoside product will have its anomeric configuration inverted. The stereochemical orientation of the glycoside will be opposite of the orientation of the sugar donor.

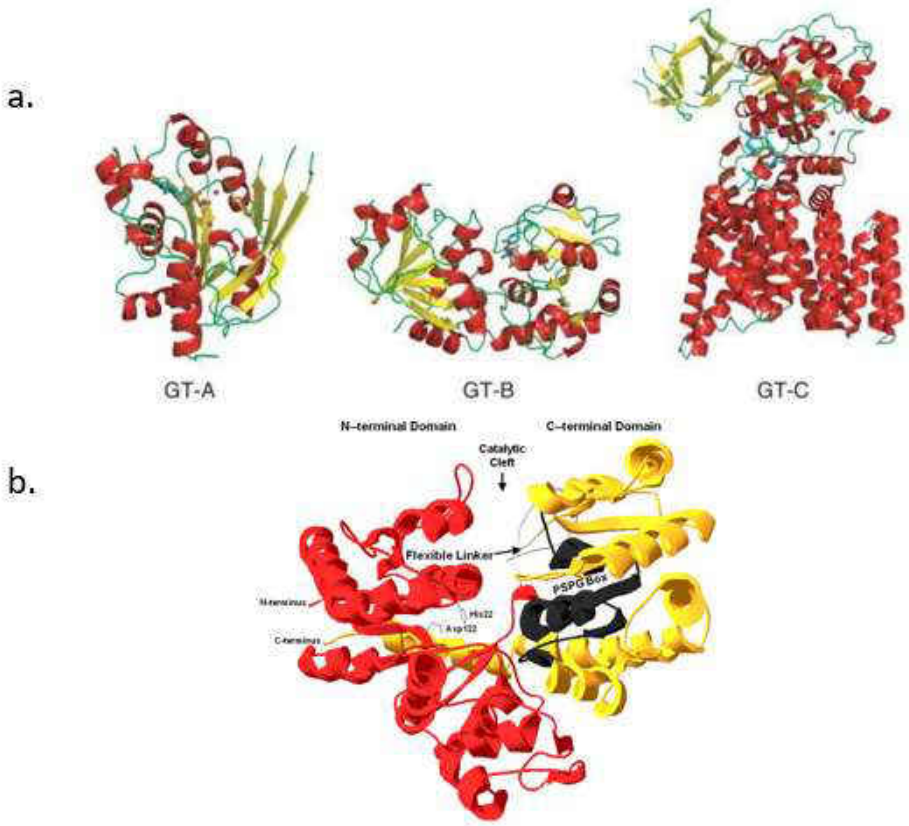


Figure 1.3: Glycosyltransferase structure and different fold types. a.) GT-A, GT-B, and GT-C fold types (Gloster 2014 used with permission). b.) Homology model of CP3GT showing GT-B fold type (McIntosh and Owens, 2016 used with permission)

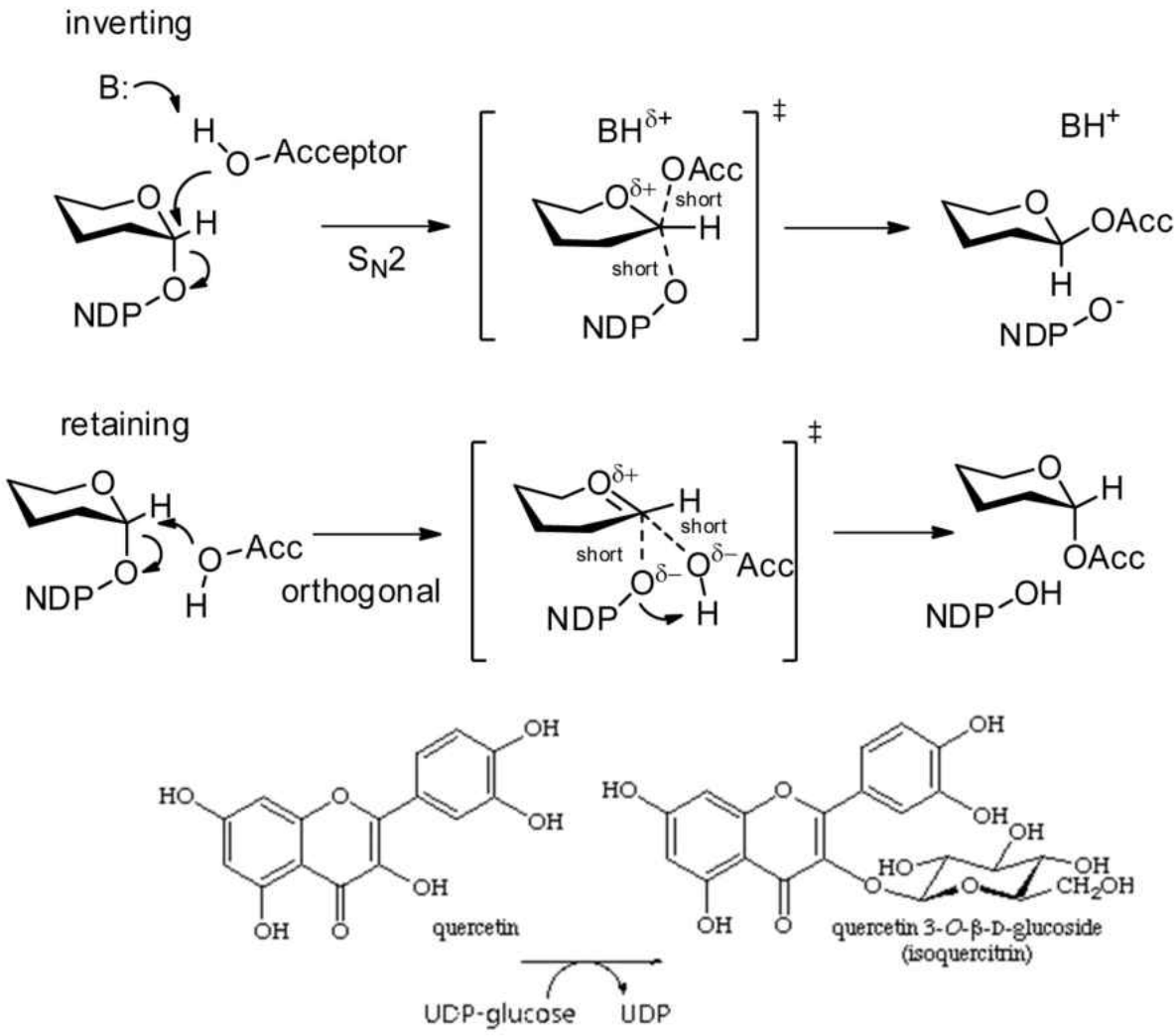


Figure 1.4: GT Stereoselectivity. a) Inverting and retaining stereoselectivity (Schuman 2013 used with permission). b) A GT family 1 enzyme demonstrates inverting stereoselectivity in the formation of quercetin glucoside (adapted from IUBMB 2004).

Glucosyltransferases classified in the GT1 family have a characteristic structural motif near the C terminus called the Plant Secondary Product Glucosyltransferase (PSPG) Box. This motif is composed of 44 amino acids and is semi-conserved in the GT1 family of enzymes. It has been determined that in GT-B enzymes this motif represents the primary UDP-sugar binding domain (Hughes and Hughes 1994; Jones and Vogt 2001; McIntosh and Owens 2016 and ref. therein). Additionally, most C-terminal alpha helices assist in the formation of the N-terminal domain by crossing over the catalytic cleft (Figure 1.3b) (McIntosh and Owens, 2016).

The most prevalent system for classifying these enzymes is the CAZy (Carbohydrate-Active Enzymes) database, that catalogues "families of structurally-related catalytic and carbohydrate-binding modules of enzymes that degrade, modify, or create glycosidic bonds" (Lombard et al. 2014; Laboratory Architecture and Function of Biological Macromolecules (AFMB) 2018). GT's are classified by amino acid sequence similarity, evolutionary relationships, and product stereoselectivity (Lairson et al. 2008). Despite growing bioinformatic databases such as CAZy, the sequence data obtained is not sufficient to predict GT function (Jackson et al. 2011; McIntosh and Owens 2016;). Due to the varying degrees of specificity of GT's and their potential applications, interest has grown towards understanding how the structure of a GT enzyme impacts its substrate and regio-specificity. Predictive models used in characterizing new enzymes rely on existing data regarding structure function relationship of enzymes, thus contributions to this database may increase reliability of predictive models.

Biotechnological Applications of Enzyme Research

Enzyme biotechnology often begins with a thorough understanding of the structure of the protein in question. More specifically, each individual amino acid of the protein must be known. This knowledge presents opportunities for site-directed mutagenesis that would allow the protein to be modified in ways that make it more valuable. An excellent, plant-based example of this method was demonstrated in 1997, when 5 amino acids of a Δ^6-palmitoyl (16:0)-ACP desaturase were mutated based on the positional equivalent amino acids of a Δ^9-stearoyl (18:0)-ACP desaturase (Cahoon et al. 1997). These mutations resulted in altered functionality of the Δ^6-palmitoyl (16:0)-ACP desaturase, allowing it to form double bonds at both the Δ^6 and Δ^9 positions (Cahoon et al. 1997). More recently, strictosidine synthase, a highly substrate-specific enzyme found in *Rauvolfie serpentine,* was crystalized in complex with its substrate to determine mutation sites that would broaden substrate specificity (Loris et al. 2007). A mutant enzyme was successfully engineered to accept a broader range of tryptamine derivatives, allowing for mass production of novel indole alkaloids that could have pharmacological significance (Loris et al. 2007).

Much of the industrial potential for enzyme applications is offset by the problems associated with keeping enzymes stable (Iyer and Ananthanarayan 2008 and ref. therein). Relatively minor changes in temperature and pressure can result in the permanent denaturing of an enzyme rendering it useless. Some of the industrial processes that use enzymes, such as the conversion of glucose to high fructose corn syrup by glucose isomerase, are more efficient at higher temperatures, yet the use of higher temperatures causes glucose isomerase to denature faster, thus requiring active enzyme replenishment (Klibanov 1983 and ref. therein). Current research routes to enhance enzyme stability include screening and exploring enzymes used by extremophiles, genetically modifying already stable enzymes, and attempting to stabilize intrinsically unstable enzyme using protein engineering and chemical modification (Iyer and Ananthanarayan 2008 and ref. therein). The latter two of these methods require a thorough understanding of enzyme structure/function relationships. This research can lead to greater enzyme stability and better activity retention making industrial processes more efficient and less costly.

There is a growing interest as to what biotechnological applications can be developed towards designing customized enzymes, specifically glucosyltransferases, capable of producing a specific product (McIntosh and Owens 2016). These enzymes could be custom designed to produce specific classes or glucosylated derivatives of flavonoids that have health benefits. Furthermore, optimization of custom enzyme syntheses could allow mass production of these health products for future study. They could also modify flavonoids in such a way as to make them more bioavailable.

Recent strides have been made towards employing custom enzymes to reduce the cost of converting biomass, such as corn or wood pulp, to biofuel. One technique employs an initial treatment phase designed to loosen the lignin/cellulose entanglement in the biomass, followed by exposure to custom cellulolytic enzymes that further digest the biomass and release hydrolysis products for fermentation (Stephanopoulos 2007). The custom enzymes used in this and similar processes cannot be engineered without a thorough understanding of enzyme structure function relationship.

Interest among commercial industries includes how flavonoids can be efficiently extracted and synthetically modified to suit a broad range of purposes (Panche et al. 2016). One study showed that the

first 4 steps in sugar production from sugarcane retain significant levels of various flavonoids within the sugarcane juice and resulting filtrates (Li et al. 2011). The same study successfully employed a comprehensive new strategy for recovering flavonoids at each step in the sugar making process, making the utilization of a highly profitable, industrial crop even more efficient, while simultaneously increasing the availability of flavonoids for industrial and research purposes (Li et al. 2011).

Additional biotechnological applications are being considered as to how protein engineering can improve crop production. Widespread use of crops carrying the genetic information to produce an insecticidal endotoxin found in *B. thuringiensis* (Bt) have raised concerns that insect resistance will develop, and indeed, at least two cases have been reported where crop pests developed resistance to Bt endotoxins (Gould 1998; Soberon et al. 2007). After elucidation of the mechanism responsible for acquired insect resistance, a Bt protein was engineered that successfully countered the insect's ability to metabolize it, restoring Bt's insecticidal potential (Soberon et al. 2007).

Tangible research developments such as these employ a straightforward approach that begins with the identification of a protein of interest that is typically responsible for one or more beneficial traits within the organism that produces it. Next, the three-dimensional structure of the protein is elucidated either by homology modeling or protein crystallization, the latter of the two providing more robust data. The 3D structure is used to identify candidate sites for amino acid substitution, carried out using site directed mutagenesis and transformation. The mutated protein is expressed and characterized to determine if the amino acid substitution would infer an enhanced benefit to the host organism. If such a benefit exists, mutant organisms can sometimes be engineered and commercialized (Rao 2008).

The background thus far given highlights a plethora of opportunities with regard to enzyme research. Glucosyltransferases found in citrus have many unique properties and are a significant research focus in this work.

The Model Plant, *Citrus paradisi*

Citrus fruits are some of the most widely consumed fruits in the world. In the United States alone an average of approximately 10 million tons of citrus are produced annually at an average value of approximately 3 billion dollars 2017). Citrus fruits contain significant levels of flavonoids and other secondary metabolites (Owens and McIntosh, 2011; Panche et al. 2016; Yi et al. 2017). To date, over 60 different types of Citrus flavonoids are known to exist with many more being discovered using advanced techniques (Berhow et al. 1998; Lv et al. 2015). Of the many flavonoid compounds found in citrus, flavonols, flavones, and flavanones are all found in high concentrations relative to other secondary metabolites, with flavanones being particularly high in concentration (Yi et al. 2017 and ref. therein). All are found in their glycosylated form (McIntosh et al. 1990). Of the flavonoid glycosides, flavanone- and flavone-7-O-diglycosides are the most common (Owens and McIntosh, 2011 and ref therein).

Citrus fruits are consumed worldwide and include such varieties as orange, grapefruit, lemon, and pomelo. The forecasted worldwide production of citrus for 2018 is estimated to be 49.3 million metric tons, of which Brazil and the United States make up approximately half 2018). Grapefruits are one of the popular citrus fruits, accounting for approximately 10 percent of all citrus produced from 2016-17 in the United States, with an estimated value of approximately 250 million dollars 2017). Oranges are the most popular of the Citrus species in the United States, valued in 2016-17 at 1.8 billion dollars 2017). Thus, the commercial, pharmaceutical, and nutraceutical importance of citrus fruits should not be understated.

Citrus GT's

Glucosyltransferases found in citrus share similar properties with other plant GT's. The citrus glucosyltransferases whose sizes have been reported reside in the 49-56 kDa range (McIntosh and Owens 2016 and ref therein). Studies on optimal pH for activity showed an optimal range of 6.5-8.0, and for citrus GT's whose pI has been studied, a pI in the acidic range was reported (McIntosh and Owens, 2016 and ref therein). Additionally, all biochemically characterized Citrus GT's are inhibited by some divalent

cations (Cu^{2+}, Fe^{2+}, Zn^{2}) and are competitively inhibited by UDP (McIntosh and Owens, 2016 and ref therein). Reports vary with regards to substrate specificity, but highly enriched, homogeneous GT extracts from *Citrus paradisi* were reported to contain a flavanone-specific GT. (McIntosh et al. 1990). A flavonol specific glucosyltransferase from *Citrus paradisi* was cloned, heterologously expressed, and is discussed more thoroughly later in this thesis (Owens and McIntosh 2009).

When compared with similar glucosyltransferases in other plants, the question arose as to what structural features in these Citrus GT's account for their specificity. This question can be best evaluated by obtaining a 3-dimensional picture of the enzyme using x-ray crystallography, but there are many steps that must first be taken before crystallization can be a viable option. First, the enzyme must be thoroughly characterized.

Flavonol Specific 3-O Glucosyltransferase from Grapefruit

A glucosyltransferase found in Duncan Grapefruit (*Citrus paradisi*) was previously identified, recombinantly expressed, and shown through biochemical characterization to exclusively glucosylate the flavonol class of flavonoids at the 3-OH position (Figure 1.2) (Owens and McIntosh, 2009). This enzyme is called *Citrus paradisi* flavonol specific 3-O-glucosyltransferase (CP3GT). It has a molecular weight of 51.2 kDa and an isoelectric point of 6.27. CP3GT demonstrated Michaelis-Menten kinetics with a K_m of 67μM for quercetin and a K_m of 669μM for UDP-glucose. Reactions proceeded optimally when incubated at 37° C and at a pH of 7.5. The use of a Tris-HCl buffer reduced enzyme activity by 51% under optimal pH, thus reactions are now performed in phosphate buffer (Owens and McIntosh, 2009). CP3GT enzyme activity was shown to be significantly inhibited by the divalent metal ions Cu^{2+} and Zn^{2+} as is reported for most GT's that have been tested (reviewed in Vogt and Jones 2000; Owens and McIntosh, 2009; reviewed in Owens and McIntosh 2011). Additionally, the reaction product UDP inhibited CP3GT enzyme activity with a K_i of 69.5μM (Owens and McIntosh, 2009).

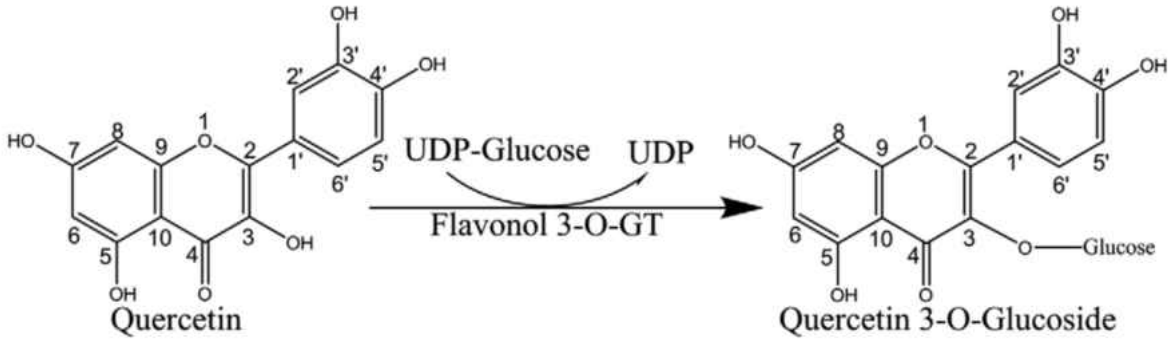


Figure 1.5: Characteristic reaction catalyzed by CP3GT (Owens and McIntosh, 2009 used with permission)

CP3GT was first expressed using *E. coli* but after finding significant amounts of insoluble proteins within *E. coli* inclusion bodies, the expression system was transferred to *Pichia pastoris*. (Owens and McIntosh, 2009; Devaiah et al. 2016). CP3GT was screened for activity with flavanone, dihydroflavonol, flavone, flavonol, isoflavone, and anthocyanidin classes of flavonoids. Of those screened, only the flavonol compounds kaempferol, quercetin, and myricetin were glucosylated. All were glucosylated at the 3-OH position (Owens and McIntosh, 2009). Various enzyme modifying reagents were employed to infer what amino acid residues are involved directly or indirectly with the active site. The data suggested that cysteine, histidine, arginine, tyrosine, and tryptophan all played some role, possibly by helping to maintain tertiary structure (Owens and McIntosh, 2009). Furthermore, a catalytic histidine residue acts as a Bronsted base, as in other GT's, and is conserved in all six crystallized plant GT's (Lairson et al. 2008; reviewed in McIntosh and Owens 2016).

Crystal Structures of Plant GT's

After characterizing a newly discovered enzyme, the next logical step was to compare it with other plant secondary product glucosyltransferases. There are currently six solved crystal structures of plant GT's for comparison with CP3GT (Table 1.1). The similarities between CP3GT and these 6 GT's

can be seen in Table 1.2. The first plant GT crystal structure to be solved was in 2005 from a small legume plant, *Medicago truncatula* (Shao et al. 2005). This enzyme, UGT71G1, facilitates the synthesis of triterpene saponins through the glucosylation of sapogenins but *in vivo* favors the synthesis of quercetin glycosides (Achnine et al. 2005). The 3D structure supports the hypothesis that a His-22 residue acts as a catalytic base and an Asp-121 residue assists deprotonation by forming an electron transfer chain with the histidine residue. Additionally, it was hypothesized that a Glu-381 residue recognizes the UDP donor substrate by forming hydrogen bonds with the 3-OH and 4-OH atoms of the glucose ring (Shao et al. 2005). These results were further confirmed by site directed mutagenesis, which showed that when His-22, Asp-121, and Glu-381 were substituted with Ala in three respective mutants, complete loss of enzyme activity was observed (Shao et al. 2005).

Table 1.1: Plant Glucosyltransferases with Solved Crystal Structures

GTs	Function	Reference
UGT71G1, *Medicago truncatula*	Triterpenes and flavonols	Shao, et al. 2005
VvGT1, *Vitis vinifera*	Anthocyanidins, flavonols	Offen, et al. 2006
UGT72B1, *Aribidpsis thaliana*	Dichloroanilines and trichlorophenols	Braizer-Hicks, et al. 2007
UGT85H2, *Medicago truncatula*	Flavonol, isoflavone	Li, et al. 2007
UHT78G1, *Medicago truncatula*	Flavonols, isoflavones, anthocyanidins	Modolo, et al. 2009
UGT78K6, *Clitoria ternatea*	Anthocyanidins	Hiromoto, et al. 2013, 2015

Table 1.2: Percent similarity and identity of CP3GT with crystalized plant GT's

Crystal Structures	Plant Source	Identity with CP3GT	Similarity to CP3GT	Reference
UGT71G1	*Medicago truncatula*	32.5%	65.6%	Shao, et al. 2005
VvGT1	*Vitis vinifera*	56.7%	82.3%	Offen, et al. 2006
UGT71B1	*Arabidopsis thaliana*	28.3%	58.7%	Brazier-Hicks, et al. 2007
UGT85H2	*Medicago truncatula*	25.5%	61.7%	Li, et al. 2007
UGT78G1	*Medicago truncatula*	47.5%	78.2%	Modolo, et al. 2009
UGT78K6	*Clitoria ternatea*	43.0%	74.4%	Hiromoto, et al. 2013, 2015

The second plant GT to be crystallized was in 2006 from the red grape plant *Vitis vinifera* (VvGT1) (Offen et al. 2006). This enzyme is specific for the glucosylation of anthocyanidins and the flavonols quercetin and kaempferol, although relative activity rates towards flavonols are about 100 times lower than anthocyanidins (Ford et al. 1998). Using the 3D structure, this study reinforced the previously determined function of the PSPG box in that it defines the location of the nucleotide-sugar binding site (Hughes and Hughes 1994; Offen et al. 2006). Three key residues were identified that interacted directly with the -OH groups of the glucose moiety of UDP-Glucose, Asp374, Gln375, and Thr141. A His-20 residue was also identified whose function is analogous with the His-22 residue found in UGT71G1. Recombinant mutant protein confirmed this; when each of these 4 amino acids was mutated, all mutant proteins showed decreased or total loss of activity (Offen et al. 2006).

A third plant GT was crystallized in 2007 and came from the model weed *Arabidopsis thaliana* (UGT72B1) (Brazier-Hicks et al. 2007). This GT glucosylates chloroanilines, an important precursor to many plant pesticides, and is unique in that it is the only GT so far identified to catalyze the formation of both N and O-glucosides. Structurally this GT differs from the other 5 crystallized GT's in that the

catalytic histidine residue His19 interacts with the Ser14 residue as opposed to interacting with the Asp117 residue. This difference may account for UGT72B1's ability to form N-glucosides and its unique chemistry with the acceptor substrate (Brazier-Hicks et al. 2007).

The fourth plant GT to be crystalized was in 2007 from the model legume *Medicago truncatula* (UGT85H2) (Li et al. 2007). This multifunctional glucosyltransferase was shown to glucosylate kaempferol and quercetin (flavonols), genistein and biochanin A (isoflavones), and iso-liquiritigenin (chalcone) (Li et al. 2007). UGT85H2 was shown to produce 7-O glucosides of biochanin A and genistein and the 3-O glucoside of kaempferol (Li et al. 2007). The sugar donor binding site within the pocket of UGT85H2 was very similar with the two previously crystallized plant GT's. Crucial amino acid residues that interacted with the sugar donor were found in the PSPG box and included Cys361, Gln363, and Glu386, that formed hydrogen bonds with the uracil and ribose ring, respectively. One significant difference observed in the PSPG box between UGT85H2 and UGT71G1 or VvGT1 included a Trp360 residue that faced the opposite direction, suggesting a conformational change after sugar donor acceptance (Li et al. 2007).

A fifth plant GT was crystallized in 2009 also from *Medicago truncatula* (UGT78G1) (Modolo et al. 2009). This GT showed activity with kaempferol, myricetin, formononetin, pelargonidin, and cyanidin. An interesting characterization of this GT was its ability to also behave like a glucosidase, converting the biochanin, genistein, kaempferol, and quercetin glucosides back to their respective aglycone form, but not regenerating UDP-glucose (Modolo et al. 2009). A glutamate at position 192 was identified as a key residue for the glucosidase reaction. Similar properties of the binding site were reported that involved most residues within the PSPG box (residues Trp334-Gln377) and some outside the PSPG motif (Ser308, Thr25, and Ser282). Four amino acids were identified (Trp355, Asp376, Gln377, and Thr141) that interact directly with the glucose moiety and are postulated to be crucial for determining sugar donor specificity for the enzyme (Modolo et al. 2009).

The sixth and most recent glycosyltransferase to be crystallized, UGT78K6, came from *Clitoria ternatea* and was shown to glucosylate the anthocyanidins delphinidin, malvidin, peonidin, and petunidin.

Trace glucosylation activity with flavonols was detected with isorhamnetin, quercetin, kaempferol, and myricetin, but percent relative activity was very low (Hiromoto et al. 2015). This indicates that UGT78K6 glucosylates anthocyanidins but does not glucosylate flavonols to an appreciable extent. In the donor binding site, the side chains of Asp367 and Gln368 interact with the glucose moiety on UDP-glucose (Hiromoto et al. 2015). These residues are conserved in VvGT1, which also glucosylates anthocyanidins, and have the same function. The catalytic residues His17 and Asp114 were also conserved. In the acceptor-binding site, Pro78, Asp 181, and Asp367 all interacted with delphinidin (Hiromoto et al. 2013; Hiromoto et al. 2015).

Strong indications of the structural basis that accounts for selectivity of glucosyltransferase has been determined by protein crystallization in these six plant glucosyltransferases, especially when co-crystallized with substrates (Shao et al. 2005; Offen et al. 2006; Li et al. 2007; Brazier-Hicks et al. 2007; Modolo et al. 2009; Hiromoto et al. 2013; Hiromoto et al. 2015). These crystal structures have provided strong clues as to the molecular basis for their respective enzyme specificities. This includes insights such as which amino acid residues are crucial for docking, stabilization, and transfer of sugar donor to aglycone substrate. The established reaction mechanism for all 6 GT's involves an imidazole side chain from a histidine residue that acts as a Bronsted base. The negative charge that develops from this interaction is stabilized by an aspartate residue. Both amino acids are conserved in all 6 crystallized glucosyltransferases (Lairson et al. 2008; McIntosh and Owens, 2016).

Out of the six, VvGT1 and UGT78K6 have been of the greatest interest to our research. VvGT1 has been shown to glucosylate both flavonols and anthocyanidins, while UGT78K6 only glucosylates anthocyanidins (Offen et al. 2006; Hiromoto et al. 2015). VvGT1 shares 57% sequence identity and 82% similarity with CP3GT while UGT78K6 shares 43% sequence identity and 74% similarity with CP3GT (Table 1.3). Despite their structural similarities, UGT78K6 does not glucosylate flavonols and CP3GT does not glucosylate anthocyanidins. VvGT1, however, glucosylates both. The structural similarities and complementary specificities that CP3GT shares with these crystallized GT's make CP3GT an excellent candidate for crystallization. Analysis of a flavonol – specific GT crystal structure in conjunction with an

anthocyanin – specific and a flavonol/anthocyanin – specific GT crystal structure will increase the likelihood of identifying key structural features that account for flavonol and anthocyanin specificity.

Despite the massive potential obtaining a crystal structure for CP3GT holds, this would be enhanced by including further study of mutants designed to test substrate and regiospecifity. Work done on the structure – function relationship of GT's has been successful using tertiary structures obtained from x-ray crystallography and homology modeling, as well as homology modeling with other GT's with different properties (McIntosh and Owens, 2016 and ref. therein). Due to its high similarity, VvGT1 was used to identify candidate amino acid residues for site directed mutagenesis in previous work. This and other modeling has resulted in the production and characterization of over 30 CP3GT mutants (McIntosh and Owens, 2016; Kandel 2016; Sathanantham 2015; Adepoju 2014; Devaiah et al. 2018). Three mutants stand out due to their unique properties and are outlined below.

CP3GT Mutants of Interest

The S20G-T21S mutant was designed in 2014 in which serine at residue 20 was mutated to glycine and the threonine at residue 21 mutated to serine (Sathanantham 2015). This glucosyltransferase showed broadened substrate specificity to include the flavanone naringenin and favored the flavonol kaempferol over quercetin. Additionally, the mutant showed 16% greater overall activity compared to wild type (Sathanantham 2015). Docking analysis with kaempferol suggested that the catalytic residue His-22 was shifted approximately 2 Å closer to the 3-OH site of glucosylation. When docked with quercetin, the histidine residue was only 0.3 Å closer (Sathanantham 2015). This is postulated to be the reason why the mutant enzyme prefers kaempferol over quercetin. Docking with naringenin suggested that the catalytic residue is 4 Å closer to the 7-OH site of glucosylation relative to wild type and therefore could account for the broadening of substrate specificity (Sathanantham 2015).

Biochemical assays showed that P297F, in which a proline located at residue 297 was mutated to a phenylalanine (P297F), lost all activity with the flavonol quercetin and did not show broadened substrate specificity (Adepoju 2014). *In silico* modeling suggested the mutation shifted a loop near the

catalytic cleft such that it blocked the binding pocket, thus eliminating enzyme activity (Adepoju 2014), however this needs to be experimentally verified. The mutation in P297F was selected based on comparisons of amino acid residue sequences between two flavonoid 3-O GTs and two flavonoid 7-O GT's, with the hypothesis that this mutation might broaden regiospecificity in CP3GT to glucosylate at the 7-OH position. Three proline residues that have position equivalent residues in VvGT1 and UGT78K6 were selected. The prolines are at position 284, 285, and 286 in VvGT1, and 277, 278, and 279 in UGT78K6 (Adepoju 2014).

The P145T mutant arose from modeling CP3GT with VvGT1 (Kandel 2016). A threonine at position 145 in VvGT has a positionally equivalent proline residue in CP3GT. Threonines are common in the functional centers of proteins and would impose less conformational rigidity on the enzyme than a proline. A CP3GT-P145T mutant was expressed and characterized and shown to have broadened substrate specificity for the flavanone naringenin in a screening assay (Kandel 2016). Despite docking models predicting that no reaction could occur at the 7-OH site, P145T glucosylated naringenin at the 7-OH position, demonstrating broadened regiospecificity (Kandel 2016). Additionally, screening assays showed that kaempferol had 119 percent activity relative to quercetin (Kandel 2016). These result highlights the need for follow up of *in silico* models with experimental testing as some of the models were not accurate. The crystallization and analysis of three CP3GT mutants with altered enzyme activity will further enhance our analysis of CP3GT enzyme structure function relationship and may help inform modeling programs.

Following characterization of an enzyme, one must determine conditions necessary to produce the enzyme at a sufficient quantity, purity, and activity for crystallization to occur. CP3GT is currently expressed in *Pichia pastoris* and contains tags that allow it to be purified using cobalt column chromatography and visualized using western blot. These two tags, the c-myc and the 6x His, are used for reliably obtaining pure CP3GT, but should not be present when attempting to crystallize; crystallization of native enzyme is preferred as one cannot be sure of impact of tags on protein structure. Consequently, methods for removing the tags while maintaining enzyme activity must be determined. A well-

documented way of removing tags is to insert an amino acid sequence upstream of the tags that can be recognized and cleaved by a protein; in this case we used bovine thrombin. Thrombin cleavage is good for our enzyme because there are no other sites in the CP3GT sequence recognized by thrombin, therefore we do not have to worry about aberrant internal cleavage (Owens and McIntosh 2009). Additionally, conditions for thrombin digestion must not degrade the enzyme to the extent that it is no longer active. Therefore, a range of conditions should be explored to determine what concentration of thrombin is necessary to cleave tags at a temperature and time that will preserve CP3GT activity. A significant portion of the work presented here aimed to first insert the thrombin cleavage sequence and then determine the conditions necessary for cleavage. The following background highlights the current research on the effect of tags.

Effect of Tags on Enzyme Activity

Some studies have indicated that vector tags attached to enzymes can alter their activity while others suggest that no effect exists. For example, a C-terminal tagged form of topinone reductase, a short chain dehydrogenase, was functionally impaired by a hexahistidine tag (Freydank et al. 2008). This impairment occurred only when the tag was inserted at the C-terminus. No impairment was observed when the tag was placed on the N-terminus (Freydank et al. 2008). In another study, an N-terminal hexahistidine tag attached to the catalytic sub-unit of phosphoinositide-3-kinase had no impact on lipid kinase activity but did alter protein kinase activity (Dickson et al. 2013).

Regarding glycosyltransferases, one study showed that a hexahistidine tag produced no effect on the catalytic activity of DesVII, a glycosyltransferase involved in the biosynthesis of some macrolide antibiotics (Borisova and Liu 2010). In another study, a hexahistidine tag present in a commercially available rhGalNAcT2 glycosyltransferase, tested as a control against a codon optimized HisDapGalNAcT2 glycosyltransferase, may have resulted in a lower specific activity of the HisDapGalNAcT2 during assays (Lauber et al. 2015).

CP3GT has also been tested for tag effects. Under optimal conditions, the N-terminal thioredoxin and hexahistidine tags, used when CP3GT was expressed in *E. coli*, presented no effects on enzyme activity. When assays were performed using Tris-HCl (pH 7.5) however, a 70% decrease in activity for tagged CP3GT was observed (Owens and McIntosh 2009). CP3GT is currently expressed using *Pichia pastoris* and uses c-myc and 6x-His tags on the C terminus. The impact of these tags has not yet been assessed.

Once reliable conditions for digesting the c-myc and 6x His tags were determined, the next logical step would be to verify that the enzyme is still active. One would do this by treating a CP3GT protein sample with thrombin while subjecting another CP3GT protein sample to the same conditions (without adding thrombin) and conducting an activity assay on both samples. After verifying activity, the next step toward crystallization would be to begin screening conditions for obtaining crystallized CP3GT protein. Once viable conditions for crystal growth are identified, the process can be further improved until crystals suitable for x-ray diffraction are obtained. The work done in this thesis was designed to make future research on CP3GT crystallization possible. Decoding the diffraction pattern for said crystal and solving the 3-D structure of CP3GT will allow us to answer questions regarding the structure function relationship of CP3GT's flavonol specificity.

As previously highlighted, protein crystallography is an invaluable tool for determining the subtle differences that account for the varying degrees of specificities in glucosyltransferases. The solving of a crystal structure allows a 3D image of the protein in question to be resolved at the molecular level, making the molecular interactions of each amino acid and its functional parts visible to the researcher (Smyth and Martin 2000). For example, CP3GT is quite similar (57%) to VvGT1, and both can glucosylate flavonols, but only VvGT1 is capable of glucosylating anthocyanins. Because the 3D structure of VvGT1 is already known, CP3GT is an excellent candidate for protein crystallization, potentially shedding light on the structural basis for flavonol specificity. Despite the known advantages of protein crystallography, solved crystal structures have only been obtained for six plant

glucosyltransferases, with the first occurring in 2005 and the most recent in 2015 (Table 1.1). Genomic databases have catalogued thousands of GT sequences across many different plant species, yet to date, approximately 200 putative secondary product plant glucosyltransferases have been identified and about 40 biochemically characterized (Vaistij et al. 2009; McIntosh and Owens, 2016). Given this information, a significant gap exists in our understanding of the structural basis for glucosyltransferase regio- and substrate specificity (McIntosh and Owens, 2016). Protein crystallography is an invaluable resource for bridging this gap.

Hypothesis

Understanding the structure function relationship of CP3GT's substrate and regiospecificity for flavonols at the 3-OH position will contribute to a diverse and growing range of biotechnological applications that include the custom design of enzymes to make desired compounds. Additionally, this research will contribute to expanding the classifications that define the different plant glycosyltransferases. This research is designed to test two hypotheses. The first is that conditions for thrombin digestion can be determined that will result in minimal loss of CP3GT activity. The second is that thrombin digestion will have no impact on enzyme activity.

CHAPTER 2

PREPARATION OF A FLAVONOL SPECIFIC GLUCOSYLTRANSFERASE FOUND IN GRAPEFRUIT AND SITE DIRECTED MUTANTS FOR CRYSTALLIZATION USING A THROMBIN CLEAVAGE SITE

Aaron Birchfield, Cecilia A. McIntosh

Department of Biological Sciences, East Tennessee State University, Johnson City, TN, 37614

Corresponding Author: Cecilia McIntosh

Department of Biological Sciences

East Tennessee State University

Box 70703, Johnson City, TN, 37614

Phone (423)-439-5838

e-mail: mcintosc@etsu.edu

Abstract

Citrus and other plants produce secondary metabolites that are modified by enzymes called glucosyltransferases. Glucosyltransferases are ubiquitous in plants, animals, and bacteria, but vary greatly in their specificity. Glucosides of flavonoids are a critical part of plant metabolism and many have health benefits when consumed. One glucosyltransferase (CP3GT), found in grapefruit (*Citrus paradisi*), was previously shown to exclusively glucosylate flavonols at the 3-OH position. This research was designed to prepare CP3GT for crystallization by determining the conditions necessary to collect native, active protein that is 95-99% pure. This requires the removal of a c-myc and a 6x-His tag by inserting a thrombin cleavage site upstream of the tags and digesting pure rCP3GT protein with thrombin. Wild type CP3GT and P297F mutant underwent site-directed mutagenesis to insert a thrombin cleavage site. rCP3GT was expressed in yeast using methanol induction and was purified using cobalt column chromatography. Conditions for digestion were determined using 30µg of pure protein and a thrombin range of 3-200U at 4° C and 10° C. It was determined that 3.3U of thrombin is sufficient to digest 1µg of protein when the digestion is carried out at 4° C for 2 hours. Storage at 4° C for 2 hours resulted in approximately 70% retention of CP3GT activity. The effect of thrombin treatment on CP3GT activity was tested by assaying pure CP3GT with and without tags using the flavonol quercetin as a substrate. Data showed no significant difference in activity between tagged and native enzyme. Future experiments will further test the effect of a c-myc and a 6x-His tag on enzyme activity and conditions for crystallization will be explored.

1. Introduction

Citrus fruits are some of the most widely consumed fruits in the world and contain significant levels of flavonoids and other secondary metabolites (Berhow et al. 1998; Ghasemi et al. 2009). To date, over 60 different types of Citrus flavonoids are known to exist with many more being discovered using advanced techniques (Berhow et al. 1998; Lv et al. 2015). These flavonoids have been shown in multiple studies to have beneficial biological effects in plants and animals in both *in vitro* and *in vivo* studies (Ferreyra et al. 2012; Chandrika et al. 2016). The antioxidant activity of Citrus flavonoids is one of the most important markers for their bioactive potential and is usually associated with anti-cancer and anti-inflammatory properties (Gulcin 2012). Thus, the commercial, pharmaceutical, and nutraceutical importance of citrus fruits cannot be understated.

Flavonoids are critical to plant survival and confer many unique benefits to plant defense and adaptability (Kumar and Pandey 2013 and ref. therein). They provide protection from both abiotic and biotic stressors that can include UV light damage and attacks from microbial pathogens (Winkel-Shirley 2002 and ref. therein). Some flavonoids such as anthocyanidins are responsible for pigmentation and can give plant tissues distinctive colors, promoting insect attraction (Winkel-Shirley, 2001). This, in turn, increases the likelihood of pollination and positively impacts plant reproduction. Some flavonoids such as luteolin are used by legumes to promote the symbiotic infection of nitrogen-fixing bacteria by inducing nodulation gene expression (Peters et al. 1986).

In Citrus and other plants, flavonoids are most often found in their glycosylated form. In this form they are more stable and have more bioavailability when consumed (McIntosh and Owens 2016 and ref. therein). The enzyme responsible for the formation of a glucoside is called a glucosyltransferase (GT). These GT's make up a diverse class of enzymes ubiquitous throughout the plant and animal kingdoms and vary greatly in their structural properties. There is a growing interest as to what biotechnological applications can be developed towards designing customized enzymes, specifically glucosyltransferases, capable of producing a specific product (McIntosh and Owens 2016). These enzymes could be custom

designed to produce specific classes of flavonoids that have health benefits or that can modify taste profiles of fruit juices. They could also modify flavonoids in such a way as to make them more bioavailable. Industrial processes that make use of enzymes must consider enzyme stability in their processes and would benefit from a customized enzyme capable of withstanding a variety of extreme conditions (Klibanov 1983 and ref. therein). Customizing enzymes requires a robust understanding of the structural features that account for an enzyme's specificity. This research is designed to help elucidate the structure/function relationship that gives rise to a grapefruit flavonol-specific glucosyltransferase's substrate and regio-specificity.

A flavonol specific grapefruit glucosyltransferase (CP3GT) has been cloned, recombinantly expressed, and shown through biochemical characterization to exclusively glucosylate the flavonol class of flavonoids at the 3-OH position (Devaiah et al. 2016; Owens and McIntosh 2009). Because of its specificity, this enzyme has served as a reliable model for determining how the structural identity of an enzyme can impact its functions (McIntosh and Owens 2016 and ref. thererin; Devaiah et al 2018). The similarity CP3GT has with many other plant GT's further highlights the importance of this work. To date, some 30 mutants have been designed and characterized, producing useful insight into how flavonol-specific GT's function (Kandel 2016; Devaiah et al. 2018; Adepoju 2014; Sathanantham 2015).

Homology modeling can be a first line of tests for hypothesizing the structural basis for enzyme specificity. These methods have been employed successfully with CP3GT, however, sequence homologies are sometimes too low to create reliable models. Such was the case with a recently characterized mutant, P145T, which *in silico* docking models predicted would not produce nariginen-7-O-glucoside, yet when activity assays were carried out, this product did form (Kandel 2016). X-ray crystallography of crystallized GT enzymes and co-crystallized with substrate are alternatives that allow a 3D image of the protein in question to be resolved at the molecular level, making the molecular interactions of each amino acid and its functional parts visible. To date only 6 plant secondary product glucosyltransferases have had crystal structures solved (Shao et al. 2005; Offen et al. 2006; Li et al. 2007; Brazier-Hicks et al. 2007; Modolo et

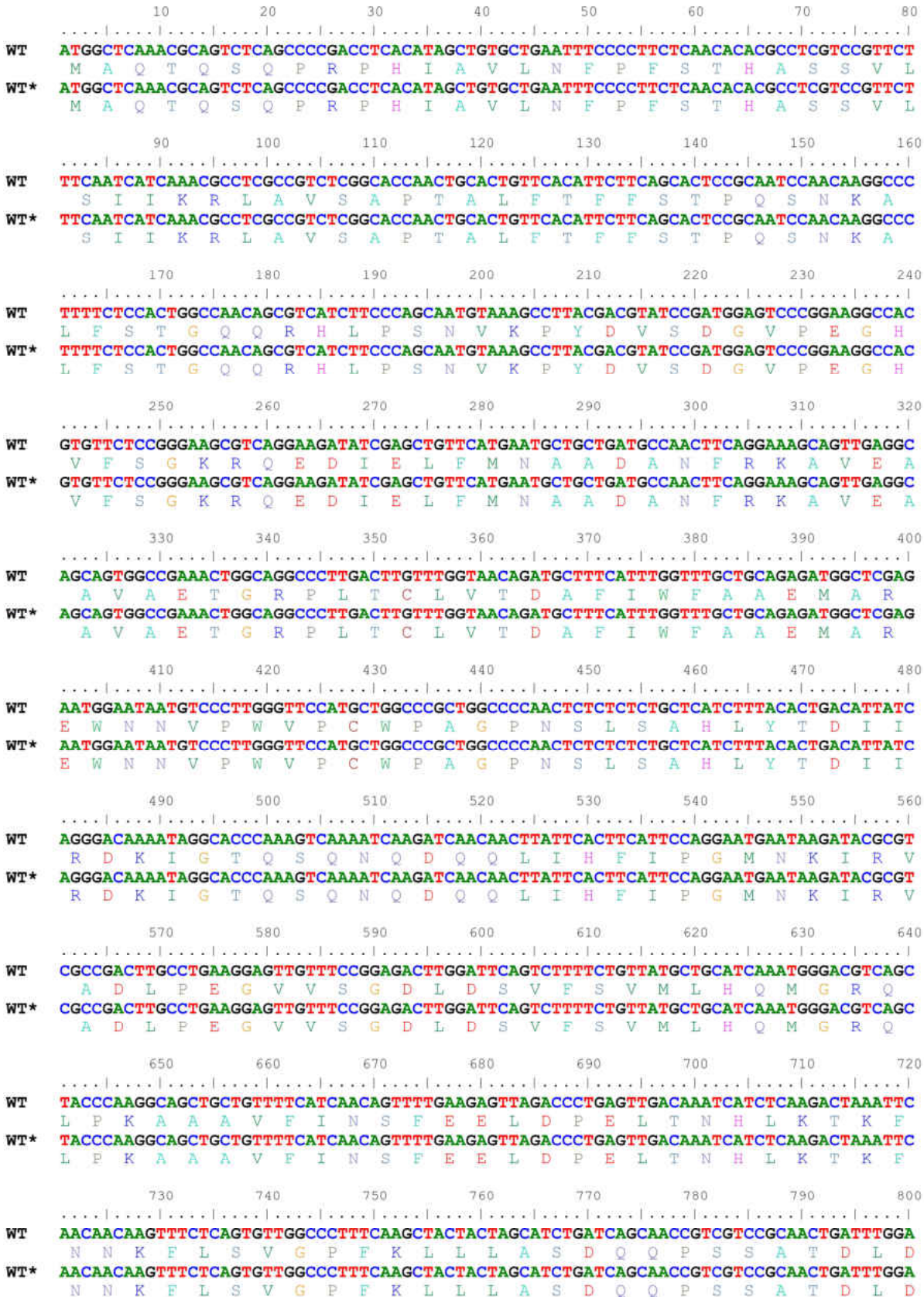
al. 2009; Hiromoto et al. 2013; Hiromoto et al. 2015). The structural similarities and differing specificities that CP3GT shares with these crystallized GT's make CP3GT an excellent opportunity for elucidating its structure function relationship using crystallization. Crystallography work has provided strong insights into the structural basis for a glucosyltransferase that is anthocyanidin specific (UGT78K6) and for a glucosyltransferase that is both anthocyanidin and flavonol specific (VvGT1), yet no structural basis currently exists for a flavonol specific enzyme such as CP3GT (Offen et al. 2006; Hiromoto et al. 2013; Hiromoto et al. 2015). If a solved crystal structure for CP3GT was obtained, it could provide significant insights into the structural basis for flavonol specificity.

Crystallizing a protein is usually undertaken only after thorough characterization studies have been carried out. As previously mentioned, CP3GT and some 30 site-directed mutants have already been characterized, yet questions still exist that only a solved crystal structure of CP3GT can answer. While it is indeed an excellent candidate for crystallization, obtaining crystals suitable for x-ray diffraction takes significant preparation. The enzyme must be at least 95% pure and be in its native form (Dessau and Modis 2011). The enzyme should also retain normal activity levels (Ferguson et al. 2002). Pure CP3GT protein can be reliably obtained using cobalt column chromatography, yet the presence of c-myc and 6x-His tags necessitates further exploration as to how native CP3GT protein can be obtained while preserving activity. The amino acid sequence L-V-P-R-G-S is recognized and cleaved by bovine thrombin and was inserted into the CP3GT gene sequence upstream of the c-myc and 6x his tags (Kandel 2016). When CP3GT is treated with thrombin, the tags are removed, and native protein can then be obtained, however, cleaving tags with thrombin should be employed in such a way that enzyme activity is maintained. Previous work has shown that CP3GT does not contain any internal thrombin cleavage sites (Owens and McIntosh 2009). One significant focus of this work involved testing conditions for thrombin cleavage so that the CP3GT enzyme would retain activity. Furthermore, this work presented the opportunity to test whether the presence of c-myc and 6x His tags impacts the activity of CP3GT.

2. Results and Discussion

2.1 Verification of CP3GT and P297F Mutant Sequencing

Before transforming CP3GT-thrombin (rCP3GT) into yeast, it was necessary to verify clone identity and the presence of the c-myc tag, 6x-his tag, and thrombin-cleavage site, the latter of which was previously inserted (Kandel 2016). Cell stocks of rCP3GT in *E. coli* were plated, a single colony was grown in LB broth, and the plasmid extracted for sequencing. The plasmid insert (asterisk) was sequenced and aligned with the gene sequence for CP3GT (Genbank Protein ID: ACS15351), as seen in Figure 2.1, to show the presence of tags and verify the sequence with thrombin cleavage site was in frame. The sequence that thrombin recognizes and cleaves, Leu-Val-Pro-Arg-Gly-Ser, can be clearly seen and is bracketed in red. The c-myc tag sequence, Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu, binds to the primary antibody we use for western blotting, and is shown by the blue set of brackets. The 6x histidine tag is necessary to bind the protein to the cobalt on the IMAC column and can be seen bracketed in green. All inserts were in frame.


10 20 30 40 50 60 70 80
WT ATGGCTCAAACGCAGTCTCAGCCCCGACCTCACATAGCTGTGCTGAATTTCCCCTTCTCAACACACGCCTCGTCCGTTCT
M A Q T Q S Q P R P H I A V L N F P F S T H A S S V L
WT* ATGGCTCAAACGCAGTCTCAGCCCCGACCTCACATAGCTGTGCTGAATTTCCCCTTCTCAACACACGCCTCGTCCGTTCT
M A Q T Q S Q P R P H I A V L N F P F S T H A S S V L
90 100 110 120 130 140 150 160
WT TTCAATCATCAAACGCCTCGCCGTCTCGGCACCAACTGCACTGTTCACATTCTTCAGCACTCCGCAATCCAACAAGGCCC
S I I K R L A V S A P T A L F T F F S T P Q S N K A
WT* TTCAATCATCAAACGCCTCGCCGTCTCGGCACCAACTGCACTGTTCACATTCTTCAGCACTCCGCAATCCAACAAGGCCC
S I I K R L A V S A P T A L F T F F S T P Q S N K A
170 180 190 200 210 220 230 240
WT TTTTCTCCACTGGCCAACAGCGTCATCTTCCCAGCAATGTAAAGCCTTACGACGTATCCGATGGAGTCCCGGAAGGCCAC
L F S T G Q Q R H L P S N V K P Y D V S D G V P E G H
WT* TTTTCTCCACTGGCCAACAGCGTCATCTTCCCAGCAATGTAAAGCCTTACGACGTATCCGATGGAGTCCCGGAAGGCCAC
L F S T G Q Q R H L P S N V K P Y D V S D G V P E G H
250 260 270 280 290 300 310 320
WT GTGTTCTCCGGGAAGCGTCAGGAAGATATCGAGCTGTTCATGAATGCTGCTGATGCCAACTTCAGGAAAGCAGTTGAGGC
V F S G K R Q E D I E L F M N A A D A N F R K A V E A
WT* GTGTTCTCCGGGAAGCGTCAGGAAGATATCGAGCTGTTCATGAATGCTGCTGATGCCAACTTCAGGAAAGCAGTTGAGGC
V F S G K R Q E D I E L F M N A A D A N F R K A V E A
330 340 350 360 370 380 390 400
WT AGCAGTGGCCGAAACTGGCAGGCCCTTGACTTGTTTGGTAACAGATGCTTTCATTTGGTTTGCTGCAGAGATGGCTCGAG
A V A E T G R P L T C L V T D A F I W F A A E M A R
WT* AGCAGTGGCCGAAACTGGCAGGCCCTTGACTTGTTTGGTAACAGATGCTTTCATTTGGTTTGCTGCAGAGATGGCTCGAG
A V A E T G R P L T C L V T D A F I W F A A E M A R
410 420 430 440 450 460 470 480
WT AATGGAATAATGTCCCTTGGGTTCCATGCTGGCCCGCTGGCCCCAACTCTCTCTCTGCTCATCTTTACACTGACATTATC
E W N N V P W V P C W P A G P N S L S A H L Y T D I I
WT* AATGGAATAATGTCCCTTGGGTTCCATGCTGGCCCGCTGGCCCCAACTCTCTCTCTGCTCATCTTTACACTGACATTATC
E W N N V P W V P C W P A G P N S L S A H L Y T D I I
490 500 510 520 530 540 550 560
WT AGGGACAAAATAGGCACCCAAAGTCAAAATCAAGATCAACAACTTATTCACTTCATTCCAGGAATGAATAAGATACGCGT
R D K I G T Q S Q N Q D Q Q L I H F I P G M N K I R V
WT* AGGGACAAAATAGGCACCCAAAGTCAAAATCAAGATCAACAACTTATTCACTTCATTCCAGGAATGAATAAGATACGCGT
R D K I G T Q S Q N Q D Q Q L I H F I P G M N K I R V
570 580 590 600 610 620 630 640
WT CGCCGACTTGCCTGAAGGAGTTGTTTCCGGAGACTTGGATTCAGTCTTTTCTGTTATGCTGCATCAAATGGGACGTCAGC
A D L P E G V V S G D L D S V F S V M L H Q M G R Q
WT* CGCCGACTTGCCTGAAGGAGTTGTTTCCGGAGACTTGGATTCAGTCTTTTCTGTTATGCTGCATCAAATGGGACGTCAGC
A D L P E G V V S G D L D S V F S V M L H Q M G R Q
650 660 670 680 690 700 710 720
WT TACCCAAGGCAGCTGCTGTTTTCATCAACAGTTTTGAAGAGTTAGACCCTGAGTTGACAAATCATCTCAAGACTAAATTC
L P K A A A V F I N S F E E L D P E L T N H L K T K F
WT* TACCCAAGGCAGCTGCTGTTTTCATCAACAGTTTTGAAGAGTTAGACCCTGAGTTGACAAATCATCTCAAGACTAAATTC
L P K A A A V F I N S F E E L D P E L T N H L K T K F
730 740 750 760 770 780 790 800
WT AACAACAAGTTTCTCAGTGTTGGCCCTTTCAAGCTACTACTAGCATCTGATCAGCAACCGTCGTCCGCAACTGATTTGGA
N N K F L S V G P F K L L L A S D Q Q P S S A T D L D
WT* AACAACAAGTTTCTCAGTGTTGGCCCTTTCAAGCTACTACTAGCATCTGATCAGCAACCGTCGTCCGCAACTGATTTGGA
N N K F L S V G P F K L L L A S D Q Q P S S A T D L D

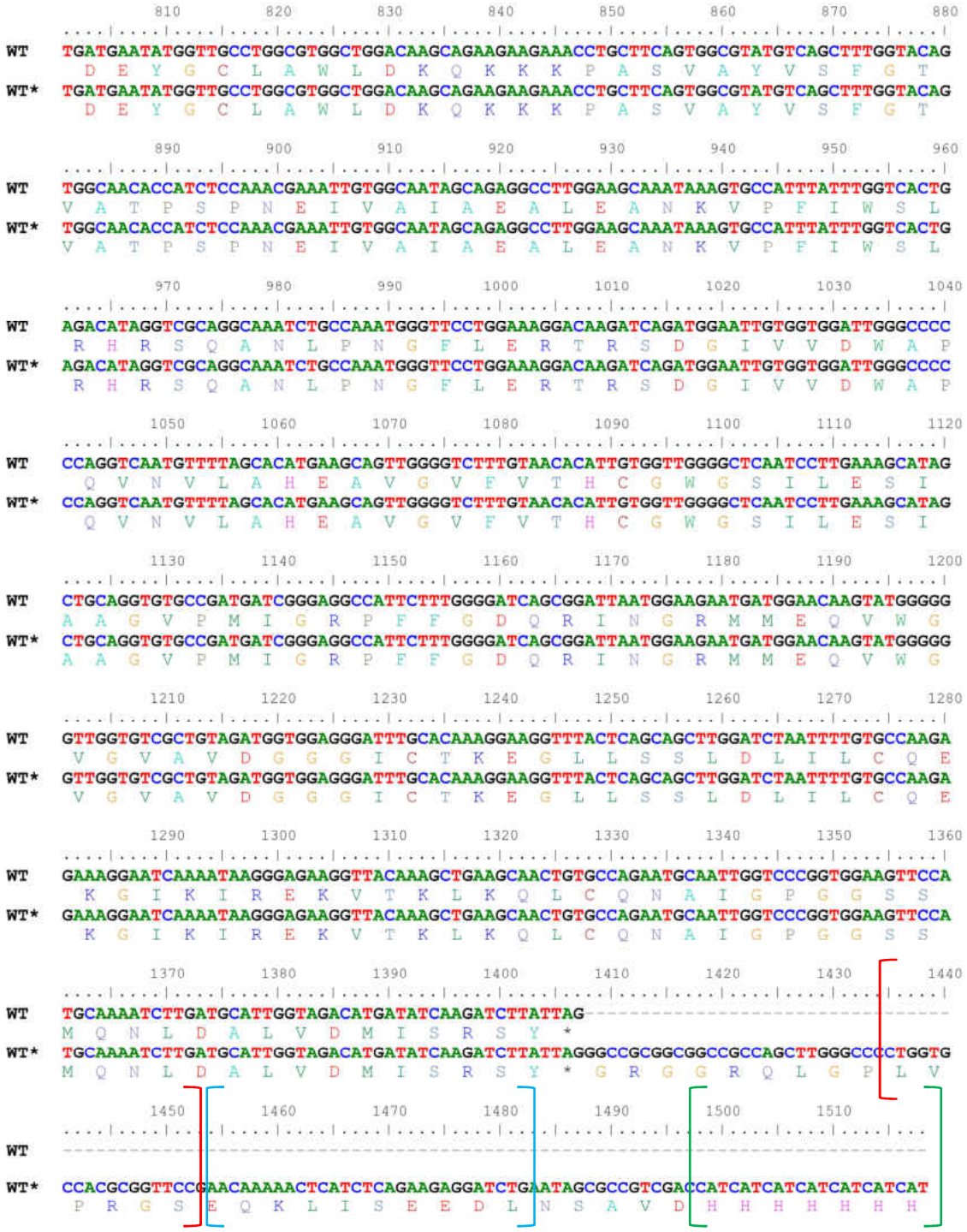
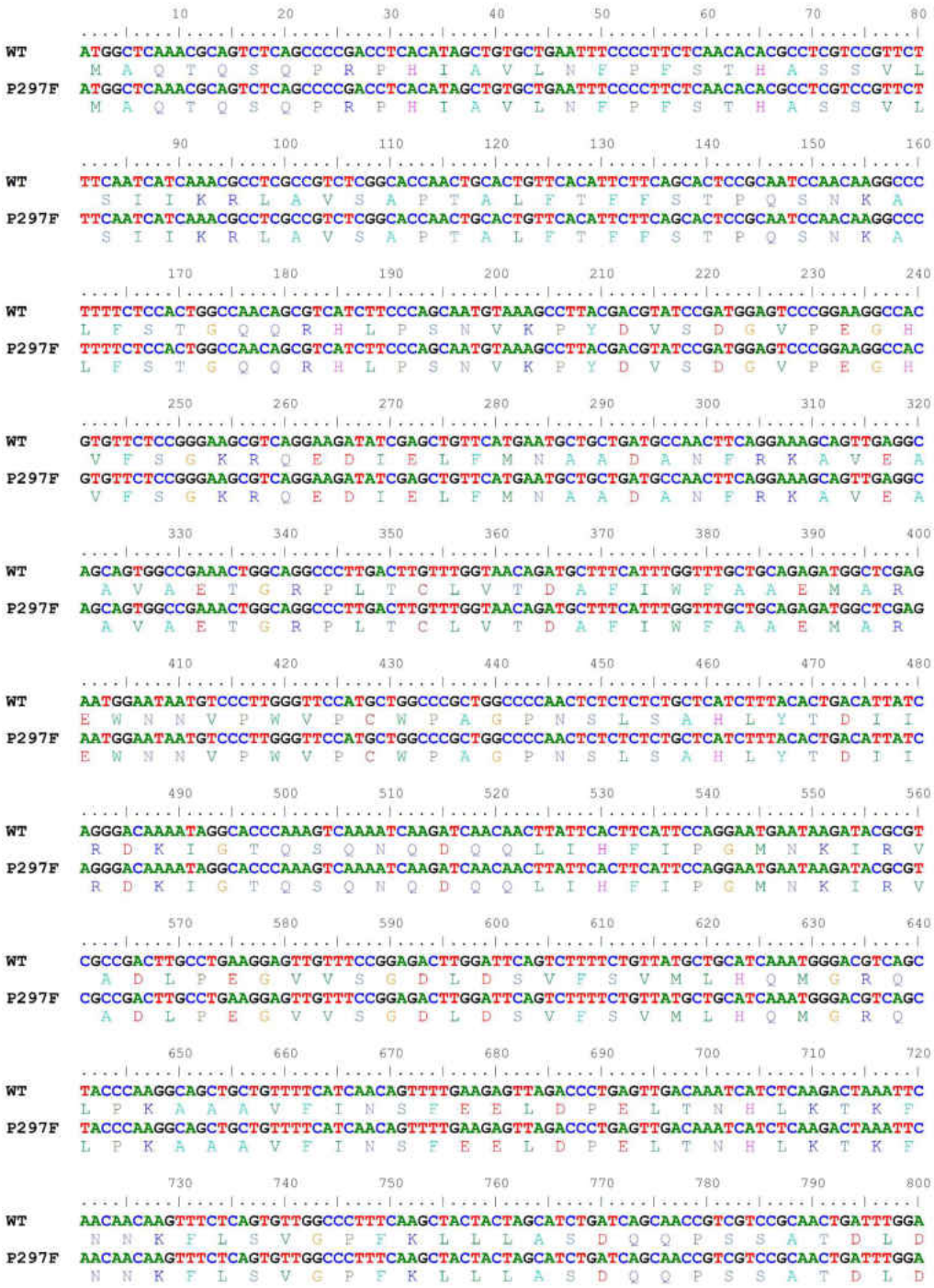


Figure 2.1: Sequence alignment of WT CP3GT and recombinant WT CP3GT (asterisk) showing thrombin-cleavage site (red bracket), c-myc tag (blue bracket), and 6x-his tag (green bracket).

The P297F mutant was previously cloned in the pPICZα plasmid using site directed mutagenesis (Adepoju 2014). This single amino acid mutation resulted in complete loss of GT activity. After site directed mutagenesis was performed to insert the thrombin cleavage sequence, the rP297F mutant plasmid (asterisk) was sequenced and aligned with the WT CP3GT sequence to verify that the mutation was present as well as the thrombin cleavage site and that everything was in frame. In Figure 2.2, the thrombin-cleavage site, c-myc tag, and 6x His tag are all visible and are bracketed as in Figure 2.1. In both WT and P297F CP3GT the sequences are all in frame.


10 20 30 40 50 60 70 80
WT ATGGCTCAAACGCAGTCTCAGCCCCGACCTCACATAGCTGTGCTGAATTTCCCCTTCTCAACACACGCCTCGTCCGTTCT
M A Q T Q S Q P R P H I A V L N F P F S T H A S S V L
P297F ATGGCTCAAACGCAGTCTCAGCCCCGACCTCACATAGCTGTGCTGAATTTCCCCTTCTCAACACACGCCTCGTCCGTTCT
M A Q T Q S Q P R P H I A V L N F P F S T H A S S V L
90 100 110 120 130 140 150 160
WT TTCAATCATCAAACGCCTCGCCGTCTCGGCACCAACTGCACTGTTCACATTCTTCAGCACTCCGCAATCCAACAAGGCCC
S I I K R L A V S A P T A L F T F F S T P Q S N K A
P297F TTCAATCATCAAACGCCTCGCCGTCTCGGCACCAACTGCACTGTTCACATTCTTCAGCACTCCGCAATCCAACAAGGCCC
S I I K R L A V S A P T A L F T F F S T P Q S N K A
170 180 190 200 210 220 230 240
WT TTTTCTCCACTGGCCAACAGCGTCATCTTCCCAGCAATGTAAAGCCTTACGACGTATCCGATGGAGTCCCGGAAGGCCAC
L F S T G Q Q R H L P S N V K P Y D V S D G V P E G H
P297F TTTTCTCCACTGGCCAACAGCGTCATCTTCCCAGCAATGTAAAGCCTTACGACGTATCCGATGGAGTCCCGGAAGGCCAC
L F S T G Q Q R H L P S N V K P Y D V S D G V P E G H
250 260 270 280 290 300 310 320
WT GTGTTCTCCGGGAAGCGTCAGGAAGATATCGAGCTGTTCATGAATGCTGCTGATGCCAACTTCAGGAAAGCAGTTGAGGC
V F S G K R Q E D I E L F M N A A D A N F R K A V E A
P297F GTGTTCTCCGGGAAGCGTCAGGAAGATATCGAGCTGTTCATGAATGCTGCTGATGCCAACTTCAGGAAAGCAGTTGAGGC
V F S G K R Q E D I E L F M N A A D A N F R K A V E A
330 340 350 360 370 380 390 400
WT AGCAGTGGCCGAAACTGGCAGGCCCTTGACTTGTTTGGTAACAGATGCTTTCATTTGGTTTGCTGCAGAGATGGCTCGAG
A V A E T G R P L T C L V T D A F I W F A A E M A R
P297F AGCAGTGGCCGAAACTGGCAGGCCCTTGACTTGTTTGGTAACAGATGCTTTCATTTGGTTTGCTGCAGAGATGGCTCGAG
A V A E T G R P L T C L V T D A F I W F A A E M A R
410 420 430 440 450 460 470 480
WT AATGGAATAATGTCCCTTGGGTTCCATGCTGGCCCGCTGGCCCCAACTCTCTCTCTGCTCATCTTTACACTGACATTATC
E W N N V P W V P C W P A G P N S L S A H L Y T D I I
P297F AATGGAATAATGTCCCTTGGGTTCCATGCTGGCCCGCTGGCCCCAACTCTCTCTCTGCTCATCTTTACACTGACATTATC
E W N N V P W V P C W P A G P N S L S A H L Y T D I I
490 500 510 520 530 540 550 560
WT AGGGACAAAATAGGCACCCAAAGTCAAAATCAAGATCAACAACTTATTCACTTCATTCCAGGAATGAATAAGATACGCGT
R D K I G T Q S Q N Q D Q Q L I H F I P G M N K I R V
P297F AGGGACAAAATAGGCACCCAAAGTCAAAATCAAGATCAACAACTTATTCACTTCATTCCAGGAATGAATAAGATACGCGT
R D K I G T Q S Q N Q D Q Q L I H F I P G M N K I R V
570 580 590 600 610 620 630 640
WT CGCCGACTTGCCTGAAGGAGTTGTTTCCGGAGACTTGGATTCAGTCTTTTCTGTTATGCTGCATCAAATGGGACGTCAGC
A D L P E G V V S G D L D S V F S V M L H Q M G R Q
P297F CGCCGACTTGCCTGAAGGAGTTGTTTCCGGAGACTTGGATTCAGTCTTTTCTGTTATGCTGCATCAAATGGGACGTCAGC
A D L P E G V V S G D L D S V F S V M L H Q M G R Q
650 660 670 680 690 700 710 720
WT TACCCAAGGCAGCTGCTGTTTTCATCAACAGTTTTGAAGAGTTAGACCCTGAGTTGACAAATCATCTCAAGACTAAATTC
L P K A A A V F I N S F E E L D P E L T N H L K T K F
P297F TACCCAAGGCAGCTGCTGTTTTCATCAACAGTTTTGAAGAGTTAGACCCTGAGTTGACAAATCATCTCAAGACTAAATTC
L P K A A A V F I N S F E E L D P E L T N H L K T K F
730 740 750 760 770 780 790 800
WT AACAACAAGTTTCTCAGTGTTGGCCCTTTCAAGCTACTACTAGCATCTGATCAGCAACCGTCGTCCGCAACTGATTTGGA
N N K F L S V G P F K L L L A S D Q Q P S S A T D L D
P297F AACAACAAGTTTCTCAGTGTTGGCCCTTTCAAGCTACTACTAGCATCTGATCAGCAACCGTCGTCCGCAACTGATTTGGA
N N K F L S V G P F K L L L A S D Q Q P S S A T D L D

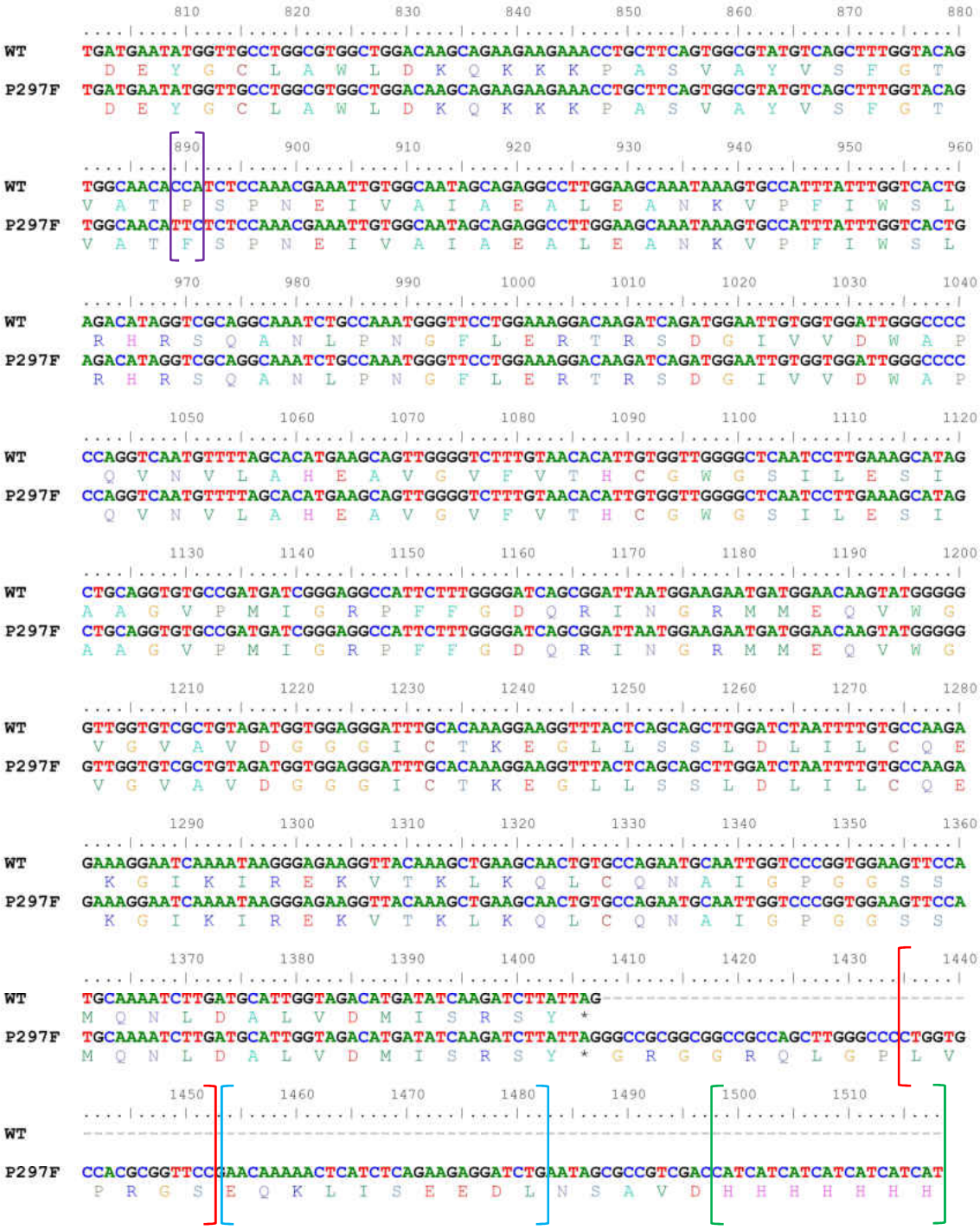


Figure 2.2: Sequence alignment of P297F CP3GT and recombinant WT CP3GT showing residue mutation (purple bracket), thrombin-cleavage site (red bracket), c-myc tag (blue bracket), and 6x-his tag (green bracket).

2.2 Transformation of rCP3GT and P297F into *Pichia pastoris*

Once the presence of all necessary tags was confirmed, WT CP3GT plasmid was extracted from *E. coli*, prepared, and rCP3GT transformed into *Pichia pastoris* as described. Before inducing protein expression on the newly transformed cells, a direct screen was performed using colony PCR to confirm transformation. Single colonies were picked and re-streaked on 8 YPD plates containing zeocin. From each of these, a single colony was selected for direct screening. The results of the direct screen can be seen in Figure 2.3. Lanes 6 and 7 show the gene of interest at the anticipated size of 1.7kbp.

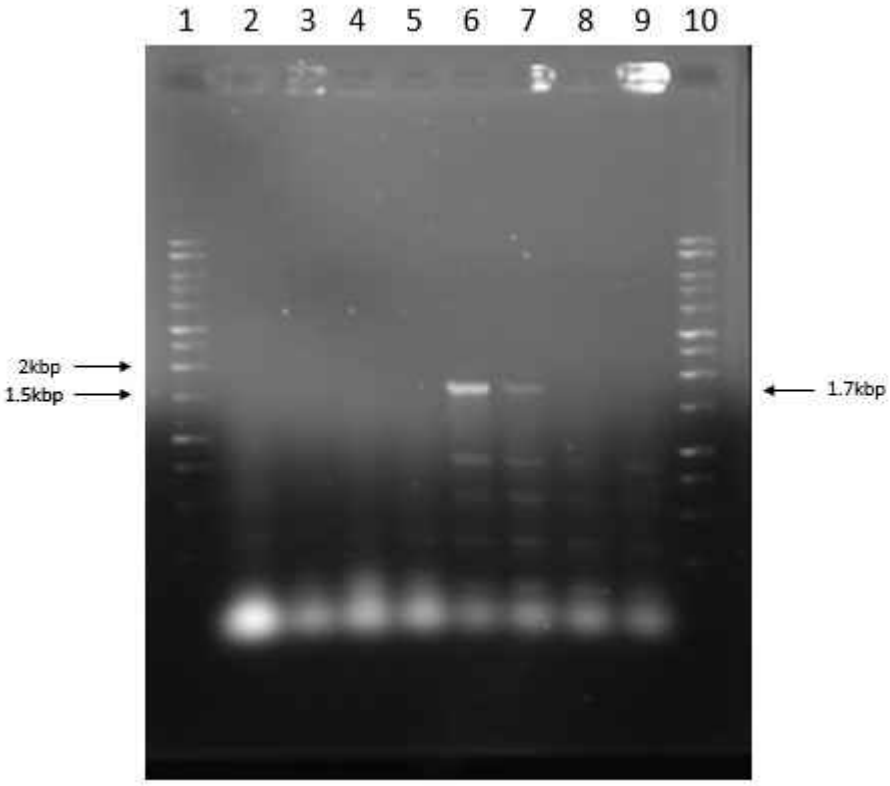


Figure 2.3: Direct screen of rCP3GT colonies showing rCP3GT insert at the anticipated size of 1.7kbp in samples in lanes 6 and 7. Lanes 1 and 10 contain DNA ladder.

Single colonies were selected from the plates that corresponded with lanes 6 and 7 from Figure 2.3 and were streaked on 4 new YPD plates with zeocin. Single colonies were taken from the re-streaked plates and were used to conduct another colony PCR. All the colonies selected had the gene insert at the

anticipated size (Fig. 2.4). Colonies from each of the plates that were verified to have the insert were cultured in YPD media and stored at -80º C in glycerol.

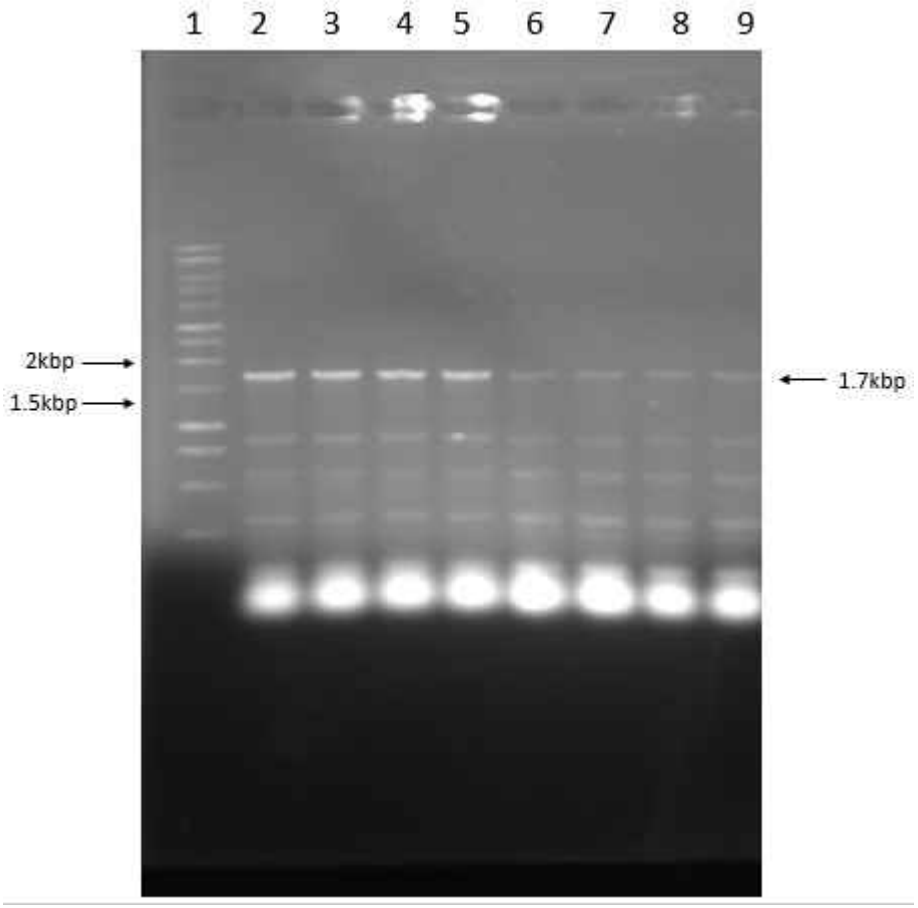


Figure 2.4: Colony PCR of colonies taken from restreaks that had the insert as visualized in Figure 2.3. Lanes 2-5 were loaded with PCR reactions made from colonies taken from 4 separate restreaks of the colony that had insert in lane 6 from Figure 2.3. Lanes 6-9 were loaded with PCR reactions made from colonies taken from 4 separate restreaks of the colony that had insert in lane 7 from Figure 2.3. Lane 1 is the DNA ladder.

2.3 Protein Expression and Purification

rCP3GT protein was induced in yeast and purified using cobalt column chromatography as described. Fractions were collected at various stages of the purification process to verify proper functioning of the purification system and were visualized using SDS-PAGE and western blot (Fig. 2.5 and 2.6). The flow through fraction contained proteins that do not interact with the IMAC column. This

first fraction contained nonspecific proteins left behind after the desalted lysate was passed through the His-Pure column. The second wash fraction contained little to no protein because most of the nonspecific proteins were washed off in the first wash step.

Initial trials showed that eluting CP3GT protein from the column in one step as suggested in the manufacturers handbook caused nonspecific proteins to contaminate the final sample (McIntosh, personal communication). Multiple wash steps were not enough to mitigate this problem, thus a low astringency, 10mM imidazole wash was employed to further clean the column of nonspecific proteins. As can be seen in Figure 2.5, nonspecific proteins were present in the 10mM eluate, but much less so in the final 150mM eluate, indicating that the low astringency wash was mostly successful in removing nonspecifically bound proteins from the column. Imidazole will also remove weakly bound rCP3GT if the binding capacity of the His-Pure resin has been exceeded, thus there is a small amount present in the 10mM fraction, as can be seen from the faint band in Figure 2.6. Nonspecifically binding proteins were also present in the 150mM eluate. In previous work, this was mitigated by using more 10mM imidazole buffer until OD_{280} reached 0. Subsequent purifications will employ a more thorough low-astringency 10mM wash until the column is completely free of nonspecifically bound proteins (Adepoju 2014; Sathanantham 2015; Kandel 2016).

rCP3GT protein (51.2 kDa) was detected in the crude and desalt fractions, but not in the flow through or in either of the 2 wash fractions (Fig. 2.6). The Coomassie stain indicates that non-specific proteins were removed from the columns during the flow through, wash, and low stringency elution steps (Fig. 2.5). The protein was not 95% pure, however, because only 5-10mL of 10mM imidazole buffer was used when washing away non-specific proteins. In future purifications, non-specific proteins will be eluted using 10mM imidazole until OD_{280} reaches 0, indicating that all weakly bound proteins have been washed away.

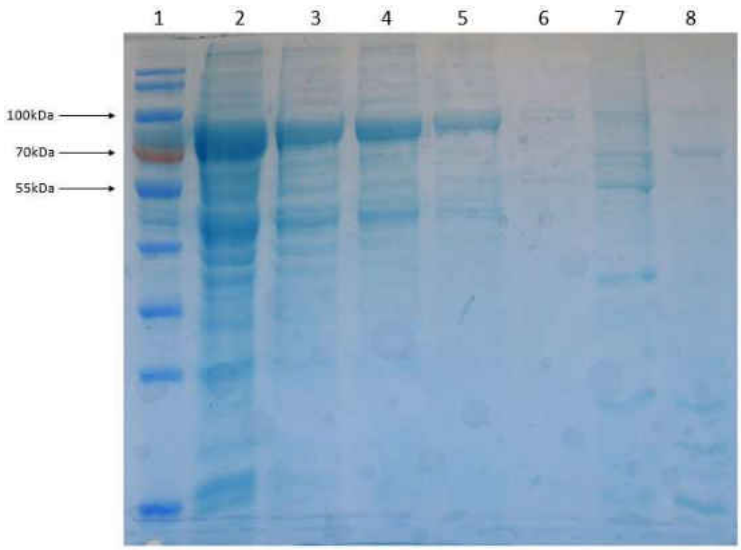


Figure 2.5: SDS-PAGE gel of different fractions from protein purification. Lane 1 contains a 10-180 kDa ladder. Lane 2 contains crude rCP3GT lysate. Lane 3 contains the desalted protein fraction. Lane 4 contains the flow through fraction. Lane 5 contains the first wash fraction and lane 6 the second wash fraction. Lane 7 contains the pooled 10mM eluate fractions. Lane 8 contains the pooled 150mM eluate fractions.

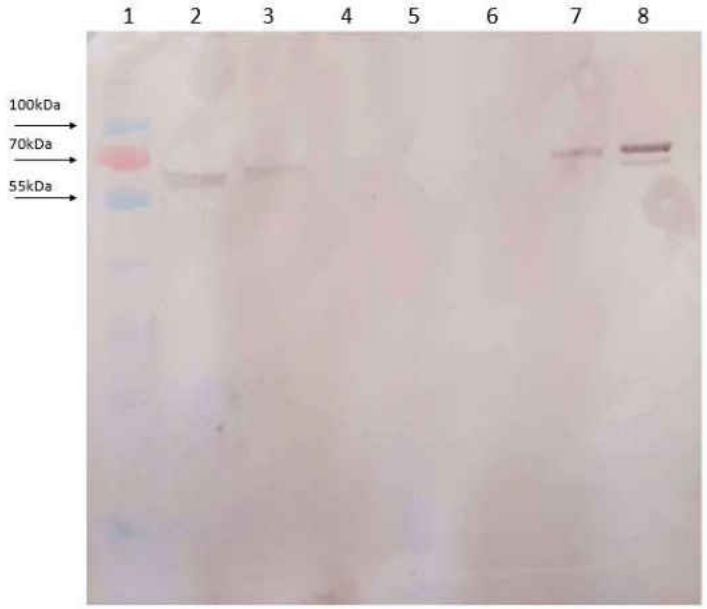


Figure 2.6: Western blot showing presence of rCP3GT protein in crude, desalt, and eluate. Lane 1 contains a 10-180 kDa ladder. Lane 2 contains crude rCP3GT lysate. Lane 3 contains the desalted protein fraction. Lane 4 contains the flow through fraction. Lane 5 contains the first wash fraction and lane 6 the second wash fraction. Lane 7 contains the pooled 10mM eluate fractions. Lane 8 contains the pooled 150mM eluate fractions.

2.4 CP3GT Activity Assay

While thrombin digestion conditions were being optimized, a time-course GT activity assay was carried out to determine conditions for which CP3GT activity is linear in preparation for future kinetic studies. This allows us to see what concentration of pure protein is needed to achieve a linear reaction and for how long the reaction is linear. These data allow us to standardize future kinetic assays to assure we are testing at initial velocity. The assay uses a radiolabeled UDP-glucose donor substrate and a quercetin acceptor substrate (McIntosh et al 1990). Quercetin is used because previous reactions showed that CP3GT has the highest affinity for this flavonol (Owens and McIntosh, 2009; Devaiah et al 2016). Initial results showed that linearity occurs for up to 15 minutes using 2ug of pure protein. All subsequent reactions were run using 2µg of pure CP3GT protein for 10 minutes at 37° C.

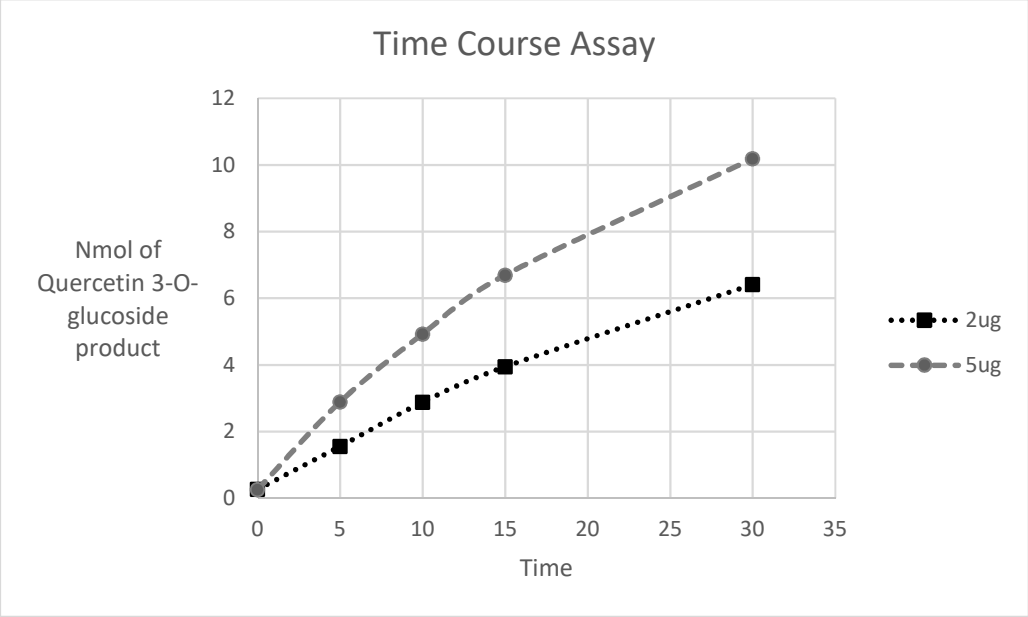


Figure 2.7: Time course assay of CP3GT using 2 and 5ug of pure protein at 37° C.

Assays were performed using freshly purified CP3GT enzyme and CP3GT enzyme that was stored for 2 hours at 4° C following purification. Data showed that subjecting pure CP3GT to these storage conditions resulted in an approximately 70 percent retention in activity (Fig 2.8). These conditions for storage were selected based upon the incubation conditions used to cleave tags with thrombin. It was

necessary to determine what impact thrombin incubation conditions had on CP3GT activity because maintaining CP3GT activity after removing tags is important for crystallization of active, native protein. Data showed that CP3GT enzyme can be subjected to short-term cold storage without losing significant activity.

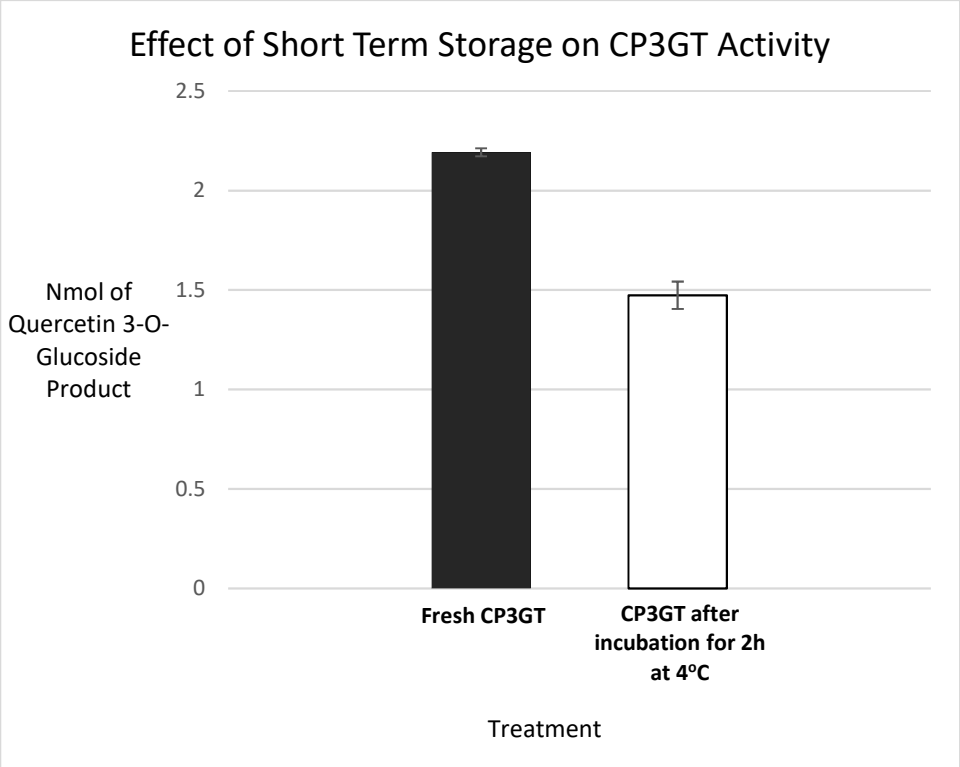


Figure 2.8: Effect of short term storage on CP3GT activity after storage for 2 hours at 4°C, 70% activity is retained. Data are mean ± standard deviation, N = 3

2.5 Thrombin Digestion

Once the CP3GT protein was purified, it was digested with thrombin for 2 hours using increasing concentrations of thrombin and at temperatures of 4° C and 10° C. At 4° C, 100U of thrombin was sufficient to cleave tags off 30µg of protein (Fig. 2.9). When no digestion occurred, the c-myc tag remained and a positive result on the western blot was detected, such as in lane 3 where 0U of thrombin

was used. As the concentration of thrombin was increased, more of the c-myc tag was cleaved causing less of the protein to be detected via western blot. The same of amount of protein was loaded into each well, therefore the absence of a band does not indicate that no protein is present, rather it indicates that protein is simply not detectable due to the removing of the c-myc tag and 6x-His tag.

Figure 2.9: Digestion of rCP3GT using increasing concentrations of thrombin at 4° C for 2 hours. Lane 1 contains a 10-180 kDa ladder. Lane 2 contains 30µg of pure rCP3GT control. Lane 3-8 contain 30µg of purified rCP3GT protein and increasing concentrations of thrombin in the amounts of 0, 3, 6, 30, 60, and 100 units, respectively.

This step was crucial to our research goals because it was hypothesized that digestion with thrombin would result in minimal loss of CP3GT activity. Before an activity assay could be conducted to test this hypothesis, digestion conditions had to be optimized. The data shown here allowed us to carry out activity assays on pure CP3GT with and without tags. Additionally, this step is important because our research has the long-term goal of crystallization, and all previously crystallized plant GT's had tags

removed prior to crystallizing. Therefore, optimization of tag removal was critical to both the short-term goal of testing the effects of tags on activity and the long-term goal of crystallizing CP3GT.

2.6: Verification of Thrombin Digestion

Conditions for digesting CP3GT with thrombin were validated following the activity assay of CP3GT with and without tags. Aliquots of rCP3GT were digested with 0, 1.65, 3.3, and 6.6 Units of thrombin per microgram of rCP3GT, respectively, for 2 hours at 4° C using 50µg of rCP3GT. The digestion was stopped with 10mM PMSF and the products were visualized using SDS PAGE and western blot (Figure 2.11 and 2.12). Each lane contained 20µL of the digested products (approximately 20µg of rCP3GT). On the Coomassie stain, lanes 2-5 show the increasing concentration of thrombin indicated by the thick bands at approximately 35 kDa. Lane 6 and 10 were left empty but leakage from lanes 7, 8, and 9 contaminated these wells. Lane 7 contained 20µg of undigested rCP3GT and lane 8 contained 20µg rCP3GT and 3.3U/µg thrombin. Lane 9 contained a CP3GT control.

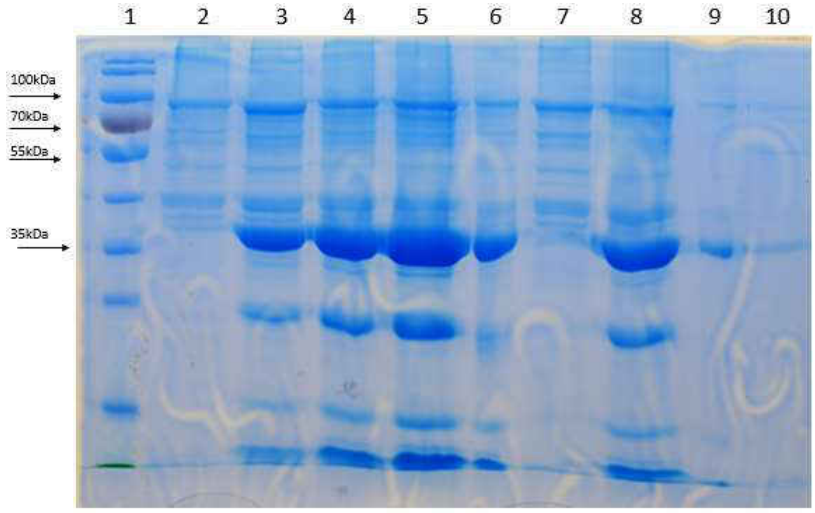


Figure 2.10: Coomassie stain of thrombin digest. Lane 1 contains a 10-180 kDa ladder. Lanes 2-5 contain 50µg rCP3GT digested with 0, 1.65, 3.3, and 6.6 Units thrombin/µg rCP3GT, respectively. Lanes 6 and 9 were left empty. Lane 7 contained 50µg untreated rCP3GT. Lane 8 contained 50µg rCP3GT treated with 3.3U/µg. Lane 10 contained a CP3GT control.

The western blot shows complete digestion of rCP3GT at thrombin concentrations of 1.65, 3.3, and 6.6 U/µg, indicated by the lack of bands in lanes 3-5. As can be seen in the Coomassie stain, protein is still present in these lanes, but the cleavage of the c-myc tag by thrombin causes no band to appear on the western blot for digested products. The faint bands in lane 6 and 8 indicate that undigested protein from lane 7 leaked into these lanes. It is still clear, however, that rCP3GT loaded in lane 8 was digested. The appearance of two bands in lane 2 and 7 is likely caused by protein denaturing which can occur when purified rCP3GT is stored at 4° C for longer than a few days in the absence of glycerol (McIntosh, personal communication). The CP3GT control in lane 10 was 95-99% pure and was stored in glycerol.

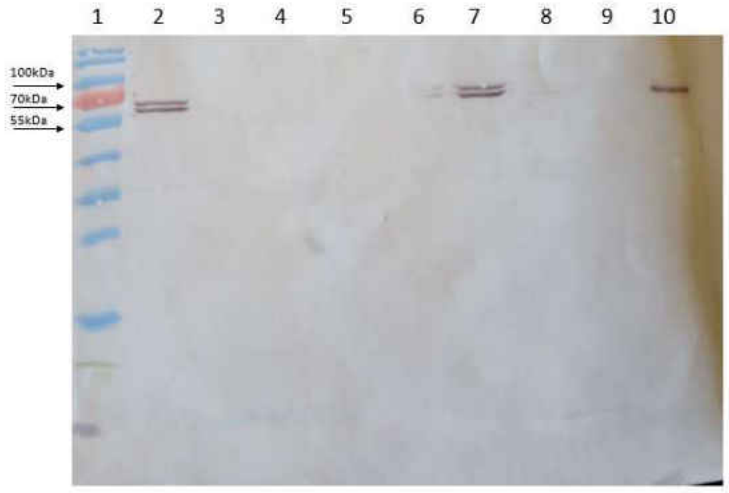


Figure 2.11: Western blot of thrombin digest. Lane 1 contains a 10-180 kDa ladder. Lanes 2-5 contain 50µg rCP3GT digested with 0, 1.65, 3.3, and 6.6 Units thrombin/µg rCP3GT, respectively. Lanes 6 and 9 were left empty. Lane 7 contained 50µg untreated rCP3GT. Lane 8 contained 50µg rCP3GT treated with 3.3U/µg. Lane 10 contained a CP3GT control.

2.7: Impact of thrombin on enzyme activity

Activity assays were carried out on samples of purified CP3GT that had been digested with 0, 1.65, 3.3, and 6.6 units of thrombin per microgram of rCP3GT, respectively for 2 hours at 4°C. All samples were brought to the same volume using thrombin buffer and kept under the same conditions. The data showed no significant difference in activity between untreated rCP3GT and rCP3GT treated with

thrombin (Fig. 2.10). Additionally, increasing the concentration of thrombin to twice the amount needed did not have a significant effect on CP3GT activity. From this data it was concluded that thrombin digestion has no deleterious effects on CP3GT enzyme activity. Using 3.3U per microgram of rCP3GT resulted in complete cleavage of tags under these conditions (Figure 2.11 and 2.12)

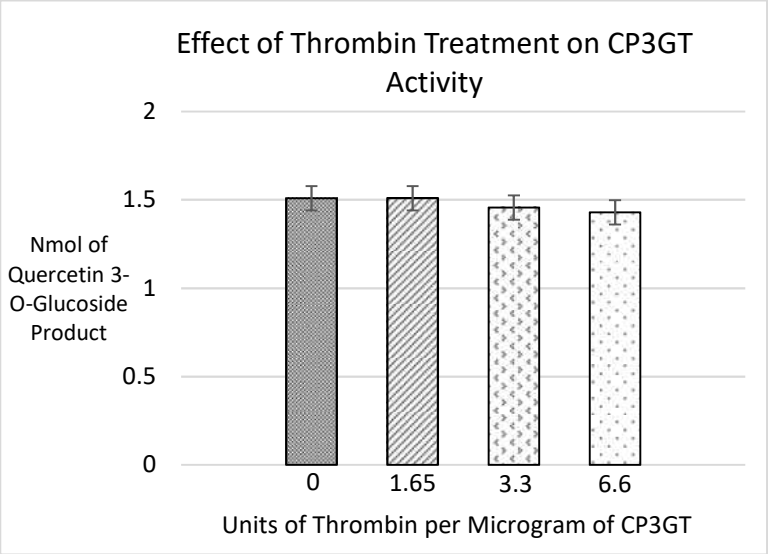


Figure 2.12: Effect of thrombin treatment on CP3GT activity. Data are mean ± standard deviation, N = 3

3. Conclusions

This work has expanded upon CP3GT characterization studies and has paved the way for more robust exploration into the structural basis for flavonol specificity in this grapefruit GT. Previous studies in our lab used homology modeling and site directed mutagenesis to design and biochemically characterize 30 CP3GT mutants, in addition to using *in situ* docking analysis to predict how these mutants interact with their acceptor/donor substrates. Biochemical assays were then performed to test hypotheses generated from *in silico* analyses. These methods are often the first-line approach for enzyme characterization, and have provided a simple, reliable way to explore CP3GT's unique properties. This enzyme is well characterized, yet the structural basis for its flavonol specificity is still not known.

Current methods are not sufficient to answer this question and our work has shifted towards preparing CP3GT for crystallization. Additionally, the effect of tags on enzyme activity was tested. To the end of recognizing both these goals, a thrombin cleavage site was inserted into WT CP3GT by Kandel in 2016 and was in this work transformed into *Pichia pastoris*. A thrombin cleavage site was inserted into the P297F mutant and will be transformed into yeast in the near future.

Reliable conditions for WT CP3GT expression and purification were identified and the enzyme was tested for activity. Purification trials did not give rCP3GT that was 95-99% pure due to inadequate elution of nonspecific proteins. In these purifications, only 5-10mL of 10mM imidazole buffer was used which proved to be insufficient as nonspecific proteins remained on the column. Future purifications will use 10mM elution buffer until eluted fractions show zero absorbance when analyzed at OD_{280}. Additionally, optimal conditions for thrombin digestion were identified. Short term storage of CP3GT for 2 hours at 4^{o}C resulted in an approximately 70% retention of activity. Thrombin treatment using 3.3U/µg of rCP3GT for 2 hours at 4^{o}C was sufficient to cleave tags. Neither the 3.3U/µg digestion nor double that concentration had any significant effect on CP3GT activity.

Future studies should focus on inserting and verifying thrombin cleavage site in all mutants of interest and transforming all into yeast. WT CP3GT and all mutants should be digested with thrombin and screened for crystallization conditions. Additional work should be done to test the effect of tags on CP3GT activity. Kinetics assays and assays with different substrates should be performed. Crystallization trials should be conducted following screening to generate CP3GT and mutant crystals for x-ray diffraction. CP3GT should also be co-crystallized with substrate to further enhance analysis. Decoding and analysis of diffraction patterns should be conducted to determine the 3D structure of CP3GT.

4. Experimental

4.1 Reagents

Quercetin, kaempferol, and their corresponding 3-O glucosides were from Sigma (St. Louis, MO, USA). Gene JET mini-prep kit was from Thermo Scientific. Wizard Plus Midipreps kit, SacI and DpnI restriction enzymes were from Promega (Madison, WI). Bovine thrombin was from MP Biomedicals. All other reagents were from Fischer Scientific.

4.2 Expression of WT rCP3GT

Two YPD agar plates containing 100mg/mL zeocin were streaked with yeast cell stocks stored at -80ºC that contained wild type Cp3GT with thrombin cleavage tags inserted. The plates were incubated for 24-48 hours at 28ºC. Two 5mL aliquots of BMGY in 20mL glass test tubes were inoculated from a single colony picked from the incubated YPD plates. The tubes were placed in an incubator-shaker at an angle to allow adequate aeration and were incubated at 30C and 250rpm until an OD_{600} gave a reading of 2, approximately 18-20 hours. From each of these cultures, 200uL was taken to inoculate two, 50mL aliquots of BMGY in 500mL baffled flasks, which were incubated as before to an OD_{600} of 2. The cells were then centrifuged for 5 minutes at 3000 x *g* at room temperature. The cells were resuspended in BMMY to wash away any remaining glycerol and centrifuged again for 5 minutes at 3000 x *g* at room temperature. The cells were resuspended in an amount of fresh BMMY to final OD_{600} reading of 1. One 50mL culture of BMGY grown cells required approximately 500mL of BMMY to achieve an OD_{600} reading of 1. Two and a half milliliters of methanol was added to make the final concentration of the BMMY media 0.5% and the cells were divided equally into two, 1L sterile baffled flasks. The cells were incubated at 30C and 250rpm for 18 hours as determined to be the optimum time for protein expression. After growing, the cultured cells were aliquoted into 50mL tubes and the tubes were centrifuged at 3-

5000 x *g* for 3-5 minutes. The supernatant was discarded, and the pelleted cells were stored at -80C for later use or used immediately for protein extraction, purification, and assay.

4.3 Protein Extraction and Purification

The frozen or fresh cells were thawed or placed on ice before being resuspended in 20mL of cold 50mM phosphate, pH 7.5, buffer that contained 5mM BME. The cells were centrifuged at 3000 x g for 5 minutes at 4^{o} C and then resuspended in 10mL of fresh buffer. The cells were lysed using a French press stored at 4^{o}C at 1250psi for 4 cycles. Samples were kept on ice for no more than 1 minute between cycles. PD-10 desalting columns were prepared by equilibrating each with 25mL equilibration buffer prior to desalting proteins. The first fraction was collected after rCP3GT crude lysate was passed through a PD-10 desalting gravity column, called the desalted fraction. Approximately 3mL of cell lysate was passed through each PD-10 column. Desalted protein was eluted from each column by adding 4-5mL of equilibration buffer. The flow through from each column was collected and pooled. An IMAC column was prepared using 2.5mL bed volume of HisPur cobalt resin and was equilibrated with three resin bed volumes of equilibration buffer. The next fraction, called the flow through fraction, was collected after desalted rCP3GT lysate was passed through a His-Pure Cobalt Resin gravity column. The desalted cell lysate was passed through the column and the flow through was collected in 2mL fractions. The next fractions collected were the wash fractions, which were collected in two stages. The first wash fraction was collected after 10mL of equilibration buffer was passed through the His-Pure column. The second wash fraction was collected after an additional 10mL of equilibration buffer was passed through the His-Pure column. This step was repeated until the wash fractions showed zero absorbance at 280nm using a nanodrop spectrophotometer. A low astringency wash was performed using 10mM imidazole to remove nonspecifically bound proteins. To the column was added 10mL of 10mM imidazole and was collected in 2mL fractions. Pure CP3GT protein was eluted by adding three resin bed volumes of elution buffer with 150mM imidazole. The eluate was collected in 2mL fractions. To remove imidazole from the protein

elution, the fractions showing the greatest absorbance were pooled into an Amicon 30 MWCO filter and centrifuged at 4C until the volume reached 500uL (about 15 minutes). Two mL of assay buffer or crystallization buffer was added and again centrifuged until the volume reached 500uL (about 5 minutes). Again, 2mL of assay or crystallization buffer was added and centrifuged until the volume reached 500uL. Aliquots from all purification fractions (desalt, flow through, wash, and elution) were collected for analysis by SDS-PAGE and western blot.

Scale up of protein purification was attempted using 100-200mL of culture and 7-10mL resin beds, but the extended time necessary to elute from such large columns resulted in loss of enzyme activity. Therefore, multiple 2.5mL resin beds were used concurrently to save time and preserve enzyme activity.

4.4 Thrombin Digestion

A 5kU stock of bovine thrombin was obtained from MP Biomedicals (CAS Number 9002-04-4) and diluted to a concentration of 25U/µL. The diluted stock was divided into 20µL aliquots and stored at -80° C. Optimal conditions for thrombin digestion were determined by incubating 30µg of purified protein with increasing concentrations of thrombin enzyme at 4° C for 2 hours. The reaction was stopped by adding PMSF to a final concentration of 10mM. It was determined by SDS-PAGE and western blot that 100U of bovine thrombin digested 30µg of pure protein in 2 hours at 4° C. Digestions were also carried out at 10° C, but the data showed no benefit to digesting at a higher temperature. Therefore, all thrombin digestions were carried out at 4° C using 3.3 units of thrombin per microgram of rCP3GT protein.

4.5 CP3GT Activity Assay

Glucosyltransferase assays were performed to determine the amount of reaction product catalyzed by the purified enzyme. The first assay was designed to determine the amount of enzyme necessary to achieve a linear reaction rate and for how long this rate lasted before plateauing. A 75uL reaction was prepared by mixing 5uL of quercetin aglycone (50nmol/5uL ethylene glycol monomethyl ether), 10uL of 100nmoles UDP-^{14}C glucose containing 40,000cpm, either 2 or 5ug of purified CP3GT enzyme (added upon starting), and enough 50mM phosphate buffer (pH 7.5 containing 14mM BME) to make a total reaction volume of 75uL. A total of 4 reaction vessels were prepared. The reaction vessel was scaled up such that 4, 75uL aliquots could be taken at time points 0, 5, 10, 15, 30 minutes after the reaction was started. The reaction was incubated at 37° C and proceeded for 30 minutes. A 75 µL aliquot was removed at time points 0, 5, 10, 15, and 30 minutes and the reaction was stopped by adding 15uL of 6M HCl to each aliquot. To each aliquot was added 250uL of ethyl acetate followed by vortexing for 20 seconds. Brief centrifugation was used to separate the organic and aqueous layers. A 125uL aliquot of the organic layer was carefully pipetted off and placed into a vial with 2mL of CytoScint scintillation fluid (Thermo Fisher). Incorporation of the reaction product was measured using a Beckman LS 6500 scintillation counter. The counts obtained were multiplied by two because only half of the reaction product was pipetted into the scintillation fluid. The amount (nmoles) of quercetin 3-O-glucoside produced was calculated by dividing the cpm incorporated by 400cpm/nmole.

Assays were performed on CP3GT treated and untreated with thrombin. A 75µL reaction was prepared as previously described for 4 separate, 2µg aliquots of CP3GT that had been treated 0, 1.65, 3.3, and 6.6U/µg respectively. The reaction was incubated at 37°C for 10 minutes. Each reaction was stopped using 15µL of 6M HCl. Extraction of incorporated product using ethyl acetate was carried out as before and counts were measured using a Perkin Elmer TriCarb 2810 TR.

REFERENCES

Adepoju, O. A., 2014. Using site-directed mutagenesis to determine Impact of amino acid substitution on substrate and regiospecificity of grapefruit flavonol 3-O-glucosyltransferase. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. ETSU School of Graduate Studies, Johnson City, Tennessee, USA.

Agati, G., Matteini, P., Goti, A., Tattini, M., 2007. Chloroplast-located flavonoids can scavenge singlet oxygen. New Phytol. 174, 77-89.

Berhow, M., Tisserat, B., Kanes, K., Vandercook, C., 1998. Survey of phenolic compounds produced in citrus. USDA.

Brazier-Hicks, M., Offen, W. A., Gershater, M. C., Revett, T. J., Lim, E., Bowles, D. J., Davies, G. J., Edwards, R., 2007. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. PNAS, 20238-20243.

Chandrika, B. B., Steephan, M., Kumar, T. R., Sabu, A., Haridas, M., 2016. Hesperetin and Naringenin sensitize HER2 positive cancer cells to death by serving as HER2 Tyrosine Kinase inhibitors. Life Science 160, 47-56.

Coutinho, P. M., Deleury, E., Davies, G. J., Henrissay, B., 2003. An Evolving Hierarchical Family Classification for Glycosyltransferases. J. Mol. Biol. 328, 307-317.

D'Arcy, A., Elmore, C., Stihle, M., Johnston, J. E., 1996. A novel approach to crystallising proteins under oil. J. Cryst. Growth 168, 175-180.

Dessau, M. A., Modis, Y., 2011. Protein crystallization for x-ray crystallography. Journal of Visualized Experiments 47, 2285.

Devaiah, S. P., Owens, D. K., Sibhatu, M. B., Sarkar, T. R., Strong, C. L., Mallampalli, V. K. P. S., Asiago, J., Cooke, J., Kiser, S., Lin, Z., Wamucho, A., Hayford, D., McIntosh, C. A., 2016. Identification, Recombinant Expression, and Biochemical Analysis of Putative Secondary Product Glucosyltransferases from Citrus Paradisi. J. Agric. Food Chem., 1957-1969.

Devaiah, S. P., Tolliver, B. M., Zhang, C., Owens, D. K., McIntosh, C. A., 2018. Mutational analysis of substrate specificity in a Citrus paradisi flavonol-3-O-glucosyltransferase. J. Plant Biochem. Biotechnol. 27, 13-17.

Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., Helm, D. v. d., Deisenhofer, J., 2002. Structural Basis of Gating by the Outer Membrane Transporter FecA. Science 295, 1715-1719.

Figueiredo, C. A., Barroso, J. G., Pedro, L. G., Scheffer, J. J., 2008. Factors affecting secondary metabolite production in plants: volatile components and essential oils. Flavour Fragrance J. 23, 213-226.

Gulcin, I., 2012. Antioxidant activity of food constituents: an overview. Arch. Toxicol. 86, 345-391.

Heller, W., Hahlbrock, K., 1980. Highly purified "flavanone synthase" from parsley catalyzes the formation of naringenin chalcone. Arch. Biochem. Biophys. 200, 617-619.

Hiromoto, T., Honjo, E., Noda, N., Tamada, T., Kazuma, K., Suzuki, M., Blaber, M., Kuroki, R., 2015. Structural basis for acceptor-substrate recognition of UDP-glucose: anthocyanidin 3-O-glucosyltansferase in clitoria ternatea. The Protein Society 24, 395-407.

Hiromoto, T., Honjo, E., Tamada, T., Noda, N., Kazuma, K., Suzuki, M., Kuroki, R., 2013. Crystal structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase from clitoria ternatea. Journal of Synchotron Radiation 20, 894-898.

Kandel, S., 2016. Structural and Functional Analysis of Grapefruit Flavonol-Specific-3-O-GT Mutant P145T. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. ETSU School of Graduate Studies, Johnson City.

Li, L., Modolo, L. V., Escamilla-Trevino, L. L., Achnine, L., Dixon, R. A., Wang, X., 2007. Crystal structures of medicago truncatula UGT85H2 - insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. J. Mol. Biol. 370, 951-963.

Lv, X., Zhao, S., Ning, Z., Zeng, H., Shu, Y., Tao, O., Xiao, C., Lu, C., Liu, Y., 2015. Citrus fruits as a treasure trove of active natural metabolites that potentially provide benefits for human health. Chem. Cent. J. 9, 1-4.

Mann, K. G., Batt, C. W., 1969. The Molecular Weights of Bovine Thrombin and Its Primary Autolysis Products. The journal of biological chemistry 244, 6555-6557.

McGarvey, D. J., Croteau, R., 1995. Terpenoid Metabolism. The Plant Cell 7, 1015-1026.

McIntosh, C. A., Latchinian, L., Mansell, R. L., 1990. Flavanone specific 7-Oglucosyltransferase activity in Citrus paradisi seedlings: purification and characterization. Arch. Biochem. Biophys. 282, 50-57.

McIntosh, C. A., Owens, D. K., 2016. Advances in flavonoid glycosyltransferase research: integrating recent findings with long-term citrus studies. Phytochem Review 15, 1075-1091.

Miller, B. J., Pieterse, T., Marais, C., Bezuidenhoudt, B. C., 2012. Ring-closing metathesis as a new methodology for the synthesis of monomeric flavonoids and neoflavonoids. Tetrahedron Lett. 53, 4708-4710.

Modolo, L. V., Li, L., Pan, H., Blount, J. W., Dixon, R. A., Wang, X., 2009. Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. J. Mol. Biol. 392, 1292-1302.

Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., Davies, G. J., 2006. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. The Embo Journal 25, 1396-1405.

Owens, D. K., McIntosh, C. A., 2009. Identification, recombinant expression, and biochemical characterization of a flavonol 3-O-glucosyltransferase clone from Citrus paradisi. Phytochemistry 70, 1382-1391.

Peters, N. K., Frost, J. W., Long, S. R., 1986. A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. Science 233, 977-980.

Sathanantham, P., 2015. Impact of mutations of targeted serine, histidine, and glutamine residues in Citrus paradisi flavonol specific glucosyltransferase activity. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. ETSU School of Graduate Studies, Johnson City, Tennessee, USA.

Shao, H., He, X., Achnine, L., Blount, J. W., Dixon, R. A., 2005. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. Plant Cell 17, 3141-3154.

Winkel-Shirley, B., 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126, 485-493.

Winkel-Shirley, B., 2002. Biosynthesis of flavonoids and effects of stress. Curr. Opin. Plant Biol. 5, 218-223.

CHAPTER 3

SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

In this research, a wild type flavonol specific glucosyltransferase from grapefruit (CP3GT) was sequenced from plasmid stock and verified to possess an inserted thrombin cleavage site in frame and upstream of a c-myc and a 6x-His tag. Site directed mutagenesis was performed on a P297F mutant to insert a thrombin cleavage site. Sequencing data for the recombinant P297F mutant was analyzed to verify the presence of the P-F residue mutation and that a thrombin cleavage site was inserted in frame and upstream of a c-myc identification tag and a 6x-His purification tag. The P145T mutant was previously verified to have a thrombin cleavage site and the two tags in frame. Another mutant of interest, S20G-T21S did not undergo site directed mutagenesis. The recombinant WT CP3GT had previously undergone thrombin insertion and was verified to possess all inserts in frame.

Wild type CP3GT was transformed into yeast and expressed using methanol induction. Initial inductions sought to scale up protein production and purification in preparation for harvesting highly concentrated CP3GT protein for use in crystallography screens, however, inductions that used large-batch yeast cultures (3-5L) were often unsuccessful due to both contamination and unreliability of large-scale incubators. Those inductions that were successful encountered additional problems during protein purification. Scale up of purification required larger IMAC resin beds to be poured on the order of 7-10mL, 3-5 times larger than what was normally used. The time required to bind, wash, and elute such large bed volumes delayed CP3GT acquisition and increased the chance of CP3GT activity loss. It was concluded that a more efficient way of obtaining concentrated protein was to induce cultures no greater than 1L at a time and use multiple 2.5mL IMAC resin beds. Pure protein eluted from each of the columns would then be pooled, dialyzed, and concentrated until desired concentration was obtained.

Activity assays of CP3GT were conducted using quercetin as the acceptor substrate and radioactive UDP-glucose as the donor substrate. The first assay was designed to determine initial velocity

and proceeded for 30 minutes. Data from this assay showed that CP3GT activity was linear for 15 minutes when 2µg of protein was used, therefore, all subsequent assays used 2µg of protein and were conducted for 10 minutes at 37C.

Pure CP3GT was subjected to a variety of digestion conditions with thrombin at various concentrations. Aliquots of 30µg of were treated with 0, 3, 6, 30, 60, and 100 units of thrombin. Respective samples were incubated at both 4ºC and 10ºC. A concentration of 3.3U/µg of protein was determined to be sufficient for cleaving tags when the digestion proceeded at 4º C for 2 hours. To determine the impact these storage conditions had on CP3GT activity, an assay was performed using freshly purified CP3GT and CP3GT following storage for 2 hours at 4ºC. Neither of these two samples were digested with thrombin. Results showed that 70% of activity was retained during these conditions and this treatment will be used to prepare protein for crystallization experiments.

To determine the impact of thrombin treatment on CP3GT activity, assays were performed on CP3GT digested with 0, 1.65, 3.3, and 6.6U/µg respectively for 2 hours at 4ºC. Data showed that thrombin treatment did not impact CP3GT activity. Doubling the concentration of thrombin necessary to digest CP3GT showed no significant impact on CP3GT activity. From these data it was concluded that digestion with thrombin even at high concentrations will not significantly decrease CP3GT activity, therefore, it is a reliable means of removing tags in preparation for crystallization.

Future work should focus on preparing the P297F, S20G-T21S, and P145T CP3GT mutants for crystallization. Site directed mutagenesis should be performed on the S20G-T21S mutant to insert a thrombin cleavage site. Insertion in frame should be verified by plasmid sequencing. The P297F and P145T mutants have already been verified to possess the thrombin cleavage site in frame. Once all mutants possess the thrombin cleavage site, they should each be transformed into yeast. From each mutant should be generated reliable, well-labeled cell stocks stored at -80ºC.

A time course of expression for each mutant should be performed to determine peak protein production. Once expression is optimized, protein can be harvested, and purification can be optimized. The methods used in this work for protein purification did not produce 95-99% pure CP3GT due to inadequate washing of proteins weakly bound to the IMAC column. Only 5-10mL of 10mM imidazole buffer was used to elute nonspecific proteins which was not sufficient. Future experimentation can address these obstacles by using 20mL or more of the 10mM imidazole to completely remove nonspecific proteins from the column. Low-astringency, 10mM washing should continue until collected fractions show zero absorbance at 280nm. Following purification, thrombin digestion should be optimized for each mutant. Activity assays should be performed to determine the impact of short-term storage and impact of thrombin treatment on each of the mutant proteins.

Future experimentation should determine conditions to separate cleaved tags and thrombin from CP3GT protein following digestion. This step is critical to determining conditions for crystallization, as that step requires a native protein that is at least 95% pure. Contamination from thrombin and the cleaved tags might be addressed using size exclusion chromatography. Bovine thrombin has a molecular weight of approximately 33 kDa and the c-myc and 6x-His tag has a combined molecular weights of approximately 2.5 kDa (Mann and Batt 1969). The protein of interest, CP3GT, has a molecular weight of 51.2 kDa. Therefore, a size-exclusion column can be made and calibrated, then used to separate pure CP3GT protein.

Further experimentation should be conducted to explore the effect of the c-myc and 6x-His tags on CP3GT stability when assayed for optimal pH. Previous experimentation showed that when CP3GT was expressed in yeast it had a broader optimal pH curve than when it was expressed in bacteria (Owens and McIntosh 2009; Devaiah et al 2016). This change may have been the result of a thioredoxin tag used instead of the c-myc/6x-His tags when expressed in bacteria. Thioredoxin/6x-His tags were located at the N terminus while c-myc/6x-His tags were located at the C terminus when expressed in yeast, thus the difference in position may play a role as well. Kinetic assays on CP3GT with and without tags may

determine what impact tags have on CP3GT activity. Additional assays should be performed that use flavanone and anthocyanidin substrates, particularly the P145T mutant, which produced naringenin 7-O-glucoside in a screening assay contrary to interpretation of predictive models that showed this reaction would not occur. Other assays should be carried out to determine Km, Vmax, Ki, optimal pH, and optimal temperature. CP3GT should be assayed with UDP to determine the extent at which feedback inhibition occurs.

Following tag removal and separation, crystallization screens should be performed to optimize conditions for producing CP3GT crystals. Screening kits have already been purchased that use a wide array of reagents at varying concentrations and pH's commonly used to crystallize proteins. Once conditions for crystallization have been narrowed down using the screening kits, crystallization trials should begin.

There exist many methods for protein crystallization but the most widely used techniques can be divided into three categories: vapor diffusion, microbatch, and microdialysis (Chayen and Saridakis, 2008). In principle, vapor diffusion involves placing a drop composed of protein sample and reagent containing half the concentration of precipitant present in the reservoir in a container that has both a reservoir and a platform. The drop is positioned based on the technique employed and the reservoir is filled with reagent. Over time, due to the difference in the concentration of the precipitant, diffusion drives water vapor from the drop toward the reservoir effectively increasing the saturation of protein in the drop. The slow increase in protein saturation assists with crystal formation, but other parameters (protein concentration, pH, thermal stability, genetic modifications, etc.) will impact crystal formation. Vapor diffusion methods include the sitting and hanging drop techniques (Davies and Segal 1971).

The sitting drop method uses a raised platform in the center of the reservoir tank upon which the reagent/sample drop sits. The method provides a stable sitting position for the drop. Vapor diffusion techniques in general use multi-well plates each with their own drop and reservoir so loading them can be time consuming. Sitting drop has the benefit of being easier to load and functions well with liquid

handling robots (Chayen and Saridakis, 2008). There is a reduced risk with this method of the crystal falling into the reservoir solution (Stevens 2000).

The hanging drop method uses multi well plates with a reservoir and a cover slip upon which the reagent/sample drop sits. The drop is placed on the cover slip and the cover slip is then carefully placed over the well with a silicone seal. The cover slip is oriented such that the drop hangs above the reservoir. This technique has the benefit of allowing multiple drops to be placed over a single reservoir. Additionally, the cover slip makes examination and crystal collection much easier.

Microbatch methods of crystallization involve combining the sample and reagent and sealing the mixture in some form of container under a layer of paraffin oil. Due to the small volumes used, the oil layer is necessary to prevent quick evaporation. Microbatch can be done without oil but this typically requires larger volumes of reagent and protein, the latter of which may not be available or is too laborious to obtain (D'Arcy et al. 1996). This technique has the benefit of being very simple and cost effective, but it only works with about 50% of proteins (Chayen et al. 1990).

Microdialysis involves separating the protein sample from the reagent with a semi-permeable membrane. This membrane allows reagents to pass through but prevents protein sample from doing so. Crystallization occurs when the proper equilibrium is reached through the simple diffusion of reagent in or out of the sample, while the protein sample remains at a constant concentration. Like microbatch, microdialysis only works with certain proteins and is typically used once vapor diffusion methods have been exhausted (McPherson 2004).

When attempting to crystallize, the protein of interest should be 95-99% pure and homogenous (Dessau and Modis 2011). Several methods exist to assist with crystallization. Dynamic Light Scattering can be used to test homogeneity and identify potential nucleation that may lead to crystallization (Ferre-D'Amare and Burley 1997). Differential Scanning Fluorimetry can be used to measure the temperature stability of the protein in the presence of different pH, buffers, precipitants, ionic strengths (NaCl) or other additives (Niesen et al. 2007). Sparse-matrix screens can be employed and are designed to use samplings of reagent formulations that previously crystallized a protein (Thakur et al. 2007).

The hanging drop method of vapor diffusion should be tested first because all plant secondary product glucosyltransferases crystallized to date have used this method. Additionally, supplies are readily available in the collaborative lab and our crystallography expert has the most experience with this method. Conditions under which the six plant GT's have been crystallized are summarized in Table 1.2. In addition to all having employed the vapor diffusion method, protein concentrations during crystallization were in the range of 5-20mg/mL with 5mg/mL being the most common. All use reservoir solutions with polyethylene glycol at percentages from 18-40% with approx. 25% being most common. A wide range of buffers were used with sodium citrate and ammonium acetate appearing twice. The reservoir solutions had pH ranges from 5.5-8 with slightly acidic 5.5 being most common. Some crystals appeared as early as 24 hours while other proteins took up to two weeks to crystallize (Table 1.2).

Table 1.2: Crystallization Specifications for Six Plant Glucosyltransferases

Year – Author	Name	Conditions
2005 – Shao et al.	UGT71G1	**Method:** Hanging Drop **Protein Concentration:** 5mg/mL **Reservoir solution:** 40% (w/v) polyethylene glycol 3350 0.2M ammonium acetate 0.1M sodium citrate, pH 5.6 **Temperature:** 20 C **Time:** 2-5 days
2006 – Offen et al.	VvGT1	**Method:** Hanging drop **Protein Concentration:** 4.8mg/mL **Reservoir solution:** 18-22% polyethylene glycol 10000 (Fluka) 0.1M Bis-tris propane, pH 7.0 0-0.5% (v/v) Pluronic F-68 (Hampton) **Temperature:** not specified **Time:** 24h
2007 – Li et al.	UGT85H2	**Method:** Hanging Drop **Protein Concentration:** 5mg/mL **Reservoir solution:** 20% (w/v) PEG 3350 0.2M magnesium formate (pH5.9) **Temperature:** 20C **Time:** 1-2 weeks
2007 – Brazier – Hicks et al.	UGT72B1	**Method:** Hanging drop **Protein Concentration:** 5mg/mL **Reservoir solution:** 23% (w/v) PEG 3350 0.1M MMT buffer, pH 8.0 **Temperature:** not specified **Time:** not specified
2009 – Modolo et al.	UGT78G1	**Method:** Hanging Drop **Protein Concentration:** 15mg/mL **Reservoir solution:** 25% (w/v) polyethylene glycol 100 mM Bis-Tris, pH 5.5 **Temperature:** 20C **Time:** not specified
2015 – Hiromoto et al.	UGT78K6	**Method:** Hanging Drop **Protein Concentration:** 20mg/mL **Reservoir solution:** 26% (w/v) PEG 4000 0.1M Sodium Citrate Tribasic Dihydrate (pH 5.6) 0.2M Ammonium Acetate **Temperature:** 293K **Time:** not specified

In addition to wild type, all mutant proteins should be crystallized because these mutants have altered activity and specificity. CP3GT should be co-crystallized with substrate to better understand residue-substrate interactions. Solving these structures will add data that may improve reliability of predictive models and provide more insight into the structural elements that account for flavonol specificity and the altered activities of each CP3GT mutant.

REFERENCES

Achnine L, Huhman DV, Faraq MA, Sumner LW, Blount JW, Dixon RA. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. *The Plant Journal: for cell and molecular biology* 2005;41(6):875-87.

Adepoju OA. Using site-directed mutagenesis to determine Impact of amino acid substitution on substrate and regiospecificity of grapefruit flavonol 3-O-glucosyltransferase. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. Johnson City, Tennessee, USA: ETSU School of Graduate Studies; 2014.

Agati G, Matteini P, Goti A, Tattini M. Chloroplast-located flavonoids can scavenge singlet oxygen. *New Phytologist* 2007;174:77-89.

Agati G, Stefano G, Biricolti S, Tattini M. Mesophyll distribution of antioxidant flavonoids in Ligustrum vulgare leaves under contrasting sunlight irradiance. *Annals of Botany* 2009;104:853-861.

Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *The Plant Cell* 1998;10(7):1135-1149.

Benavente-Garcia O, Castillo J. pdate on uses and properties of citrus flavonoids: new findings in anticancer, cardiovascular, and anti-inflammatory activity. *Journal of Agricultural and Food Chemistry* 2008;56(15):6185-205.

Bennet RN, Wallsgrove RM. Secondary metabolites in plant defence mechanisms. *New Phytologist* 1994;127(4):617-633.

Berhow M, Tisserat B, Kanes K, Vandercook C. Survey of phenolic compounds produced in citrus. USDA; 1998.

Borisova SA, Liu H. Characterization of the glycosyltransferase desVII and Its auxiliary partner protein desVIII in the methymycin/pikromycin biosynthetic pathway. *Biochemistry* 2010;49(37):8071-8084.

Bourgaud F, Gravot A, Milesi S, Gontier E. Production of plant secondary metabolites: a historical perspective. *Plant Science* 2001;161:839-851.

Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim E, Bowles DJ, Davies GJ, Edwards R. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *PNAS* 2007:20238-20243.

Cahoon EB, Lindqvist Y, Schneider G, Shanklin J. Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94(10):4872-4877.

Chandrika BB, Steephan M, Kumar TR, Sabu A, Haridas M. Hesperetin and Naringenin sensitize HER2 positive cancer cells to death by serving as HER2 Tyrosine Kinase inhibitors. *Life Science* 2016;160:47-56.

Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods* 2008;5(2):147-153.

Chayen NE, Shaw-Stewart PD, Maeder DL, Blow DM. An automated system for micro-batch protein crystallization and screening. *Journal of Applied Crystallography* 1990;23:297-302.

Cody V. Crystal and molecular structures of flavonoids. *Progress in Clinical and Biological Research* 1988;280:29-44.

Coutinho PM, Deleury E, Davies GJ, Henrissay B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* 2003;328(2):307-317.

Croteau R, Kutchan TM, Lewis NG. Natural products (secondary metabolites). Biochemistry and Molecular Biology of Plants. Hoboken: Wiley; 2000. p 1250-1318.

Cushnie T, Lamb AJ. Antimicrobian activity of flavonoids. *International Journal of Antimicrobial Agents* 2005;26(5):343-356.

D'Arcy A, Elmore C, Stihle M, Johnston JE. A novel approach to crystallising proteins under oil. *Journal of crystal growth* 1996;168(1-4):175-180.

Davies DR, Segal DM. Protein crystallization: micro techniques involving vapor diffusion. *Methods in Enzymology* 1971;22:266-269.

Dessau MA, Modis Y. Protein crystallization for x-ray crystallography. *Journal of Visualized Experiments* 2011;47:2285.

Devaiah SP, Owens DK, Sibhatu MB, Sarkar TR, Strong CL, Mallampalli VKPS, Asiago J, Cooke J, Kiser S, Lin Z, Wamucho A, Hayford D, McIntosh CA. Identification, Recombinant Expression, and Biochemical Analysis of Putative Secondary Product Glucosyltransferases from Citrus Paradisi. *Journal of Agricultural and Food Chemistry* 2016:1957-1969.

Devaiah SP, Tolliver BM, Zhang C, Owens DK, McIntosh CA. Mutational analysis of substrate specificity in a Citrus paradisi flavonol-3-O-glucosyltransferase. *Journal of plant biochemistry and biotechnology* 2018;27:13-17.

Dickson JMJ, Lee W, Shepherd PR, Buchanan CM. Enzyme activity effects of N-terminal His-tag attached to catalytic sub-unit of phosphoinositide-3-kinase. *Bioscience Reports* 2013;33(6):857-863.

Dower JI, Geleijnse JM, Gijsbers L, Schalkwijk C, Kromhout D, Hollman PC. Supplementation of the pure flavonoids epicatechin and quercetin affects some biomarkers of endothelial dysfunction and inflammation in (pre)hypertensive adults: a randomized double-blind, placebo-controlled, crossover trial. *The Journal of Nutrition* 2015;145:1459-1463.

Eisenreich W, Rohdich F, Bacher A. Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Science* 2001;6(2):78-84.

Fayoux SC, Hernandez RJ, Holland RV. The debittering of navel orange juice using polymeric films. *Journal of Food Science* 2007;72(4):E143-54.

Ferguson AD, Chakraborty R, Smith BS, Esser L, Helm Dvd, Deisenhofer J. Structural Basis of Gating by the Outer Membrane Transporter FecA. *Science* 2002;295(5560):1715-1719.

Ferre-D'Amare AR, Burley SK. Dynamic light scattering in evaluating crystallizability of macromolecules. *Methods in Enzymology* 1997;276:157-166.

Ferreyra MLF, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers in plant science* 2012;3(222).

Figueiredo CA, Barroso JG, Pedro LG, Scheffer JJ. Factors affecting secondary metabolite production in plants: volatile components and essential oils. *Flavour and Fragrance Journal* 2008;23(4):213-226.

Ford CM, Boss PK, Hoj PB. 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. *The Journal of Biological Chemistry* 1998;273(15):9224-33.

Foreign Agriculture Service. Citrus: World Markets and Trade. Washington, D.C.: United States Department of Agriculture; 2018.

Freydank AC, Brandt W, Drager B. Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins* 2008;72(1):173-183.

Fukuchi-Mizutani M, Okuhara H, Fukui Y, Nakao M, Katsumoto Y, Yonekura-Sakakibara K, Kusumi T, Hase T, Tanaka Y. Biochemical and molecular characterization of a novel UDP-glucose: anthocyanin 3′-O-glucosyltransferase, a key enzyme for blue anthocyanin biosynthesis, from Gentian. *Plant Physiology* 2003;132:1652-1663.

Gachon CM, Langlois-Meurinne M, Saindrenan P. Plant secondary metabolism glycosyltransferases: the emerging functional analysis. *Trends in Plant Science* 2005;10(11):542-9.

Ghasemi K, Ghasemi Y, Ebrahimzadeh MA. Antioxidant activity, phenol and flavonoid contents of 13 citrus species peels and tissues. *Pakistan Journal of Pharmaceutical Sciences* 2009;22(3):277-81.

Gloster TM. Advances in understanding glycosyltransferases from a structural perspective. *Current opinion in structural biology* 2014;28:131-141.

Gould F. Sustainability of transgenic insecticidal cultivars: integrating pest genetics and ecology. *Annual Review Entomology* 1998;43:701-726.

Gulcin I. Antioxidant activity of food constituents: an overview. *Archives of Toxicology* 2012;86(3):345-91.

Harborne JB. Ecological chemistry and biochemistry of plant terpenoids. Proceedings of the Phytochemical Society of Europe. New York: Clarendon Press; 1991. p 399-426.

Heller W, Hahlbrock K. Highly purified "flavanone synthase" from parsley catalyzes the formation of naringenin chalcone. *Archives of Biochemistry and Biophysics* 1980;200(2):617-9.

Hiromoto T, Honjo E, Noda N, Tamada T, Kazuma K, Suzuki M, Blaber M, Kuroki R. Structural basis for acceptor-substrate recognition of UDP-glucose: anthocyanidin 3-O-glucosyltansferase in clitoria ternatea. *The Protein Society* 2015;24:395-407.

Hiromoto T, Honjo E, Tamada T, Noda N, Kazuma K, Suzuki M, Kuroki R. Crystal structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase from clitoria ternatea. *Journal of Synchotron Radiation* 2013;20(6):894-898.

Hooper L, Kroon PA, Rimm EB, Cohn JS, Harvey I, Cornu KAL, Ryder JJ, Hall WL, Cassidy A. Flavonoids, flavonoid-rich foods, and cardiovascular risk: a meta-analysis of randomized controlled trials. *The American Journal of Clinical Nurtition* 2008;88(1):38-50.

Huang WY, Cai YZ, Zhang Y. Natural phenolic compounds from medicinal herbs and dietary plants: potential use for cancer prevention. *Nutrition and Cancer* 2010;62(1):1-20.

Hughes J, Hughes MA. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. *DNA sequence* 1994;5(1):41-9.

International Union of Biochemistry and Molecular Biology. Quercetin 3-O-Glycoside Derivatives Biosynthesis. 2004.

Irwin RE, Cook D, Richardson LL, Manson JS, Gardner DR. Secondary compounds in floral rewards of toxic rangeland plants: impacts on pollinators. *Journal of Agricultural Food Chemistry* 2014;62(30):7335-44.

Iyer PV, Ananthanarayan L. Enzyme stability and stabilization—Aqueous and non-aqueous environment. *Process biochemistry* 2008;43(10):1019-1032.

Jackson R, Knisley D, McIntosh C, Pfeiffer P. Predicting flavonoid UGT regioselectivity. *Advanced Bioinformatics* 2011.

Jia Q. Generating and Screening a Natural Product Library for CYclooxygenase and Lipoxygenase Dual Inhibitors. *Studies in natural products chemistry* 2003;29:643-718.

Jones P, Vogt T. Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* 2001;213(2):164-174.

Jux A, Gleixner G, Boland W. Classification of terpenoids according to the methylerythritolphosphate or the mevalonate pathway with natural 12C/13C isotope ratios: dynamic allocation of resources in induced plants. *Angewandte Chemie International Edition* 2001;40(11):2091-2094.

Kandel S. Structural and Functional Analysis of Grapefruit Flavonol-Specific-3-O-GT Mutant P145T. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. Johnson City: ETSU School of Graduate Studies; 2016.

Kelly EH, Anthony RT, Dennis JB. Flavonoid antioxidants:chemistry, metabolism and structure-activity relationships. *Journal of Nutritional Biochemistry* 2002;13:572-584.

Khoddami A, Wilkes MA, Roberts TH. Techniques for analysis of plant phenolic compounds. *Molecules* 2013;18(2):2328-2375.

Kirca A, Cemeroglu B. Degradation kinetics of anthocyanins in blood orange juice and concentrate. *Food Chemistry* 2003;81(4):583-587.

Klepacka J, Gujska E, Michalak J. Phenolic compounds as cultivar- and variety-distinguishing factors in some plant products. *Plant Foods for Human Nutrition* 2011;66(1):64-69.

Klibanov AM. Stabilization of Enzymes against Thermal Inactivation. *Advances in Applied Microbiology* 1983;29:1-28.

Knaggs AR. The biosynthesis of shikimate metabolites. *Natural Product Reports* 2003;20(1):119-36.

Kramer CM, Prata RTN, Willits MG, De Luca V, Steffens JC, Graser G. Cloning and regiospecificity studies of two flavonoid glucosyltransferases from allium cepa. *Phytochemistry* 2003;64:1069-1076.

Kumar S, Pandey AK. Chemistry and Biological Activities of Flavonoids: An Overview. *The scientific world journal* 2013.

Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annual Review of Biochemistry* 2008;77:521-555.

Lauber J, Handrick R, Sebastian L, Durre P, Gaisser S. Expression of the functional recombinant human glycosyltransferase GalNAcT2 in escherichia coli. *Microbial Cell Factories* 2015;14(3).

Le Roy J, Huss B, Creach A, Hawkins S, Neutelins G. Glycosylation Is a major regulator of phenylpropanoid availability and biological activity in plants. *Frontiers in Plant Science* 2016;7:735.

Li J, Ou-Lee TM, Raba R, Amundson, R. G., Last RL. Aribidopsis flavonoid mutants are hypersensitive to UV-B irradiation. *The Plant Cell* 1993;5(2):171-179.

Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. Crystal structures of medicago truncatula UGT85H2 - insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *Journal of Molecular Biology* 2007;370(5):951-963.

Li X, Song H, Yao S, Jia C, Yang Y, Zhu W. Quantitative analysis and recovery optimisation of flavonoids and anthocyanins in sugar-making process of sugarcane industry. *Food Chemistry* 2011;125(1):150-157.

Liang D, Liu J, Wu H, Wang B, Zhu H, Qiao J. Glycosyltransferases: mechanisms and applications in natural product development. *Chemical Society Reviews* 2015;44(22):8350-8374.

Lombard V, Golaconda RH, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* 2014;42:D490-D495.

Loris EA, Panjikar S, Ruppert M, Barleben L, Unger M, Schubel H, Stockigt J. Structure-based engineering of strictosidine synthase: auxiliary for alkaloid libraries. *Chemistry and Biology* 2007;14(9):979-985.

Lv X, Zhao S, Ning Z, Zeng H, Shu Y, Tao O, Xiao C, Lu C, Liu Y. Citrus fruits as a treasure trove of active natural metabolites that potentially provide benefits for human health. *Chemistry Central Journal* 2015;9:1-4.

Maier VP, Bennet RD, Hasegawa S. Limonin and other limonoids. Citrus Science and Technology. Volume 1: Avi Publishing Company; 1977.

Mann KG, Batt CW. The Molecular Weights of Bovine Thrombin and Its Primary Autolysis Products. *The journal of biological chemistry* 1969;244:6555-6557.

Markham KR, Ternai B, Stanley R, Geiger H, Mabry TJ. Carbon-13 NMR studies of flavonoids-III: Naturally occuring flavonoid glyosides and their acylated derivatives. *Tetrahedron* 1978;34(9):1389-1397.

Matsuura HN, Fett-Neto AG. Plant alkaloids: main features, toxicity, and mechanisms of action. Toxinology. Netherlands: Springer; 2015. p 1-15.

McGarvey DJ, Croteau R. Terpenoid Metabolism. *The Plant Cell* 1995;7:1015-1026.

McIntosh C, Mansell RL. Three-dimensional distribution of limonin, limonoate A-ring monolactone, and naringin in the fruit tissues of three varieties of Citrus paradisi. *Journal of Agriculture and Food Chemistry* 1997;45(8):2876-2883.

McIntosh CA. Quantification of Limonin and Limonoate A-Ring Monolactone During Growth and Development of Citrus Fruit and Vegetative Tissues by Radioimmunoassay. Citrus Lemonoids: American Chemical Society; 2000. p 73-95.

McIntosh CA, Latchinian L, Mansell RL. Flavanone specific 7-Oglucosyltransferase activity in Citrus paradisi seedlings: purification and characterization. *Archives of Biochemistry and Biophysics* 1990;282(1):50-57.

McIntosh CA, Mansell RL. Biosynthesis of naringin in Citrus paradisi: UDP-glucosyltransferase activity in grapefruit seedlings. *Phytochemistry* 1990;29(5):1533-1538.

McIntosh CA, Owens DK. Advances in flavonoid glycosyltransferase research: integrating recent findings with long-term citrus studies. *Phytochem Review* 2016;15(6):1075-1091.

McPherson A. Introduction to protein crystallization. *Methods* 2004;34(3):254-265.

Miler M, Zivanovic J, Ajdzanovic V, Orescanin-Dusic Z, Milenkovic D, Konic-Ristic A, Blagojeciv D, Milosevic V, Sosic-Jurjevic B. Citrus flavanones naringenin and hesperetin improve antioxidant status and membrane lipid compositions in the liver of old-aged Wistar rats. *Experimental Gerontology* 2016;84:49-60.

Miller BJ, Pieterse T, Marais C, Bezuidenhoudt BC. Ring-closing metathesis as a new methodology for the synthesis of monomeric flavonoids and neoflavonoids. *Tetrahedron Letters* 2012;53(35):4708-4710.

Modolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X. Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *Journal of Molecular Biology* 2009;392(5):1292-1302.

National Agricultural Statistics Service. Citrus Fruits: 2017 Summary. Washington, DC: USDA - NASS; 2017.

Newman DJ, Gordon MC. Natural products as sources of new drugs from 1981 to 2014. *Journal of Natural Products* 2016;79(3):629-661.

Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* 2007;2(9):2212-21.

Nimse SB, Pal D. Free radicals, natural antioxidants, and their reaction mechanisms. *RSC Advances* 2015;5:27986-28006.

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *The Embo Journal* 2006;25(6):1396-1405.

Owens DK, McIntosh CA. Identification, recombinant expression, and biochemical characterization of a flavonol 3-O-glucosyltransferase clone from Citrus paradisi. *Phytochemistry* 2009;70(11-12):1382-1391.

Owens DK, McIntosh CA. Biosynthesis and function of Citrus glycosylated flavonoids. The Biological Activity of Phytochemicals. New York: Springer; 2011. p 67-95.

Pal S, Saha C. A review on structure-affinity relationship of dietary flavonoids with serum albumins. *Journal of Biomolecular Structure and Dynamics* 2014;32(7):1132-47.

Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. *Journal of Nutritional Science* 2016;5(47):1-15.

Peters NK, Frost JW, Long SR. A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. *Science* 1986;233(4767):977-80.

Qiu S, Sun H, Zhang AH, Xu HY, Yan GL, Han Y, Wang XJ. Natural alkaloids: basic aspects, biological roles, and future perspectives. *Chinese Journal of Natural Medicines* 2014;12(6):401-6.

Rao AG. The outlook for protein engineering in crop improvement. *Plant Physiology* 2008;147(1):6-12.

Rodriguez-Mateos A, Vauzour D, Krueger CG, Shanmuganayagam D, Reed J, Calani L, Mena P, Del Rio D, Crozier A. Bioavailability, bioactivity and impact on health of dietary flavonoids and related compounds: an update. *Archives of Toxicology* 2014;88(10):1803-1853.

Sahu BD, Kalvala AK, Koneru M, Mahesh KJ, Kuncha M, Rachamalla SS, Ramakrishna, S. Ameliorative effect of fisetin on cisplatin-induced nephrotoxicity in rats via modulation of NF-κB activation and antioxidant defence. *PLoS One* 2014;9(9):1-15.

Sathanantham P. Impact of mutations of targeted serine, histidine, and glutamine residues in Citrus paradisi flavonol specific glucosyltransferase activity. East Tennessee State University School of Graduate Studies: Electronic Theses and Dissertations. Johnson City, Tennessee, USA: ETSU School of Graduate Studies; 2015.

Schlager S, Drager B. Exploiting plant alkaloids. *Current Opinion in Biotechnology* 2016;37:155-164.

Schuman B. Displays the associative chemical mechanisms utilized by retaining and inverting glycosyltransferases. English Wikipedia; 2013.

Shao H, He X, Achnine, L., Blount JW, Dixon RA. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. *Plant Cell* 2005;17(11):3141-54.

Smith A, Oertle J, Warrn D, Prato D. Quercetin: A promising flavonoid with a dynamic ability to treat various diseases, infections, and cancers. *Journal of cancer therapy* 2016;7:83-95.

Smyth MS, Martin JHJ. X-ray crystallography. *Molecular pathology* 2000;53(1):8-14.

Soberon M, Pardo-Lopez L, Lopez I, Gomez I, Tabashnik BE, Bravo A. Engineering modified Bt toxins to counter insect resistance. *Science* 2007;318:1640-1642.

Stafford HA. Flavonoid evolution: an enzymic approach. *Plant Physiology* 1991;96:680-685.

Stephanopoulos G. Challenges in engineering microbes for biofuels production. *Science* 2007;315(5813):801-804.

Stevens RC. High-throughput protein crystallization. *Current Opinion in Structural Biology* 2000;10(5):558-563.

Subramanian S, Stacey G, Yu O. Distinct, crucial roles of flavonoids during legume nodulation. *Trends in Plant Science* 2007;12(7):282-285.

Taylor LP, Grotewold E. Flavonoids as deveolpmental regulators. *Current Opinion in Plant Biology* 2005;8:317-323.

Thakur AS, Robin G, Guncar G, Saunders NFW, Newman J, Martin JL, Kobe B. Improved success of sparse matrix protein crystallization screening with heterogeneous nucleating agents. *PLoS ONE* 2007;2(10):e1091.

Tholl D. Biosynthesis and biological functions of terpenoids in plants. Biotechnology of Isoprenoids. Advances in Biochemical Engineering/Biotechnology. Volume 148. Cham: Springer; 2015. p 63-106.

Tiwari P, Sangwan RS, Sangwan NS. Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnology Advances* 2016:714-739.

Vaistij FE, Lim E-K, Edwards R, Bowles DJ. Glycosylation of Secondary Metabolites and Xenobiotics. *Plant derived natural products* 2009:209-228.

Vilarino MP, Ravetta DA. Tolerance to herbivory in lupin genotypes with different alkaloid concentration: Interspecific differences between Lupinus albus L. and L. angustifolius L. *Environmental and Experimental Botany* 2008;63(1-3):130-136.

Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in Plant Science* 2000;5(9):380-6.

Wang SF, Liu AY, Ridsdill-Smith TJ, Ghisalberti EL. Role of alkaloids in resistance of yellow lupin to red-legged earth mite. *Journal of Chemical Ecology* 2000;26(2):429-441.

Wang T-y, Li Q, Bi K-S. Bioactive flavonoids in medicinal plants: Structure, activity and biological fate. *Asian journal of pharmaceutical sciences* 2018;13(1):12-23.

Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant physiology* 2001;126(2):485-93.

Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. *Current opinion in plant biology* 2002;5(3):218-23.

Xiao J, Muzashvili TS, Georgiev MI. Advances in the biotechnological glycosylation of valuable flavonoids. *Biotechnology Advances* 2014;32(6):1145-1156.

Yang D, Wang X, Gan L, Zhang H, Shin J, Lee K, Hong S. Effects of flavonoid glycosides obtained from a Ginkgo biloba extract fraction on the physical and oxidative stabilities of oil-in-water emulsions prepared from a stripped structured lipid with a low omega-6 to omega-3 ratio. *Food Chemistry* 2015;174:124-131.

Yi L, Ma S, Ren D. Phytochemistry and bioactivity of Citrus flavonoids: a focus on antioxidant, anti-inflammatory, anticancer and cardiovascular protection activities. *Phytochemistry Review* 2017;16:479.

Zou Z, Wanpeng X, Yan H, Chao N, Zhiqin Z. Antioxidant activity of citrus fruits. *Food Chemistry* 2016;196:885-896.

APPENDICES

APPENDIX A: RECIPES

Acceptor substrates (aglycones) for enzyme assay 50nmol/5µL (1 mL)

Quercetin
3.4 mg Quercetin
Dissolve in 1mL of ethylene glycol monomethyl ether (EGME)
Vortex to dissolve completely
Store at -20ºC

Sugar donor (UDP- glucose)

Non-radioactive for HPLC

100nmol UDP-α-D-glucose Disodium Salt (non-radioactive)/10µL, 1mL
6.1 mg of UDP-α-D-glucose
Dissolve in 1000 uL of sterile dH2O
Vortex before use and store at -20º C

Activity assays

UDP-14C-glucose (radioactive) – 40,000cpm, 100nmol/10µL, 400µL
6.78mg of UDP-glucose dissolve in 1mL of dH2O = 100nmol/9uL (Stock solution)
40uL of UDP-14C glucose (Specific activity=50 uCi/2.5 mL)
360uL of non-radioactive UDP-glucose (100nmols/9 uL)
Vortex well and verify cpm by adding 2uL of the mix to 2 mL scintillation cocktail
Store at -20 °C

Media and Buffers

Protein expression

Low salt Luria broth (LB) liquid media (200 mL)
1 g yeast extract (0.5%)
2 g tryptone (1%)
2 g sodium chloride (1%)

Dissolve ingredients in 180 mL dH2O
Adjust pH to 7.5 with 1M NaOH
Make up the volume to 200mL using milliQ dH2O
Sterilize by autoclaving for 20 minutes
Before inoculating, add zeocin to a final concentration of 25 ug/mL media
Store at 4° C

Low salt LB-Agar plates (200 mL)
1 g yeast extract (0.5%)
2 g tryptone (1%)
2 g sodium chloride (1%)

Dissolve ingredients in 150 mL dH2O
Adjust to pH=7.5 with 1M NaOH
Add 3 g agar (1.5%)
Make up volume to 200 mL with dH2O
Autoclave for 20 minutes
Once slightly cooled, add 50 uL of zeocin from stock (100mg/mL)
Pour 25mL of media into plates and allow to solidify
Store at 4°C

YPD (200mL)
2 g yeast extract (1%)
4 g peptone (2%)

Dissolve ingredients in 180 mL dH2O (Adjust final pH=6.5)
Autoclave for 20 minutes
Cool to room temperature
Add 20mL of autoclaved 20% dextrose (final concentration of 2%)
Add 200 uL of zeocin from (100 mg/mL) stock.
Store at 4°C

YPD-Agar Plate (200mL)
2 g yeast extract (1%)
4 g peptone (2%)

Dissolve ingredients in 180 mL dH2O (Final pH=6.5)
Add 4 g agar (2%)
Autoclave for 20 minutes
Cool to about 55°C and add 20 mL of autoclaved 20% dextrose
Add 200 uL of zeocin from (100 mg/mL) stock
Mix well and pour 25mL in each plate and allow it to solidify
Store plates at 4°C in dark

YPDS
2 g yeast extract (1%)
4 g peptone (2%)
36.44 g sorbitol
Dissolve in 180 mL dH2O (Final pH=6.5)
Autoclave for 20 minutes
Allow to cool to room temperature
Add 20 mL of autoclaved 20% dextrose
Add 200 uL zeocin from (100 mg/mL) stock

YPDS-Agar plates (200 mL)
2 g yeast extract (1%)
4 g peptone (2%)
36.44 g sorbitol
Dissolve in 180 mL dH2O (Final pH=6.5)
Add 4 g agar (2%)
Autoclave for 20 minutes

Allow to cool and add 20 mL of autoclaved 20% dextrose
Add 200 uL zeocin from (100 mg/mL) stock
Pour media into plates and allow to solidify at room temperature
Store plates at 4°C

Zeocin (100mg/mL) stock solution (1mL)
100mg zeocin
Dissolve in 1mL dH2O
Filter sterilize
Store at -20°C in dark

1M Sorbitol (100mL)
18.2g sorbitol
Dissolve in 80mL of dH2O
Make up to 100mL with dH20
Autoclave for 20 minutes
Store at 4ºC

20% Dextrose (100 mL)
20 g dextrose (α-D-glucose)
Dissolve in 80 mL dH2O
Make up to 100mL with dH2O
Autoclave for 20min and store at room temperature

BMGY (1L)
10 g yeast extract
20 g peptone
Dissolve in 700 mL sterile dH2O
Autoclave for 20 min
Allow to cool at room temperature

Add:
100 mL autoclaved 1M potassium phosphate buffer (pH=6.0)
100 mL filter sterilized 1M 10X YNB
100 mL filter sterilized 10X Glycerol
2 mL filter sterilized 500X Biotin
Store at 4°C

BMMY (1L)
10 g yeast extract
20 g peptone
Dissolve in 700 mL dH2O
Autoclave for 20 min
Allow to cool at room temperature

Add:
100 mL autoclaved 1M potassium phosphate buffer (pH=6.0)
100 mL filter sterilized 1M 10X YNB
100 mL filter sterilized 10X Methanol (add methanol just before use)
2mL 500X Biotin
Store at 4°C

10X GY (10% glycerol) (1000ml)
100 mL of 100Xglycerol
Mix 900 mL dH2O
Autoclave for 20 min
Store at 4º C

10X YNB (1000 mL)
17 g yeast nitrogen base
50 g ammonium sulfate
Dissolve in 700 mL dH2O
Make up to 1000mL with dH2O
Filter sterilize the solution and store at 4°C

500X Biotin (0.02% Biotin) 100ml
20 mg biotin
Dissolve in 80 mL dH2O
Make up to 100mL dH2O
Filter sterilize the solution
Store at 4°C

10X M (5% Methanol) (500 mL)
25 mL methanol
Add 475 mL of dH2O
Filter sterilize
Store at 4°C

1M potassium phosphate buffer, pH=6.0 (1000mL)
87.09g of dibasic potassium phosphate (K2HPO4)
Dissolved in 500mL of dH2O

136.09g of monobasic potassium phosphate (KH2PO4)
Dissolved in 1000mL of dH2O

Mix 132mL of 1M dibasic potassium phosphate (K2HPO4) + 868 mL of 1M monobasic potassium phosphate (KH2PO4)
Make up pH of 6.0 using Phosphoric acid of KOH
Autoclave for 20 minutes at liquid cycle and store at 4°C

Agarose gel electrophoresis

0.8% Agarose Gel (50mL)
0.4 g Agarose
Dissolve ingredient in 50 mL 1X TAE buffer
Microwave for 60 to 70 seconds until agarose dissolves
Add dH2O to make up volume of 50mL
Allow to cool and add 2.5uL of a 10mg/mL ethidium bromide
Mix well and pour 40ml into gel casting tray

50X TAE buffer (1L)
242 g Tris base
57.1 mL glacial acetic acid
Add 100 mL of 0.5M EDTA (pH=8.0)
Dissolve the ingredients in 700 mL dH2O
Make up volume to 1L with dH2O and store at room temperature

1X TAE (1L)
Take 20mL of 50X TAE stock and make up volume to 1L adding 980mL dH2O

DNA Purification

3M Sodium Acetate (100mL)
40.8g sodium acetate
Dissolve in 70mL dH2O
Adjust pH to 5.2 with glacial acetic acid
Bring total volume to 100mL with dH20
Store at room temperature indefinitely

SDS-PAGE

10X - SDS-PAGE buffer (1000mL)
30.28g Tris base
144.14g glycine
10g SDS
Dissolve in 1000ml DH2O
Store at room temperature

1X – SDS-PAGE buffer
100mL of 10X stock
Add 900mL of dH2O

4X SDS-PAGE Sample Dye (10mL)
2.4mL of 1M Tris, pH 6.8
0.8g SDS (sodium lauryl sulfate)
4mL 100% glycerol
4mg bromophenol blue
3.1mL dH2O
Store at room temperature
Add 50 uL of BME per mL of 4X sample buffer immediately before use.

10% SDS (100 mL)
10 g SDS
Dissolve in 100 mL dH2O
Store at room temperature

10% APS (Ammonium persulfate)
0.1g Ammonium persulfate
Dissolve in 1mL dH2O (use fresh each time)

Separating gel

	1 gel	2 gels
H2O……..	2.3ml	4.8ml
40% Acrylamide…….....	1.3ml	2.5ml
1.5M Tris (pH 8.8)……...	1.3ml	2.5ml
10% SDS……..	50ul	100ul
10% APS……..	50ul	100ul
TEMED…….....	2ul	4ul

Stacking gel

	1 gel	2 gels
H2O……..	730ul	1.5ml
40% Acrylamide ……..	130ul	260ul
1 M Tris (pH 6.8)……..	130ul	250ul
10% SDS……...	10ul	20ul
10% APS……..…….	10ul	20ul
TEMED…….....	1ul	2ul

Western Blotting

10X Western buffer (1000mL)
30.28g Tris base
144.14g glycine
Dissolve ingredients in 1000 mL dH2O
Store at 4oC

1X Western buffer (1L)
100 mL of 10X stock
Add 700 mL of dH2O
Add 200 mL of methanol
Mix and store at 4°C

Blocking solution (50 ml)
2.5 g non-fat milk powder
Dissolve in 50 ml of 1X TBS and add 50 µL of 20% sodium azide

20% sodium azide (1 ml)
20 mg sodium azide
Dissolve in 1 ml dH2O
Store at 4 °C

5X TBS buffer (1L)
40g NaCl
1g KCl
15g Tris base
Dissolve ingredients in 900mL dH2O
Adjust to pH=7.4 with 6N HCl
Make up to a final volume of 1L with dH2O
Store at room temperature

Primary Antibody (1:2500 dilution)
6 uL of anti-C-myc monoclonal antibody from mouse
Dissolve in 15 ml of 1X TBS

Secondary Antibody (1:10,000 dilution)
1.5 uL of anti-mouse IgG from goat
Dissolve in 15 ml of 1X TBS

Alkaline phosphate (AP) buffer (pH 9.5) (500 ml)
2.925 g NaCl
507.5 mg MgCl2.6H2O
50 ml of 1M CHES buffer (41.44g CHES dissolve in 200mL dH2O to make 200mL of 1M CHES)
Dissolve ingredients in 475 ml dH2O
Adjust pH to 9.5
Make up the volume to 500ml using dH2O
Store at room temperature

Nitro blue tetrazolium (NBT) solution (1mL)
83mg NBT dissolved in 700uL of N, N-dimethyl formamide and 300uL of dH2O
Store at -20^{o} C

5-bromo-4-chloro-3-indolyl-phosphate (BCIP) solution (1mL)
42mg BCIP dissolve in 1mL of N, N-dimethylformamide
Store at -20^{o}C

0.5 M EDTA (pH 8.0) (500 ml)
73.06 g of EDTA
Dissolve in 450 ml dH2O
Adjust the pH to 8.0
Make up the volume using dH2O
Store at room temperature

Protein extraction and purification

Breaking buffer, pH 7.5 (500ml)
50mM Sodium phosphate buffer (pH 7.5)
1mM PMSF (Phenyl methyl sulfonyl fluoride)
1mM EDTA
5% glycerol
5mM BME
(PMSF and BME should be added before use)

0.65 g monobasic Sodium phosphate
5.5 g dibasic Sodium phosphate
145mg EDTA
Dissolve in 400ml DH2O
Adjust pH to 7.5
Add 50mL of 50% glycerol
Make up volume to 500mL
Store at 4°C

Add 5ul of 200mM PMSF per ml of breaking buffer
Add 0.35uL of BME per mL of breaking buffer

200mM PMSF (1mL)
35mg of PMSF dissolved in 1mL methanol
Store at -20°C

50% glycerol (500ml)
250mL of 100%
Make up volume to 500mL adding dH2O
Mix well and store at 4° C

Equilibration/Wash buffer (pH=7.5) (500mL)
50mM Sodium phosphate buffer (pH 7.5)
300mM NaCl
5mM BME

0.65 g monobasic sodium phosphate
5.5 g dibasic sodium phosphate
8.75 g sodium chloride
Dissolve in 450mL dH2O
Adjust pH to 7.5
Make up volume to 500mL with dH2O
Store at 4°C

Add 0.35uL of BME per mL of elution buffer prior to use

Elution buffer, pH 7.5 (500 ml)
50mM Sodium phosphate buffer (pH 7.5)
300mM NaCl
5mM BME
150mM Imidazole
0.65 g monobasic sodium phosphate
5.5 g dibasic sodium phosphate
8.75 g sodium chloride
5.1 g imidazole Dissolve in 450ml dH2O and adjust pH to 7.5
Make up the volume using dH2O and store at 4 °C

Add 0.35 uL of BME per mL of elution buffer prior to use

Assay/ Final buffer, pH 7.5 (500 ml)
50mM Sodium phosphate buffer (pH 7.5)
14mM βME
0.65 g monobasic sodium phosphate
5.5 g dibasic sodium phosphate
Dissolve in 450ml dH2O and adjust pH to 7.5
Make up the volume using dH2O and store at 4 °C
Add 1µL of βME per mL of assay buffer prior to use

IMAC column regeneration and storing

20mM MES-KOH, pH 5.0 with 0.3M NaCl
0.39 g MES
1.75 g NaCl
Dissolve in 100 ml dH2O
Adjust pH to 5.0 using 1 M KOH
Store at 4 °C

IMAC Column storing solution (100 ml)

20% ethanol containing 0.1% sodium azide
20ml of 100X ethanol
Make volume to 100ml using dH2O
Add 100mg of sodium azide
Mix well and Store at 4 °C

Thrombin Buffer (100ml)

20ml of glycerol
20ml of 50mM Potassium Phosphate Buffer (pH 7.0)
60ml of dH20
0.9g NaCl

APPENDIX B: METHODS

SDS-PAGE

A separating gel mixture sufficient to pour two gels was prepared as described in appendix A. Samples were prepared by adding 1/3 sample volume of SDS-PAGE sample dye and boiling the mixture for 5 minutes. A 5ul volume of 10- 180 kDa Page Ruler pre-stained molecular weight marker (ThermoFisher CA #26616) was added to the first well, followed by the protein samples. The gel electrophoresis chamber was filled with 1X Tris-glycine-SDS buffer and the gels were submerged in the chamber after housing them in the appropriate apparatus. The gels were electrophoresed for 90 minutes at 100 volts. One of the gels was stained with Coomassie blue for one hour on a shaking platform and then destained for one hour using de-staining solution.

Western Blot

After electrophoresing, the gel not subjected to Coomassie stain was equilibrated in western buffer on a shaking platform while the western blot was prepared. The black and white western blot cassette was placed black side down in a bowl filled with western buffer. A sponge pre-soaked in western buffer was place on the cassette first, followed by a piece of Whatman filter paper cut squarely to size. The gel was then carefully placed on the filter paper with the marker oriented on the right side. A squarely cut nitrocellulose membrane was placed on top of the gel using forceps, followed by another piece of Whatman filter paper. Finally, another piece of pre-soaked sponge paper was placed on top of the membrane and the sandwich cassette was closed and latched. The completed western sandwich was then placed into the appropriate housing apparatus and submerged in an SDS chamber filled with 1X Tris-Glycine western buffer. An ice pack was placed in the chamber and electrophoresis was conducted for 90 minutes at 100 volts.

After electrophoresis was completed, the sandwich was placed again in transfer buffer black side down, opened carefully, and picked apart with tongs to access the nitrocellulose membrane. The

membrane was placed in a dish containing 1X TBS and allowed to equilibrate for 15 minutes. The membrane was then placed in a dish containing 100mL of blocking solution and placed on a shaking platform for 2 hours at room temperature or overnight at 4º C. After blocking, the membrane was rinsed well with water and then rinsed with 1X TBS. The membrane was then transferred to a clean dish and 15mL of primary antibody solution was poured over the membrane. The membrane was placed on a shaking platform and shaken vigorously for 2 hours at room temperature. After treatment with primary antibody, the membrane was rinsed thoroughly with 1X TBS 3-5 times. The membrane was transferred to a clean dish and 15mL of secondary antibody was applied. The membrane was placed on a shaking platform and shaken vigorously for 2 hours at room temperature. After treatment with secondary antibody, the membrane was rinsed thoroughly with water and then with 1X TBS. In a separate tube, 60uL of BCIP and 60uL of NBT was added to 15mL of AP buffer. The membrane was placed in a clean dish and 15 mL of the AP buffer/BCIP/NBT mix was applied to the membrane. The membrane was placed on a shaking platform and shaken vigorously for approximately 5 minutes at room temperature until bands appeared. After bands appeared, the membrane was rinsed thoroughly with water to stop the reaction. The membrane was then allowed to air dry at room temperature. After imaging the membrane, it was stored in foil to prevent light exposure.

Miniprep Plasmid Extraction for Site Directed Mutagenesis and Sequencing

Plasmid was extracted so that site directed mutagenesis for inserting thrombin cleavage site could be performed on the two CP3GT mutants: P297F and S20G-T21S. Plasmid extraction was also required for transforming DNA into competent *E. coli* cells and for verification of insert by sequencing.

A 5ml aliquot of LB Media (Appendix A) in a glass tube was inoculated with a single colony of *E. coli* that had previously been grown on zeocin selective plates and that contained plasmid carrying the CP3GT gene with the P297F and S20G-T21S mutation respectively. One microliter of zeocin stock (100mg/mL) was added to the sample and the samples were placed in an incubator shaker at 37º C and 250rpm for approximately 16 hours. A one milliliter aliquot of the culture was added to a 2ml sterile

Eppendorf tube and was centrifuged at 8,000 rpm for 2 minutes at room temperature. The supernatant was discarded and an additional 1mL of culture was added centrifugation was applied to the tubes. This process was repeated until 3-5ml of culture had been pelleted. The cells were resuspended by vortexing in 250uL of resuspension solution that contained previously added Rnase A per the manual's instructions. A 250ul aliquot of lysis solution was added and the tube was inverted 4-6 times. After inversion, 350ul of neutralizing reagent was pipetted into the tube and the mixture was again inverted 4-6 times. A white precipitate should form that has the appearance of coconut shavings. The tube was then centrifuged at 12,000 x *g* for 5 minutes at room temperature. All subsequent centrifugation steps were performed per these specifications except time differences. The supernatant was pipetted off into the GeneJET Spin column being extra careful to avoid transferring the white precipitate and being careful not to touch the white membrane inside the spin column. The spin column was centrifuged at 12,000 x *g* for 1 minute and the flow through was discarded. A 500ul aliquot of wash solution was added and the column centrifuged for 30-60 seconds. The flow through was discarded and the wash step was repeated. The column was transferred to a new 1mL Eppendorf tubes. A 50uL aliquot of elution buffer was added to the column and allowed to incubate for 2 minutes. Centrifugation was then performed for two minutes. The elution step can be repeated once more for an approximately 20% increase in plasmid yield. Plasmid concentration was tested using nanodrop (Table B.1). Plasmid was stored at 4^{o} and used as soon as possible. Thawing and refreezing of plasmid was avoided as this can cause mechanical shearing of DNA.

Table B.1: Plasmid concentration following Miniprep

CP3GT Mutant	Concentration (ng/uL)
P297F	77.45

<u>Insertion of Thrombin Cleavage Site in P297F Mutant using Site Directed Mutagenesis</u>

Polymerase chain reaction was used to add a thrombin cleavage site after the CP3GT gene and upstream of the two tags that allow for identification and western blot, C-myc and 6x His respectively.

The thrombin cleavage site codes for the amino acid sequence leucine-valine-proline-arginine-glycine-serine. When the expressed CP3GT protein is treated with thrombin, the thrombin enzyme recognizes the amino acid sequence and cleaves the peptide bond between arginine and glycine, thus removing the purification and identification tags. Gene specific primers for the thrombin cleavage site were obtained from Agilent Technologies, designed previously using Quickchange Primer Design web tool (https://www.genomics.agilent.com/primerDesignProgram.jsp?&_requestid=994370).

The PCR reaction mixture contained as follows:

CP3GT - P297F
5uL of 10X reaction buffer
1.25uL of 125ng/uL Primer (forward)
1.25uL of 125ng/uL Primer (reverse)
1uL of dNTP mix
1.5uL of Quick solution reagent:
39uL of 100x diluted P297F DNA template (30ng)
1 uL of Quick change DNA polymerase enzyme

The samples were placed into a PCR cycler and run at the following specifications advised by the Agilent site directed mutagenesis manual:

Table B.2: Specifications for Site Directed Mutagenesis PCR Cycling

Protocol Phase	Temperature (C)	Time (sec)	Cycles
Initiation	95°	60	1
Denaturation	95°	50	16
Annealing	60°	50	
Elongation	68°	90	
Extension	68°	500	1

The PCR products were electrophoresed on a 0.8% agarose gel. A band size of 1.7kbp was expected as CP3GT is 1.4kbp and the thrombin cleavage site adds approximately 300bp. The appropriate size band of 1.7kb was confirmed by measuring distance using a DNA ladder standard curve.

DpnI Digestion

1uL of DpnI enzyme (100U/uL) was pipetted into the PCR products. Mixing was performed by pipetting up and down. The samples were incubated in a 37° C water bath for 1 hour.

Transformation into competent *E. coli* cells

For each sample to be transformed a 50uL tube of Invitrogen *E. coli* competent cells was thawed on ice. Two microliters of DpnI digested DNA was added to the cells. Mixing was performed by lightly percussing the tubes on a hard surface. Pipetting up and down was avoided. The samples were incubated on ice for 30 minutes. The samples were then placed in a 42° C water bath for 30 seconds and then placed on ice for 2 minutes. A 250uL aliquot of SOC media, prewarmed to 37° C, was added to the sample. The samples were placed in an incubator shaker at 37° C and 250rpm for 1 hour. A 20-200ul aliquot of the sample was spread on zeocin-selective LB agar plates. The plates were incubated overnight at 37° C

Re-streak of Transformants and Cell Stock Collection

Transformant colonies were picked and dissolved in 10uL of LB media. A sterile inoculating loop was used to streak the transformed colonies onto new, zeocin selective LB agar plates. The plates were incubated overnight at 37° C. From the re-streaked plates, a single colony was used to inoculate 5mL of LB media containing 1uL of zeocin solution (100mg/mL). The inoculated sample was placed in an incubator shaker at 37° C and 250rpm overnight. A 500ul aliquot of culture was mixed with 500ul of a 50% glycerol solution in a sterile 2mL Eppendorf tube. The samples were labeled and stored at -80° C.

Verification of Thrombin Insertion by Plasmid Sequencing

Using the transformed cell stocks, a miniprep was performed as previously described to extract plasmid for sequencing. A 5ug sample of plasmid DNA was sent to the Molecular Biology Core Facility at University of Tennessee in Knoxville along with samples AOX primers and the thrombin cleavage primers respectively. The program Bioedit was used to analyze the sequence results. The thrombin cleavage site was verified in frame in the P297F mutant but not in the S20G-T21S mutant.

Midiprep of WT for Transformation

To transform the plasmid DNA into the *Pichia pastoris* yeast genome, a Midiprep was performed. A 5mL aliquot of LB media containing 1uL of Zeocin (100mg/mL) was inoculated with a single colony that was previously streaked from transformed cell stocks whose DNA sequence was verified to have the thrombin insert. The 5mL samples were placed in an incubator shaker at 37° C and 250rpm until an O.D. 600 of 1-2 was observed (approximately 16 hours). In a 250mL Erlenmeyer flask, 50mL of LB media containing 10uL of zeocin was inoculated with 100uL of the 5mL culture. The sample was placed in an incubator shaker at 37° C and 250rpm until an O.D. 600 of 1-2 was observed. The 50mL culture was centrifuged at 10,000g for 10min at 4° C. The supernatant was discarded, and the pellet was resuspended in 3mL of resuspension solution. To this was added 3mL of lysis solution and the sample was inverted 4 times. Three milliliters of the neutralization reagent were poured in and the solution was inverted for four cycles. A white precipitate that resembles coconut shavings should form. The samples were then centrifuged at 14,000g for 15min at 4° C. The supernatant was transferred to a new tube. Despite the precipitates appearance of delicious, toasted coconut shavings, it was avoided when transferring the supernatant to the new tube.

A 10mL aliquot of suspended purification resin was added to the supernatant. A vacuum manifold port was attached to a vacuum pump and a midicolumn was placed into the manifold port. The bound mixture of DNA and resin was placed into the midicolumn and vacuum pressure was initiated to force the mixture to flow down the column. After the entire mixture passed through the column, vacuum pressure was initiated for no more than 30sec to dry the resin. The midicolumn was then removed.

The reservoir tip portion of the midicolumn in which the resin settled was removed from the column portion of the midicolumn by cutting with scissors. The reservoir was placed into a 2mL Eppendorf tube and centrifuged at 10,000g for 2 minutes at room temperature. The reservoir was placed into a new 2mL tube. A 300uL aliquot of preheated (65-70° C) nuclease free water was added and allowed to incubate for 1 minute. The tube was then centrifuged at 10,000g for 20 seconds to elute the DNA. The manufacturer's protocol recommends centrifuging the eluate at 10,000g for 5 minutes to remove any fine

resin particles that may remain in the DNA solution. No pellet was observed after following this step. After centrifugation, the DNA was carefully transferred to a new 2mL tube. DNA concentration was determined by nanodrop spectrophotometry. Concentrations are recorded in Table B.3.

Table B.3: Concentrations of Plasmid DNA from Midiprep

CP3GT	Sample	Concentration (ng/uL)
P297F	1	191
	2	171
WT	1	300
	2	270

Linearization of Plasmid using SacI

Wild type CP3GT (rCP3GT) had previously been prepared to this point but was not successfully transformed into yeast. rCP3GT was therefore prioritized for transformation. The *Pichia pastoris* expression manual recommends linearizing the plasmid DNA before transforming into the yeast genome. The SacI restriction enzyme (Promega) was used to digest the DNA. The restriction digest protocol recommends using 5-10ug of DNA and specifies a reaction mixture composed as follows:

CP3GT WT
30 uL of Midiprep plasmid DNA
34uL of 10X Buffer J
3uL of 100X BSA
270uL Sterile H2O
3uL SacI
340uL Total volume

The reaction mixture was incubated at 37° C for 4 hours. The SacI enzyme was heat inactivated by incubating at 65° C for 10 minutes. A 1uL aliquot of the digested and undigested DNA was run on a 0.8% agarose gel respectively to confirm digestion.

DNA Purification using Phenol-Chloroform Extraction

In a 2mL tube, a 300uL aliquot of linearized DNA (approx. 9µg) was mixed with an equal volume of cold phenol-chloroform-isoamyl alcohol solution and vortexed for 30 seconds. The sample was centrifuged at 12,000 x *g* for 2-5 minutes at room temperature. From the upper phase was taken 250uL into a new Eppendorf tube. To this was added a 1/10 volume of sterile 3M sodium acetate (1/10 * 250uL = 25uL). To this was added 2.5x the total volume ice cold 100% ethanol (275uL * 2.5 = 687.5uL). The sample was incubated at -80C for 1 hour or -20C overnight. The sample was then centrifuged at 14,000rpm for 10 minutes at 4ºC. The DNA precipitate formed a very small pellet that was sometimes white and sometimes clear. The supernatant was carefully decanted or pipetted off (if pellet was loose) and discarded. A 250uL aliquot of ice cold 80% ethanol was added and the sample was centrifuged at 14,000rpm for 10 minutes at 4º C. The supernatant was decanted or pipetted off and discarded. The tube was left to airdry at room temperature with the cap off for approximately 10 minutes. A Kimwipe was placed over the tube to prevent debris from entering the tube. After drying, the DNA pellet was dissolved in 20uL of sterile milliQ water. The concentration of purified DNA was tested using nanodrop spectrophotometry (Table B.4). Purified DNA was stored at 4ºC and used for transformation as soon as possible.

Table B.4: Concentration of Purified DNA after Phenol-Chloroform Extraction

CP3GT	Sample	Concentration (ng/uL)
WT	1	2024.7
	2	1506

Transformation of Purified DNA into *Pichia pastoris* Genome

A X-33 Mut$^+$ yeast cell stock stored at -80º C was used to streak YPD plates to produce single colonies. The plates were incubated at 28ºC until colonies appeared (approximately 2 days). A single

colony was used to inoculate 5mL YPD media without zeocin in a glass test tube. The tubes were oriented at a 45-degree angle to assist aeration and placed in an incubator shaker for 28-30° C and 250rpm until an O.D. of 1-2 was observed. A 200mL aliquot of YPD media without zeocin was placed in a 1L baffled flask and inoculated with 200uL of the 5mL culture. The sample was placed in an incubator shaker at 28-30° C and 250rpm until an O.D. 600 of 1.3-1.5 was observed (approximately 16 hours).

The overnight culture was divided equally into 4, 50mL tubes. The tubes were placed in a centrifuge for 5 minutes at a force of 1500 x *g* at 4° C. The supernatant was discarded and one, 50mL culture of cells was resuspended in 50mL ice cold sterile water. The sample was centrifuged, and the supernatant was discarded. The pellet was resuspended in 25mL of ice cold sterile water and centrifuged. The supernatant was discarded, and the cells were resuspended in 2mL of ice-cold 1M sorbitol. The cells were centrifuged, and the supernatant was discarded. The cells were resuspended in 200uL of 1M sorbitol and transferred to a sterile 2mL tube and placed on ice. If cells were too viscous to accurately pipette, 25uL of 1M sorbitol was added. This process was repeated if necessary.

An electroporation cuvette and a new 2mL tube was placed on ice. In the new 2mL tube, 80uL of yeast cells was mixed with 5-10ug of purified DNA (in 5-10uL sterile water) from phenol-chloroform extraction. The mixture was pipetted into the electroporation cuvette and the cuvette was incubated on ice for 5 minutes. The electroporation cuvettes were pulsed according to the settings for yeast (*Saccharomyces cerevisiae*) as suggested by the electroporation device manufacturer (Bio-rad micro pulser). A 1mL aliquot of ice cold 1M sorbitol was immediately added after pulsing. The contents of the electroporation cuvette were transferred into a 15mL glass tube. The sample was incubated at 30° C without shaking for 1 hour. A 1mL aliquot of YPD media without zeocin was added to the sample. The samples were placed in an incubator shaker at 30° C and 200rpm for one hour. A 200ul aliquot of the culture was plated on a YPDS agar plate containing zeocin. Optionally, the remaining culture was incubated for 3 hours and then plated on YPDS agar plates in 200uL aliquots. The plates were incubated at 30° C for 2-4 days until colonies form.

A single transformed colony was selected and dissolved in 5-10uL YPD media and streaked on zeocin selective YPD agar plates. The plates were incubated at 28-30ºC for approximately 2 days. A single colony was used to inoculate 5mL YPD media containing 5uL zeocin (100mg/mL) in a glass tube. The sample was oriented at a 45-degree angle and placed in an incubator shaker at 28-30º C until an O.D. 600 of 1-2 was observed (approximately 16 hours). A 500ul aliquot of culture was mixed with 500uL of 50% glycerol solution in a sterile 2mL tube. The samples were stored at -80º C.

Direct Screening of *Pichia pastoris* clones

Eight single colonies from the *Pichia* transformation were picked and re-streaked on YPD agar and allowed to incubate for approximately 48 hours. From each of the re-streaked plates was selected a single colony to be screened for the presence of the CP3GT gene. The single colony was suspended in 10uL of dH_2O and 5uL of 5U/uL lyticase enzyme. The mixture was incubated at 30ºC for 10 minutes and then placed in a -80ºC freezer for 10 minutes to inactivate the enzyme.

A PCR reaction mix was prepared as follows:

25uL EmeraldAmp Green Mastermix (2X)
5uL 0.2uM Forward Primer
5uL 0.2uM Reverse Primer
5uL Cell Lysate
10uL dH_2O
50uL total volume

The reaction mix was placed into the PCR machine and cycled using the following parameters:

Protocol Phase	Temperature (C)	Time (min)	Cycles
Initiation	95º	5	1
Denaturation	95º	1	30
Annealing	54º	1	
Elongation	72º	1	
Extension	72º	7	1

The PCR mix was then analyzed by gel electrophoresis.

VITA

AARON BIRCHFIELD

Education:	M.S. in Biology 2019 East Tennessee State University, Johnson City, TN, USA B.S. in Chemistry 2016 East Tennessee State University, Johnson City, TN, USA
Professional Experience	Graduate Research Assistant Biological Sciences, East Tennessee State University Dr. Cecilia McIntosh Lab 2017-2018 Teaching Assistant Biological Sciences, East Tennessee State University 2016-2017
Presentations	Aaron Birchfield, Cecilia McIntosh. "Crystallization and analysis of tags of a flavonol specific glucosyltransferase found in grapefruit." Presented at Appalachian Student Research Forum, Johnson City, TN. April 2018 Aaron Birchfield, Cecilia McIntosh. "Crystallization and characterization of a flavonol specific glucosyltransferase found in grapefruit." Presented at 56th Annual Meeting of Phytochemical Society of North America, Columbia, MO, USA. August 2017 Aaron Birchfield, Cecilia McIntosh. "Crystallization and characterization of a flavonol specific glucosyltransferase found in grapefruit." Presented at Appalachian Student Research Forum, Johnson City, TN. April 2017 Aaron Birchfield, Cecilia McIntosh. "Crystallization and characterization of a flavonol specific glucosyltransferase found in grapefruit." Seminar presented to the Department of Biological Sciences, East Tennessee State University, Johnson City, TN. March 2017
Honors and Awards	Best Poster Presentation – Natural Sciences Appalachian Student Research Forum, ETSU April 2017 and 2018

	Frank and Mary Loewus Student Travel Award Phytochemical Society of North America August 2017
Professional Mentorships	Member Phytochemical Society of North America Member ETSU Graduate and Professional Student Association