12-2023

Flavonol Glucosylation: A Structural Investigation of the Flavonol Specific 3-O Glucosyltransferase Cp3GT

Aaron S. Birchfield
East Tennessee State University

Flavonol Glucosylation: A Structural Investigation of the Flavonol Specific 3-O Glucosyltransferase Cp3GT

A dissertation

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Philosophy in Biomedical Sciences

by

Aaron S. Birchfield

December 2023

Cecilia A. McIntosh, Ph.D., Chair

Ranjan Chakraborty, Ph.D.

Dhirendra Kumar, Ph.D.

Daniel Owens, Ph.D.

Robert Standaert, Ph.D.

Keywords: Flavonoid, glycosyltransferase, synthetic biology, in silico modeling, enzyme-substrate interactions, recombinant protein purification

ABSTRACT

Flavonol Glucosylation: A Structural Investigation of the Flavonol Specific 3-O Glucosyltransferase Cp3GT

by

Aaron S. Birchfield

Flavonoid glycosyltransferases (GTs), enzymes integral to plant ecological responses and human pharmacology, necessitate rigorous structural elucidation to decipher their mechanistic function and substrate specificity, particularly given their role in the biotransformation of diverse pharmacological agents and natural products. This investigation delved into a comprehensive exploration of the flavonol 3-O GT from *Citrus paradisi* (Cp3GT), scrutinizing the impact of a c-terminal c-myc/6x histidine tag on its enzymatic activity and substrate specificity, and successfully achieving its purification to apparent homogeneity. This established a strong foundation for potential future crystallographic and other structure/function analyses. Through the strategic implementation of site-directed mutagenesis, a thrombin cleavage site was incorporated proximal to the tag, followed by cloning in *Pichia pastoris*, methanol-induced expression, and cobalt-affinity chromatography for initial purification stages. Notably, the recombinant tags did not exhibit a discernible influence on Cp3GT kinetics, substrate preference, pH optima, or metal interactions, maintaining its specificity towards flavonols at the 3-OH position and favoring glucosylation of quercetin and kaempferol. Subsequent purification steps, including MonoQ anion exchange and size-exclusion chromatography, yielded Cp3GT with ≥95% homogeneity. *In silico* molecular models of Cp3GT and its truncated variants, Cp3GTΔ80 and Cp3GTΔ10, were constructed using D-I-TASSER and COFACTOR to assess binding

interactions with quercetin and kaempferol. Results indicated minimal interference of c-myc/6x-his tags with the native Cp3GT structure. This study not only lays a foundation for impending crystallographic studies, aiming to solidify the understanding of Cp3GT's stringent 3-O flavonol specificity, but also accentuates the potential of microbial expression platforms and plant metabolic engineering in producing beneficial compounds. To this end, a thorough review of four pivotal classes of plant secondary metabolites, flavonoids, alkaloids, betalains, and glucosinolates, was conducted. This will open avenues for further research and applications in biotechnological, medical, and agricultural domains.

DEDICATION

For Kelsea: “You’re getting a Ph.D.! Oh my word!”

And for Jonah: “What’s an enzyme again?”

ACKNOWLEDGEMENTS

I stand in gratitude, first and foremost, to almighty God, whose boundless grace has allowed me to delve into the wonders of His creation with such profound detail. It is my deepest belief that this dissertation stands as a modest tribute to the grand design of the Creator, and I am blessed to have had the opportunity to contribute to our understanding of this divine blueprint.

To my wife, Kelsea, my heartfelt thanks for the unwavering support and love that has sustained me on this journey to realize my dreams as a scientist. Her strength and companionship have been my refuge in the most challenging of times.

I am eternally grateful to my parents, whose nurturing instilled in me an early love for literature and inquiry. They provided me not only with a home but a sanctuary in the form of the forest surrounding it, where I spent countless hours immersed in the splendor of the natural world.

My advisor, Dr. Cecilia McIntosh, deserves a special note of appreciation for setting an uncompromising standard of scientific excellence. Her guidance has been instrumental in shaping me into the researcher I am today.

To my committee members, I extend my gratitude for their invaluable and constructive feedback, their consistent encouragement, and their unerring ability to steer me towards the right path in my academic endeavors.

Finally, I wish to express my gratitude to all the professors and fellow students in the Department of Biological Sciences at ETSU. Their camaraderie and support have created a nurturing academic environment, a veritable second home where I have flourished.

Each of you has left an indelible mark on my work and on my heart, and for this, I am forever thankful.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Citrus paradisi flavonol 3-O-glucosyltransferase (Cp3GT) is a specialized enzyme that plays a pivotal role in the glucosylation of flavonol compounds, specifically targeting the 3-OH position of flavonols such as kaempferol, quercetin, myricetin, and fisetin (Owens and McIntosh 2009; Birchfield and McIntosh 2020). This enzymatic process is part of a broader biochemical pathway that modulates the solubility, stability, and biological activity of flavonoids, a class of plant natural products known for their diverse functional roles in plants. Cp3GT's unique specificity for flavonols and its 3-O regional specificity make it an intriguing subject for enzymatic studies, shedding light on the selective processes within flavonoid biosynthesis.

Understanding Cp3GT's structure, including its tertiary structure and the amino acid residues involved in its function, offers valuable insights into enzyme design and potential applications in biotechnology and synthetic biology. Flavonols, the substrates for Cp3GT, play essential roles in plants and possess significant pharmacological and therapeutic potential in humans (Kozłowska and Szostak-Węgierek 2022; Daryanavard et al. 2023). Flavonol glucosylation imparts enhanced stability, increases water solubility, and can significantly modulate flavonol bioactivity, thereby expanding their functional versatility both within plants and in potential therapeutic applications (Hyung Ko et al. 2006; Lee et al. 2017; Slámová et al. 2018; Araghi et al. 2023). Therefore, deciphering Cp3GT's structural and enzymatic properties that govern flavonol 3-O glucosylation holds significant implications for both plant biology and human health interventions.

Functional Diversity and Applications of Plant Natural Products

Plant secondary metabolites, commonly referred to as plant natural products, are organic compounds that play pivotal roles in the plant's ecological dynamics (Clemensen et al. 2020).

They provide the plant with protection against herbivores, pathogens, and competition from other plants, assist in reproduction processes, and offer resilience against environmental stressors such as drought, extreme temperatures, and cold conditions (Schweiger et al. 2014; Richardson et al. 2015; Rashidi et al. 2022; Reimer et al. 2022; Q. Zhang et al. 2022). These compounds are categorized as “secondary” metabolites because they do not directly participate in the primary biochemical pathways involved in fundamental functions such as photosynthesis, respiration, growth, development, and reproduction (Bourgaud et al. 2001). Yet they can influence these primary functions in advantageous ways by enhancing the plant's overall fitness and survival in its ecological niche (Erb and Kliebenstein 2020).

These compounds have been the subject of extensive study due to their important roles in plants and their potential utility to humans. Many plant natural products have been found to have medicinal properties (Devi et al. 2020; Swallah et al. 2020). Some are used in traditional medicine practices, while others have been developed into pharmaceutical drugs (Chiocchio et al. 2021). The extraction, purification, and synthesis of these compounds are areas of active research in the fields of pharmacology and organic chemistry (Twaij and Hasan 2022). Plant breeding and genetic engineering are also being used to enhance the production of these valuable compounds in crops (Divekar et al. 2022).

Plant natural products are structurally diverse and can be classified into several major groups, each with unique chemical characteristics and functions. These categories include alkaloids, that are nitrogen-containing compounds often with pharmacological effects; terpenoids or isoprenoids, that are a large group comprising essential oils and pigments; phenolics, that encompass a wide array of compounds like flavonoids and tannins; glycosides, sugar-bonded molecules with varied effects; and lignans and polyketides, with roles in plant cell wall

formation and other functions (Kabera et al. 2014; Chiocchio et al. 2021). This work will focus on the roles that terpenoids, alkaloids, and flavonoids play in both plant and human biology, with a particular focus on the significance of flavonoids.

Terpenoids: Varied Structures and Roles in Plant Protection and Human Health

Terpenoids, also known as isoprenoids, represent a significant group of plant natural products, encompassing over 80,000 identified compounds (Chen et al. 2011; Winnacker 2022). These compounds are derived from isoprene units and exhibit an extensive range of structural diversity, allowing them to participate in numerous biological functions in plants, including defense against herbivores and pathogens, growth regulation, and attracting pollinators (Tholl 2015; Pichersky and Raguso 2018). Carotenoids, categorized as tetraterpenoids, are involved in the synthesis and assembly of protein subunits of photosynthetic reaction centers (Sozer et al. 2010). In humans, terpenoids have been found to possess various pharmacological properties, such as anti-inflammatory, anticancer, antimicrobial, and antioxidant activities (Salehi et al. 2019). Terpenoids have also been utilized as natural flavorings, fragrances, and bioactive agents in the cosmetic, pharmaceutical, and food industries (Tetali 2019; Yang et al. 2020; Abbas et al. 2023).

Alkaloids: Plant Defense Mechanisms and Medicinal Applications in Human Health

Alkaloids constitute another major group of plant natural products, characterized by their nitrogen-containing heterocyclic ring structure (Aniszewski 2015). Produced by a wide variety of plant species, alkaloids have evolved primarily as defense mechanisms against herbivores, insects, and pathogens due to their bitter taste and potential toxicity(Peter J. Facchini 2001; Hotti and Rischer 2017). In human applications, alkaloids have been used as important medicinal

compounds for centuries, with many exhibiting potent pharmacological effects, such as analgesic, anti-inflammatory, and anticancer activities (Cragg and Newman 2013; Heinrich et al. 2021a). Some well-known alkaloids, such as morphine, quinine, and atropine, have been utilized in medicine to treat a variety of conditions, highlighting the significance of this group of plant natural products (Heinrich et al. 2018; Heinrich et al. 2021b; Ti et al. 2021).

Flavonoids: Structure, Biological Activities, and Health Benefits in Focus

This research is focused on a glucosyltransferase (Cp3GT) that has high specificity for the flavonol class of flavonoids, which belong to the phenolic group of plant natural products (Owens and McIntosh 2009; Birchfield and McIntosh 2020). Therefore, particular focus will be placed on the structural diversity, biological activities, and potential health benefits of flavonoids. Flavonoids are ubiquitously present in plants, with over 10,000 naturally occurring compounds identified thus far (Kumar and Pandey 2013; Panche et al. 2016; Shen et al. 2022). A typical flavonoid comprises two aromatic rings, designated A and B, interconnected by a third ring, designated as the C ring (Fig. 1.1).

This ring features an oxygen atom and can have various oxidation states. One or more hydroxyl groups are present, serving as essential sites for nucleophilic substitution (Harborne 1995; Procházková et al. 2011; Pinheiro et al. 2012). The selectivity and activity of different flavonoids are determined by their structure and the location of their respective hydroxyl groups (Procházková et al. 2011; Pinheiro et al. 2012). For instance, the flavanone hesperetin possesses hydroxyl groups at the C5 and C7 positions, forming 7-O-glucosides, while the flavonol quercetin features hydroxyl groups at the C3, C5, and C7 positions, typically forming 3-O-glucosides (Owens and McIntosh 2009; Owens and McIntosh 2011; Birchfield and McIntosh 2020)

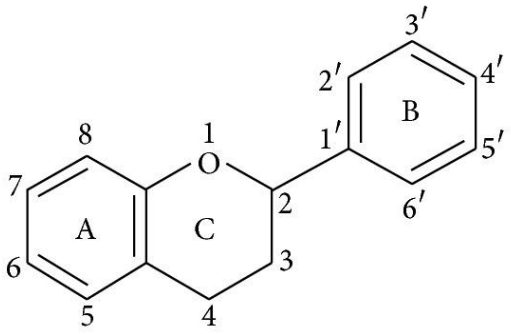

Fig. 1.1 The Fundamental Structure of a Flavonoid. Defined by a 15-carbon skeleton organized into three connected rings: A, B, and C. Rings A and C form a fused ring structure (cyclohexane-cyclopentane), and are numbered 1 to 8, initiating at the oxygen-attached carbon 1 in Ring C, which is connected to carbon 2 of Ring C, and proceeding clockwise. Ring B, an aromatic (benzene) structure, branches off from this fused structure, and is numbered from 1' to 6', with the numbering starting from the carbon attached to carbon 2 of Ring C, and moving clockwise. Used with permission (Kumar and Pandey 2013).

Flavonols: Biological Significance and Interaction with Cp3GT

Among the numerous classes of flavonoids, flavonols hold relevance for the present study, as they are the specific substrates for Cp3GT. Flavonols are chemically characterized by a ketone group at the 4th position, a hydroxyl group at the 3rd position, and a C2=C3 double bond in the C-ring, distinguishing them from other flavonoid subclasses. (Fig. 1.2) (Harborne and Williams 2000; Schutte-Smith et al. 2019; De Souza Farias et al. 2021). The presence and position of additional hydroxyl or methoxy groups on the A and B rings further determine the specific flavonol compound and its biological activity (Owens and McIntosh 2011).

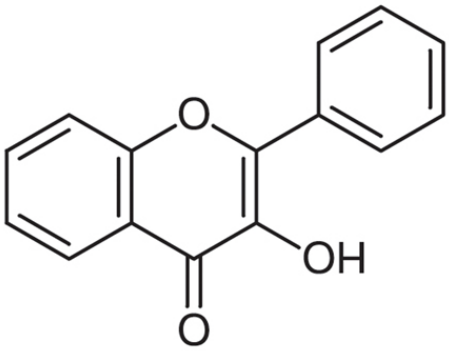

Fig. 1.2 The Fundamental Structure of a Flavonol (used with permission) (NIH National Library of Medicine)

Among the numerous flavonols, Cp3GT preferentially glucosylates quercetin, kaempferol, myricetin, and fisetin (Owens and McIntosh 2009; Devaiah et al. 2016; Devaiah et al. 2018; Birchfield and McIntosh 2020). Quercetin, widely found in leaves and fruits of vegetables such as apples, onions, and berries, plays a multifaceted role in plants by enhancing seed sprouting, supporting pollen development, boosting antioxidant defenses, and optimizing photosynthesis, thereby contributing to overall growth and resilience against environmental challenges (Singh et al. 2021). In humans, quercetin has shown numerous health benefits, one of which includes reducing diabetic complications by stimulating β-cell proliferation, thereby improving glucose metabolism and insulin secretion (Adewole et al. 2007; Lopez-Lazaro 2009; David et al. 2016; Salehi et al. 2020). Structurally, quercetin has hydroxyl groups at the C3, C5, C7, C3', and C4' positions, which contribute to its biological activity.

Kaempferol, present in foods such as broccoli, kale, and tea, acts as an indirect growth regulator in plants by coordinating with the phytohormones abscisic acid (ABA), salicylic acid (SA), jasmonic acid (JA), ethylene (ET), and auxin during stress conditions, regulating ROS

signaling, and modulating hormonal signaling, thus contributing to the control of plant development under abiotic stresses such as salt, drought, and heat (Jan et al. 2022). Like quercetin, kaempferol has numerous health benefits in humans, one of which includes inhibiting low-density lipoprotein (LDL cholesterol) oxidation, which plays a crucial role in the formation of atherosclerotic plaques (Linton et al. 2019). By scavenging reactive oxygen species (ROS), kaempferol slows down the chain of actions that may lead to coronary heart complications (Harborne 1995; Hou et al. 2004; Lin et al. 2007). It is structurally like quercetin but lacks the hydroxyl group at the C3' position (M. Calderon-Montano et al. 2011).

Myricetin, found in sources such as berries, grapes, and herbs, acts as a potent antioxidant, protecting plant cells from oxidative stress, serves as a UV-B radiation absorbent to shield plant tissues from harmful ultraviolet rays, exhibits antimicrobial properties to defend against pathogens, and may deter herbivorous insects, functioning as a natural insecticide (Li et al. 1993; Schmidt et al. 2011; Agati et al. 2012). In humans, myricetin is attributed with numerous health benefits, including the inhibition of hyperglycemia, reduction of hepatic triglycerides and oxidative stress, decrease in cholesterol levels, and protection against liver injury (Chang et al. 2012; Choi et al. 2014; Guo et al. 2015; Semwal et al. 2016). It possesses hydroxyl groups at the C3’, C4’, and C5' positions (Semwal et al. 2016).

Fisetin, prevalent in various fruits and vegetables such as strawberries, apples, and onions, plays a similar role in plant defense as other flavonols (Harborne 1995). In the context of human health, fisetin has garnered attention for its multifaceted biological activities (Park et al. 2022). Specifically, fisetin has emerged as a potent senotherapeutic agent, demonstrating the capacity to reduce senescence markers in various tissues, restore tissue homeostasis, mitigate age-related pathologies, and notably, extend median and maximum lifespan in mice, as well as

exhibiting senolytic activity in human tissues, thereby spotlighting its potential as a viable candidate for human clinical studies aimed at health span enhancement (Yousefzadeh et al. 2018). Structurally, fisetin is characterized by the presence of hydroxyl groups at the C4' and C5' (Viñas et al. 2011).

Citrus Flavonoids: Vital Role in Plant Physiology, Human Health, and Citrus Economy

Cp3GT was isolated from the citrus species *Citrus paradisi*, commonly referred to as grapefruit (Owens and McIntosh 2009). The citrus industry is a key contributor to agricultural economies worldwide, with mandarins, oranges, lemons, limes, and grapefruits commanding significant market share both in domestic and international markets. In fact, the Food and Agriculture Organization of the United Nations reported an annual global production of citrus fruits surpassing 124 million tons as of 2021, worth approximately 3.4 billion dollars, underscoring its economic importance (FAO 2021).

Citrus species are renowned for their rich diversity of flavonoids, hosting an estimated 5,000 distinct molecular variations (Harborne 1995; Berhow et al. 1998; S. Wang et al. 2017). Notably, the seeds, peels, and juice extracts of these fruits serve as substantial reservoirs for these bioactive compounds. (Daniel et al. 2011; Owens and McIntosh 2011; Addi et al. 2022). Flavonoids are intrinsic to the physiological robustness and ecological sustainability of citrus plants. These bioactive compounds fortify the plant's defensive mechanisms by exhibiting antimicrobial and antifungal activities, thereby reducing the likelihood of disease outbreaks that could compromise citrus yields (Kumar et al. 2022; Salawu et al. 2022). For example, studies have demonstrated that upon irradiation, *Citrus aurantium* exhibits elevated accumulation of the polymethoxyflavone tangeretin, suggesting a protective response against ultraviolet light-induced damage (Arcas et al. 2000). Moreover, when applied as pesticides, the citrus flavonoids

rutin and isoquercetin inhibited the growth of larval worms within the fruit and heart tissues of tobacco, showcasing their potential as biopesticides in agricultural applications (Elliger et al. 1980; Duffey and Isman 1981).

Citrus flavonoids have attracted considerable attention due to their various health benefits in humans (Mahmoud et al. 2019; Alam et al. 2022; Anbualakan et al. 2022; Sebghatollahi et al. 2022). These bioactive compounds, including hesperidin, naringin, and quercetin, among others, have been reported to prevent bone loss in post-menopausal women, positively mitigate metabolic syndrome, and reduce risk of illness after intense exercise (Nieman 2007; Horcajada et al. 2008; Alam et al. 2014; Konrad and Nieman 2015). Furthermore, some citrus flavonoids, capable of crossing the blood-brain barrier, can directly safeguard the brain through their antioxidant properties, their ability to attenuate inflammatory responses, regulate neuronal metabolism, and stimulate neuronal regeneration, while also indirectly fortifying neuroprotection by influencing gut microbiota composition and metabolites, thereby impacting the gut-brain axis function (Dey 2019; Josiah et al. 2022; Wang et al. 2023).

Citrus fruits, celebrated for bolstering plant health, offering myriad human health benefits, and driving economic growth, stand as a pivotal topic in scientific exploration. Considering the vast potential of citrus flavonoids, delving deeper into the specific interactions of Cp3GT with flavonols becomes of paramount importance. Such inquiries not only promise to expand the knowledge on the intricate biochemistry of citrus flavonoids and their manifold effects on plant and human health but also portend significant implications for the continued success of the citrus industry.

Deciphering the Intricacies of Flavonoid Biosynthesis: Steps and Modifications

Flavonoid biosynthesis is a complex and highly regulated process (Fig. 1.3). It starts with the phenylpropanoid pathway, responsible for producing a wide range of plant natural products. The first step in this pathway is the deamination of phenylalanine, to form cinnamic acid, catalyzed by the enzyme phenylalanine ammonia-lyase (PAL) (Neish 1960; Koukol and Conn 1961). Next, cinnamic acid undergoes a series of hydroxylation and O-methylation reactions, catalyzed by cytochrome P450-dependent cinnamic acid 4-hydroxylase (C4H), to first produce coumaric acid, and 4-coumarate:CoA Ligase (4CL), to then produce coumaroyl-CoA (Russell and Conn 1967; Hahlbrock and Grisebach 1970; Russell 1971; Hamberger and Hahlbrock 2004).

The committed step towards flavonoid biosynthesis is the formation of chalcone, which is catalyzed by chalcone synthase (CHS), a type III polyketide synthase that uses p-coumaroyl-CoA and three molecules of malonyl-CoA as substrates (Kreuzaler and Hahlbrock 1972; Kreuzaler and Hahlbrock 1975). Chalcone is then converted to naringenin, a flavanone, by chalcone isomerase (CHI) (Moustafa and Wong 1967; Boland and Wong 1975; Bednar and Hadcock 1988). The flavanone naringenin is a key biochemical branch point that serves as a central precursor to produce various flavonoid subclasses, including flavones, flavonols, dihydroflavonols, and anthocyanidins (Owens and McIntosh 2011).

Naringenin is hydroxylated by flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) to produce eriodyctiol and pentahydroxyflavanone, respectively (Britsch et al. 1981; Stotz and Forkmann 1982; Britsch and Grisebach 1986). Flavone synthase (FNS) catalyzes the formation of flavones from flavanones, while flavonol synthase (FLS) converts flavanones to flavonols (Sutter et al. 1975; Britsch et al. 1981; Spribille and Forkmann 1984; Forkmann et al. 1992; Martens et al. 2001). Dihydroflavonol 4-reductase (DFR) and

anthocyanidin synthase (ANS) are responsible for the production of anthocyanidins from dihydroflavonols (W. Heller et al. 1985; Werner Heller et al. 1985; Harborne 1995). Anthocyanidins are generated through the action of leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR) (Tanner et al. 2003; Xie et al. 2003). Finally, various modifications such as glycosylation, acylation, and methylation take place on the flavonoid backbone, resulting in a wide array of flavonoid structures (Owens and McIntosh 2011; Alseekh et al. 2020).

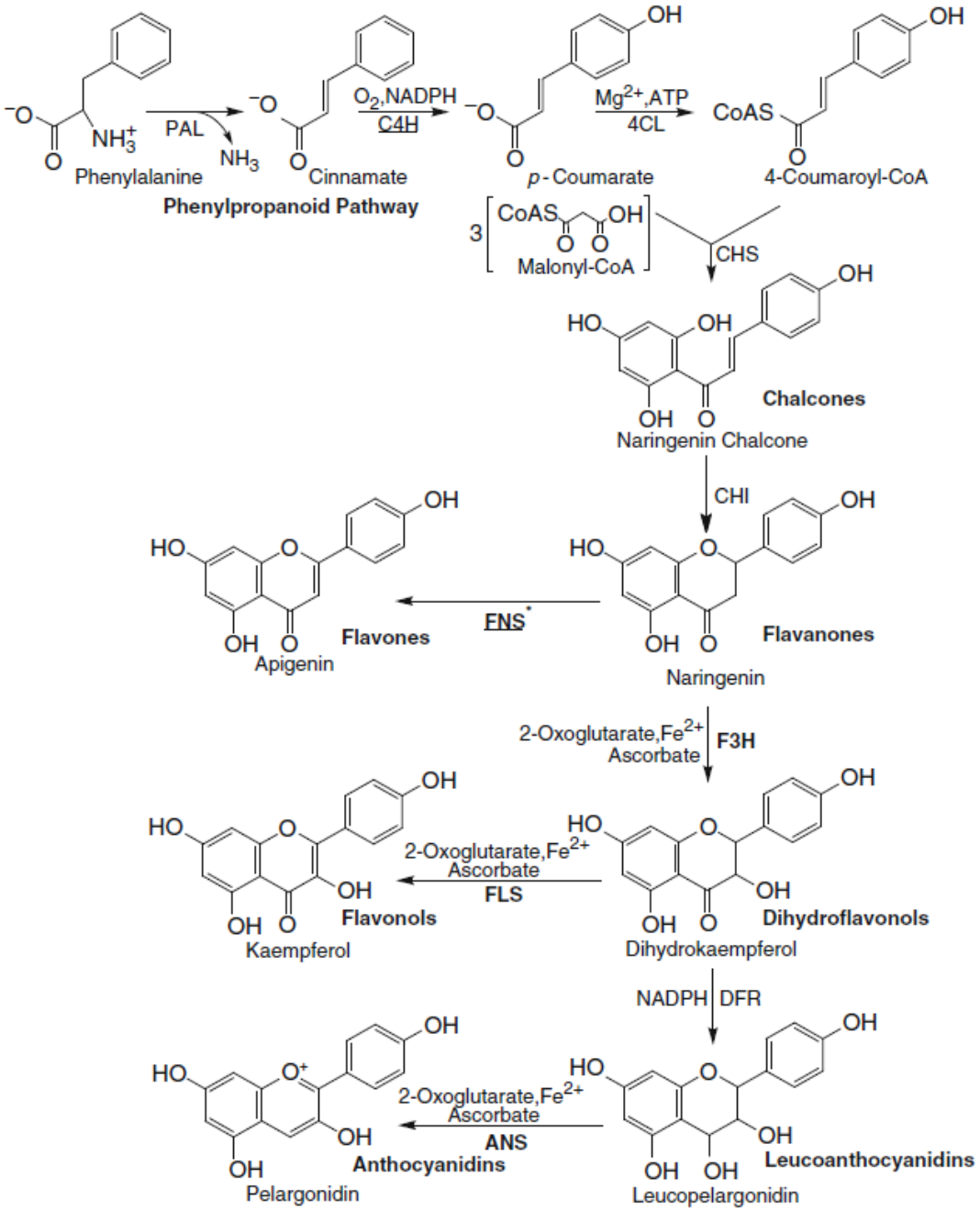

Fig. 1.3 The Central Flavonoid Biosynthetic Pathway. In the Fig., several key enzymes are represented: Phenylalanine ammonia lyase (PAL), Cinnamic acid 4-hydroxylase (C4H), 4-coumarate: CoA ligase (4CL), Chalcone synthase (CHS), Chalcone isomerase (CHI), Flavone synthase (FNS), Flavanone 3-hydroxylase (F3'H), Flavonol synthase (FLS), Dihydroflavonol 4-

reductase (DFR), and Anthocyanidin synthase (ANS). Used with permission (Owens and McIntosh 2011).

Naringenin is hydroxylated by flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) to produce eriodyctiol and pentahydroxyflavanone, respectively (Britsch et al. 1981; Stotz and Forkmann 1982; Britsch and Grisebach 1986). Flavone synthase (FNS) catalyzes the formation of flavones from flavanones, while flavonol synthase (FLS) converts flavanones to flavonols (Sutter et al. 1975; Britsch et al. 1981; Spribille and Forkmann 1984; Forkmann et al. 1992; Martens et al. 2001). Dihydroflavonol 4-reductase (DFR) and anthocyanidin synthase (ANS) are responsible for the production of anthocyanidins from dihydroflavonols (W. Heller et al. 1985; Werner Heller et al. 1985; Harborne 1995). Anthocyanidins are generated through the action of leucoanthocyanidin reductase (LAR) and anthocyanidin reductase (ANR) (Tanner et al. 2003; Xie et al. 2003). Finally, various modifications such as glycosylation, acylation, and methylation take place on the flavonoid backbone, resulting in a wide array of flavonoid structures (Owens and McIntosh 2011; Alseekh et al. 2020).

Glycosylation involves the addition of sugar moieties to flavonoid molecules, primarily at hydroxyl groups. For instance, quercetin can be glycosylated to form quercetin-3-O-glucoside (isoquercetin), which exhibits increased solubility, stability, and biological activity compared to its aglycone form (Xie et al. 2022; Araghi et al. 2023). Acylation, which is the attachment of acyl groups (such as acetyl, malonyl, or coumaroyl groups) to flavonoids, can also affect their properties (Luo et al. 2022). An example is cyanidin-3-O-(6"-O-malonyl)-glucoside, an anthocyanin derivative with enhanced color and bioavailability due to acylation (Sasaki et al. 2021; Ichiyanagi et al. 2022; P. Zhang et al. 2022). Methylation, the addition of a methyl group

to flavonoid molecules, can alter their properties as well. Kaempferol, for example, can be methylated to form isorhamnetin, which exhibits increased stability and bioactivity (Koirala et al. 2016; Bangar et al. 2022).

Unraveling the Impact of Glycosylation: Influence on Flavonoid Solubility, Stability, and Bioactivity

Glycosylation is a crucial post-translational modification of natural products in plants, affecting their solubility, stability, bioactivity, and transport (Le Roy et al. 2016). It involves the attachment of sugar moieties, such as glucose, to a wide variety of compounds, including flavonoids. Glycosylation can be classified into different types, based on the types of sugars and linkages involved. The major types of glycosylation include O-glycosylation, N-glycosylation, and C-glycosylation. O-glycosylation, a process whereby a sugar molecule forms a bond with the hydroxyl group of an acceptor molecule, constitutes the specific type of glycosylation that Cp3GT catalyzes.

Flavonoid glycosylation holds significant implications for plant biology. Glycosylation often enhances the water solubility, stability, and compartmentalization of these compounds within plant cells, which can influence their metabolic functions and roles in plant physiology (Dias et al. 2021). For example, glycosylated flavonoids like quercetin-3-O-glucoside show enhanced water solubility compared to their aglycone forms (Jan et al. 2022). This increased solubility could amplify the biological roles of these compounds, such as their participation in UV protection, signaling processes, and defense mechanisms against pathogens, thereby boosting the plant's overall health, resilience, and adaptability in various ecological contexts (Ferreyra et al. 2012).

Furthermore, glycosylation can influence the interactions of plant-derived compounds with human enzymes, receptors, and transporters, which in turn can impact their pharmacological properties and therapeutic potential (Zhang et al. 2020; Zhu et al. 2020; Xie et al. 2022). Various glycosides of cyanidin and quercetin were found to utilize hexose transporters SGLT1 and GLUT2 for cellular uptake, and this was not observed with aglycones (Zhang et al. 2020). This interaction significantly reduced the cellular uptake of D-glucose, suggesting potential therapeutic implications of these glycosides for blood sugar regulation in diabetes treatment and potential for lowering blood glucose levels (Zhang et al. 2020). Likewise, naringin, a glycosylated flavonoid in citrus fruits, outperforms its aglycone counterpart in inhibiting α-glucosidase (Zhu et al. 2020). Given that α-glucosidase suppression is a prevalent strategy in managing diabetes, this underscores the significant role of glycosylated flavonoids in promoting human health.

Within the context of human dietary consumption, a recent study examined the impact of glycosylation on the antioxidant activities and metabolic stability of flavonoids by contrasting quercetin 3-O-glucoside and luteolin 8-C-glucoside with their aglycones (quercetin and luteolin). Initial findings showed that glycosylated flavonoids possessed notably reduced activity compared to their aglycones (Xie et al. 2022). However, during *in vitro* digestion, these glycosylated compounds exhibited a marked increase in activity, surpassing their aglycone counterparts (Xie et al. 2022). This suggests that while glycosylation may diminish the innate activity of flavonoids, it enhances their stability and solubility throughout the digestion process, leading to a more consistent and sustained functional presence.

Glycosyltransferase Enzymes: Evolution, Structural Diversity, and Functional Implications

The exploration of glycosyltransferase distribution, evolution, structure, and classification presents an important context for Cp3GT research. According to the Carbohydrate-Active Enzymes (CAZy) database, GTs have been divided into over 116 families based on sequence homology and structural similarities (Drula et al. 2022). This vast and diverse classification is indicative of the extensive evolutionary processes that have shaped these enzyme families. Each GT family has evolved in response to the selective pressures of biological systems, demonstrating specificity for a diverse range of donor and acceptor substrates.

Understanding the Evolution and Diversification of Glycosyltransferases

The diversification and evolution of glycosyltransferases are driven by a combination of factors related to the biological and functional diversity of these enzymes, as well as genetic mechanisms such as mutation, natural selection, gene duplication, and horizontal gene transfer (Yonekura-Sakakibara and Hanada 2011; Taujale et al. 2020; Záveská Drábková et al. 2021). This process has contributed to the evolution of new GTs with novel substrate specificities or catalytic mechanisms. For instance, the divergence of the plant UGT family has been ascribed to ancient gene duplication events, enabling the development of a plethora of plant natural products essential for plant defense and adaptation (Krishnamurthy et al. 2020; Xu et al. 2020). .

Evolution has resulted in GTs adapted to various environmental pressures, resulting in the co-evolution of enzyme function with substrate availability (Taujale et al. 2020). The presence of specific GTs in certain organisms but not in others is likely indicative of different evolutionary pressures based on environmental conditions and survival strategies. For example, the ability to produce anthocyanin glycosides, responsible for the red, purple, and blue

pigmentation in plant tissues, has evolved independently in various plant families in response to different environmental pressures (Saigo et al. 2020).

Additionally, the formation of glycosylated flavones, flavonols, and isoflavonoids by specific GTs in different plant species reflects their distinct ecological niches and evolutionary trajectories (Yonekura-Sakakibara et al. 2019; Wen et al. 2020). For instance, soybeans (*Glycine max*) and other legume plants have evolved the ability to produce isoflavone glucosides using isoflavone-specific GTs, contributing to their symbiotic relationship with nitrogen-fixing bacteria (Kim 2022). Thus, the evolution of GTs provides a fascinating look at the co-evolution of enzyme structure and function. It reveals how these enzymes have become adapted to perform precise and complex glycosylation reactions essential for the survival and interaction of organisms with their environment. This evolutionary understanding is integral to the ability to manipulate GTs for biotechnological and therapeutic applications.

Glycosyltransferase Structural Properties and Classification

The structural framework of glycosyltransferases is marked by the presence of a conserved domain, which forms an integral part of their architecture and is pivotal for the binding of nucleotide-sugar donors (Fig. 1.4). This core domain, also known as the glycosyltransferase-B (GT-B) fold, primarily consists of two Rossmann-like α/β/α sandwich domains that hold the donor sugar nucleotide (Breton et al. 2006; Tegl and Nidetzky 2020). In plants, this conserved domain is called the Plant Secondary Product Glucosyltransferase Box (Fig. 1.5) (Hughes and Hughes 1994; Mackenzie et al. 1997). However, the structural elements of GTs are not completely conserved across all classes. The variations in structure across different GT classes stem from the diversity in their substrates and the complexity of their catalytic mechanisms. This diversity has led to the categorization of GTs into three primary

structural classes, GT-A, GT-B, and GT-C, which are characterized by distinct fold types, spatial arrangements of key catalytic residues, and subcellular localization (Lairson et al. 2008; Tegl and Nidetzky 2020; Bohl et al. 2021).

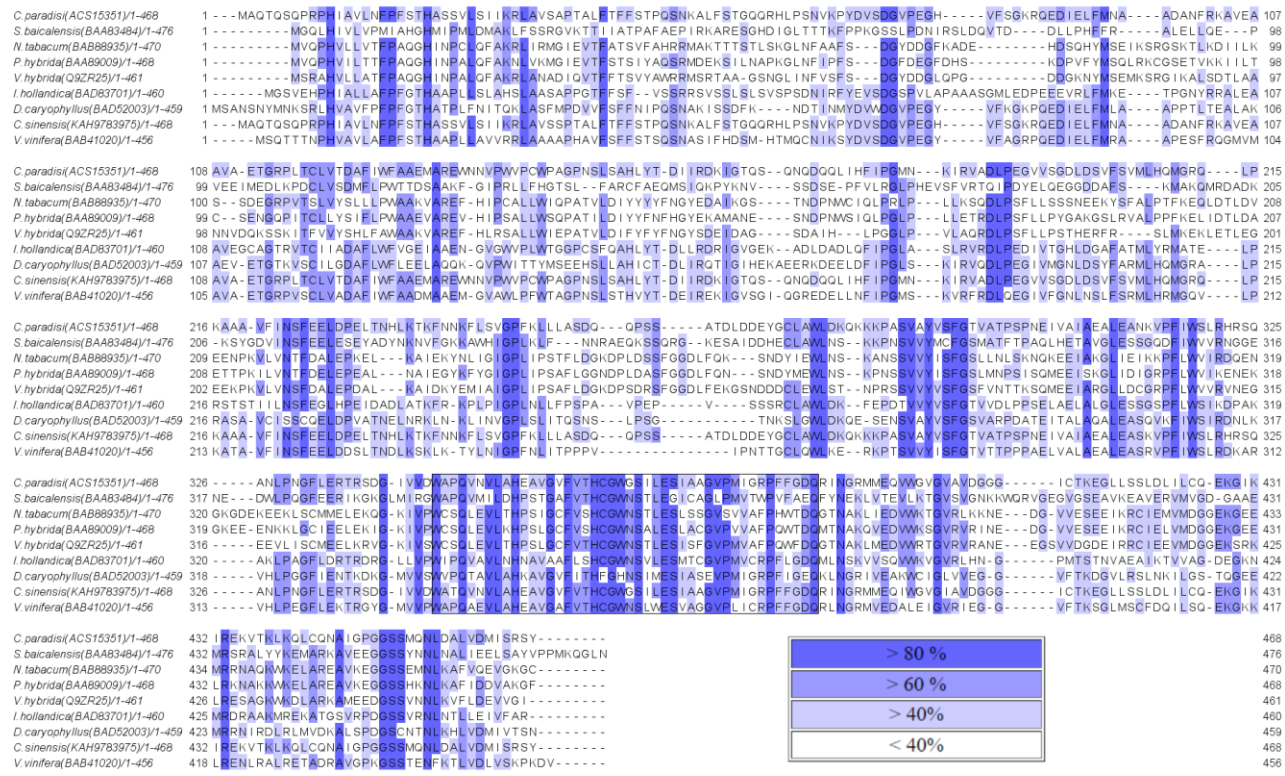

Fig. 1.4 Multiple Sequence Alignment of Flavonol-3-O GTs with Known Biochemical Function. The conserved PSPG box is outlined in red. Percent identity color scheme shown bottom right. Adapted from Owens and Mcintosh 2009 with permission (Owens and McIntosh 2009).

GT-A enzymes possess a GT-A fold, comprising a single Rossman-like domain with a characteristic α/β/α sandwich structure (Lairson et al. 2008). They typically employ a metal-dependent catalytic mechanism and have a conserved Asp X Asp motif in their active sites. UGT72E1 is a GT-A glycosyltransferase from *Arabidopsis thaliana* that participates in the

glycosylation of monolignols and is thought to play a role in the storage and transport of these compounds within the plant (Lim et al. 2005).

GT-B enzymes encompass a two-domain structure, each containing a Rossmann-fold, and a deep cleft in between the domains forms the active site. They operate via an S_N2 (nucleophilic substitution with inversion of stereochemistry) mechanism. An illustrative example of GT-B enzymes is Cp3GT from *Citrus paradisi*, which catalyzes the 3-O glucosylation of flavonols (Owens and McIntosh 2009; Birchfield and McIntosh 2020).

The GT-C superfamily is an emerging class of membrane-associated glycosyltransferases characterized by their role in catalyzing the stepwise synthesis of oligosaccharides, or the transfer of assembled glycans from lipid-linked donor substrates to acceptor proteins (Bai and Li 2021). Recent structural insights have revealed a diversity of folds within the GT-C superfamily, leading to the proposal of two subclasses: GT-CA and GT-CB. The GT-CA class retains the conserved core module of seven transmembrane helices, while the GT-CB class includes enzymes with distinct folds that lack the same conserved module (Bohl et al. 2021). The yeast ALG6 enzyme is a GT-C glycosyltransferase involved in ER-luminal N-glycan synthesis, transferring the first of three glucose moieties onto the pre-assembled GlcNAc2Man9 glycan. In humans, deficiencies in ALG6 are a frequent cause of congenital disorders of glycosylation (CDGs), in which patients have hypo-glycosylated serum glycoproteins (Haeuptle and Hennet 2009).

The classification of glycosyltransferases, including those found in plants, is based on sequence and structural similarities, and is overseen by the Carbohydrate-Active enZymes (CAZy) database (Drula et al. 2022). The plant glycosyltransferases are mostly grouped into the GT1, GT20, and GT28 families, reflecting the diversity and specialization of these enzymes in

plant species (Drula et al. 2022). It should be noted, however, that despite the considerable progress in the identification and classification of plant glycosyltransferases, the understanding of their structural and functional characteristics, as well as their regulation, is still evolving. Despite the extensive annotation of glucosyltransferases in databases like CAZy, a substantial gap persists as only a limited subset of these enzymes have undergone rigorous biochemical characterization, leaving a vast array of them with predicted but unverified functions and mechanistic pathways in plant biology. Therefore, ongoing research in this area promises to unveil new insights into the complex world of plant glycosyltransferases.

Functional Implications of Plant Glycosyltransferase Diversity

In plants, the glycosyltransferase family of enzymes showcases a striking balance of homogeneity and heterogeneity, a testament to the evolutionary forces that have sculpted their functional characteristics (Gachon et al. 2005). The presence of a conserved domain responsible for nucleotide-sugar donor binding across plant GTs, known as the Plant Secondary Product Glycosyltransferase (PSPG) box, reflects a core principle of the glycosylation process (Hughes and Hughes 1994; Mackenzie et al. 1997; Gachon et al. 2005; Lairson et al. 2008). Yet, outside this conserved region, considerable structural variation arises, contributing to the extensive repertoire of GTs' substrate and positional specificity.

The promiscuity of certain GTs further augments this functional diversity. Several plant GTs exhibit a broad substrate range, accommodating various classes of plant natural products. For example, UGT71G1 from *Medicago truncatula* (Achnine et al. 2005; Modolo et al. 2007), GuGT14 and GuGT33 from *Glycyrrhiza uralensis* (licorice) (K. Chen et al. 2019), and UGT73AE1 from *Carthamus tinctorius* (Xie et al. 2014) are plant GTs that facilitate the glycosylation of various phenylpropanoids (including coumarins, coumarones, flavones, and

isoflavones) as well as terpenoids (specifically glycyrrhetinic acid and glycyrrhizic acid). The plant glycosyltransferases GuGT10 and GuGT14, derived from *Glycyrrhiza uralensis*, were biochemically shown to glucosylate 62 licorice compounds, encompassing chemical classes such as coumarins, coumarones, pterocarpans, flavones, isoflavones, isoflavans, and chalcones (K. Chen et al. 2019).

Conversely, certain GTs have evolved to exhibit a stringent substrate and positional specificity. Cp3GT from *Citrus paradisi* exclusively glucosylates the flavonols quercetin, kaempferol, myricetin, and fisetin at the 3-OH position (Owens and McIntosh 2009; Birchfield and McIntosh 2020). The glycosyltransferase UGT78K6 from *Clitoria ternatea* exhibits strict specificity for the anthocyanidin delphinidin (Hiromoto et al. 2015). Additionally, both Rd3GT1 and Rd3GT6 from *Rhododendron delavayi* facilitate the addition of UDP-sugar to the 3-OH position of anthocyanidin. They exhibit a preference for UDP-Gal as their sugar donor and demonstrate the highest efficiency with cyanidin as the substrate (Sun et al. 2022). The co-existence of such highly specialized GTs alongside more promiscuous counterparts highlights the dynamic interplay of conservation and diversification in the evolutionary landscape of these enzymes.

Unraveling the functionality of GTs from primary sequence data is a complex endeavor, due to these overlapping structural characteristics and the spectrum of possible substrates. Hence, the necessity for comprehensive experimental validation of putative GT functions is paramount, employing techniques such as *in vitro* biochemical assays or complementation experiments (McIntosh and Owens 2016). Furthermore, enhancing these approaches with crystallographic studies can provide intricate details about the enzyme-substrate interactions and active site architecture, thereby offering a deeper understanding of the mechanistic aspects of GT catalysis

(Shao et al. 2005; Offen et al. 2006; Hiromoto et al. 2013; Hiromoto et al. 2015). These methods facilitate not only the confirmation of GT functions but also the exploration of structure-function relationships of GT enzymes. This rich tapestry of structural and functional diversity underscores the integral role of GTs in plant metabolism and survival strategy, shedding light on their evolutionary journey and offering potential applications in plant biotechnology and medicine.

Advancements in Biotechnology through Glycosyltransferase Research

In recent years, research into glycosyltransferases has illuminated promising applications in biotechnology, particularly in the fields of medicine, agriculture, and industrial production. GTs' capacity to modify a variety of bioactive compounds renders them powerful tools in drug development. For example, glycosylation can enhance the stability, solubility, and bioavailability of pharmaceutical compounds, and facilitate targeted drug delivery (Sola and Griebenow 2009; Grimsey et al. 2020). Some antibiotic substances, such as erythromycin and vancomycin, are known to improve their activity and reduce their toxicity upon glycosylation (Grimsey et al. 2020). Similarly, GTs play a crucial role in the production of glycosylated anticancer drugs, such as doxorubicin, which requires specific GTs for its biosynthesis (Mohideen et al. 2022)

In the field of agriculture, GT engineering offers promising avenues for crop improvement and protection by enhancing pest resistance, improving nutritional value, and positively mitigating environmental stressors (Yang et al. 2021; Zhou and Jander 2021; Gharabli et al. 2023). For example, the expression of glucosinolate biosynthesis genes derived from *Manihot esculenta* (cassava) in Arabidopsis increased resistance to the bacterial pathogens *Erwinia carotovora* and *Pseudomonas syringae* (Brader et al. 2006). Furthermore, by overexpressing AtUGT76C2, a cytokinin glycosyltransferase, in rice (*Oryza japonica*), cytokinin homeostasis is modified (Y. Li et al. 2020). This leads to improved root growth, heightened

adaptation to stresses such as drought and salt across development stages, decreased ion leakage during salt stress, and a significant upregulation of eight stress-responsive genes (Y. Li et al. 2020).

Industrial biotechnology also greatly benefits from GT research. GTs are used in the food and beverage industry for the modification of flavor compounds and the production of prebiotic oligosaccharides (Amin et al. 2021; Vera et al. 2021). For instance, the thermostability of glycogen synthases isolated from hyperthermophilic Archaea such as *Sulfolobus acidocaldarius* and *Thermococcus hydrothermalis* could be leveraged to catalyze polymerization reactions at high temperatures, offering potential bioengineering applications for the selective synthesis of various oligosaccharides and polysaccharides (Gruyer et al. 2002). GTs have also found application in the production of biofuels, where they help in the modification of lignocellulosic biomass to improve its fermentability (Amin et al. 2021).

Despite these advances, challenges remain in harnessing the full potential of GTs in biotechnology, due to their often-stringent donor and acceptor substrate specificities and the necessity for expensive nucleotide sugars. To overcome these issues, ongoing research is exploring enzyme engineering strategies to alter GT specificity and the use of inexpensive sugar donors (McArthur and Chen 2016; H. Chen et al. 2022; Guidi et al. 2023). Taken together, these findings underscore the immense potential and diverse applications of GTs in biotechnology, ranging from medicinal and therapeutic uses to agriculture and industrial production. Continued investigation into the biochemical properties and functional diversities of GTs is crucial to fully leverage their potential in biotechnological applications.

Unveiling Glycosyltransferases Through Structural Analysis: A Foundation for Biotechnological Applications

In the pursuit of understanding the Cp3GT enzyme at the heart of this research, obtaining structural information is critical to elucidate its catalytic mechanisms, determine its substrate specificity, and uncover the precise molecular interactions that govern its functionality. Structural insights gained from techniques such as site-directed mutagenesis, biochemical characterization, *in silico* modeling, and potentially x-ray crystallography, pave the way for rational design of Cp3GT variants with specific traits, advancing its potential biotechnological applications. (McIntosh and Owens 2016; Bi et al. 2022; M. Chen et al. 2022; H. Chen et al. 2022). Utilizing well-established enzyme characterization strategies enables the acquisition of intricate insights into Cp3GT's structure-function relationships and interactions with various substrates (Devaiah et al. 2016; McIntosh and Owens 2016; Devaiah et al. 2018). This understanding can serve as a launching pad for further investigation into GT mechanistic details and potential biotechnological applications.

Identification and Optimization of Cp3GT Expression

Cp3GT was previously identified through the design of primers against N and C-terminal residues of an annotated anthocyanidin 3-O GT (KAH9783975) from blood orange (*C. sinensis*). Amplification of the grapefruit equivalent (GQ141630) from young grapefruit leaf cDNA was followed by Topo-TA cloning and sequencing analysis (Owens and McIntosh 2009). The resulting protein, commonly referred to as Cp3GT, had a theoretical molecular mass of 51.2 kDa, which falls within the typical 45–60 kDa range observed for most plant GTs. Subsequently, GQ141630 was expressed as a recombinant fusion protein in *Escherichia coli* for in-depth biochemical characterization and analysis using the pET protein expression system (Owens and

McIntosh 2009). Utilizing the pCD1 vector, a modified form of pET32a, allowed for the advantages of an N-terminal thioredoxin fusion partner, enhancing protein solubility, and the option to eliminate all vector-encoded tags through digestion of the LeuValProArg↓GlySer cleavage site with thrombin (Owens et al. 2008; Owens and McIntosh 2009). This expression yielded a fusion protein of the appropriate mass (62 kDa), confirmed by SDS–PAGE and western blotting (Owens and McIntosh 2009). However, most of the target protein was found in the insoluble *E. coli* fraction, a characteristic observed with several recombinant glucosyltransferases (Mital et al. 2021). Consequently, Cp3GT was cloned into the pPICZα plasmid and transformed into the methylotrophic yeast *Pichia pastoris* (Devaiah et al. 2016).

The transition from the *Escherichia coli* expression system to the pPICZα system in *Pichia pastoris* proved to be a strategic decision in the production of Cp3GT. Characterized by its strong, methanol inducible AOX promoter, it showed high-level expression without losing Cp3GT to insoluble inclusion bodies (Birchfield 2019; Birchfield and McIntosh 2020). Notably, the fusion tags also differ between the two systems. While the *E. coli* system employed an N-terminal thioredoxin fusion tag, the *P. pastoris* system utilizes a C-terminal c-myc/6x His combo tag, allowing for more versatile applications including detection and purification.

To Cleave or Not to Cleave: Utilizing Recombinant Tags in Enzyme Research

Vector encoded tags have become quintessential tools in modern molecular biology and biotechnology, each with unique applications (Kosobokova et al. 2016). The c-myc tag, derived from the human c-Myc oncogene product, has found extensive use in applications such as western blotting and immunoprecipitation due to its specific antibody recognition (Zhao et al. 2013). The 6x His tag, consisting of a sequence of six histidine residues, has become an essential tool in immobilized metal affinity chromatography (IMAC), allowing for efficient protein

purification by selectively binding to nickel or cobalt ions (Hochuli et al. 1988; Kielkopf et al. 2020).

These tags offer substantial advantages, particularly in purification and detection. The 6x His tag is known for its protein purification capabilities, allowing for efficient separation from other cellular proteins, a method that has been well-established since its inception (Hochuli et al. 1988; Kielkopf et al. 2020). Conversely, the c-myc tag is primarily used for specific detection and analyses of tagged proteins using western blot and anti c-myc antibodies (Zhao et al. 2013; Kosobokova et al. 2016). However, they are not without potential drawbacks. Some studies have noted that the presence of these tags might alter substrate binding or even catalytic efficiency, affecting protein folding or functionality (Waugh 2005; Young et al. 2012). Yet, other reports found no interference with enzyme activity, emphasizing the importance of careful evaluation in specific contexts (Waugh 2005).

In the realm of enzymology, post-purification cleavage of these tags is often performed to assess the enzyme in its native form (Arnau et al. 2006; Bell et al. 2013). This is crucial for applications such as crystallization, where the native structure is desired for accurate characterization (Bell et al. 2013). A prevalent technique involves incorporating the thrombin recognition sequence, LeuValProArg↓GlySer, between the tags and the recombinant enzyme (Birchfield and McIntosh 2020). Once the protein is purified, it is treated with thrombin, an enzyme that specifically cleaves at this recognition site, effectively separating the tag from the protein (LaVallie et al. 1994; Jenny et al. 2003). However, it is worth noting that crystallization is not always performed on the tag-cleaved form. In certain cases, enzymes have been successfully crystallized with their tags intact, without hindering the determination of the enzyme structure, as observed in the crystallization of a hexahistidine-tagged kinase (Derewenda

et al. 2007). In fact, a hexahistidine tag has been recently demonstrated to enhance the formation of crystals that diffract more effectively, a result attributed to the more compact and orderly packing within the crystalline structure (Moritzer et al. 2020).

Focusing on glycosyltransferases, studies have shown mixed results concerning the effects of tags like hexahistidine on enzyme functionality. Instances such as the DesVII glycosyltransferase have shown no alteration in catalytic activity (Borisova and Liu 2010). Others, such as the rhGalNAcT2 glycosyltransferase, might have experienced a decline in specific activity due to the inclusion of a hexahistidine tag (Lauber et al. 2015). Notably, the literature presents a limited number of studies that rigorously investigate the impact of tags on glycosyltransferases, indicating a gap in systematic analyses and comprehensive reviews in this domain. Consequently, a more exhaustive exploration and systematic documentation of tag effects on various glycosyltransferases, considering different substrates and reaction conditions, would be instrumental in establishing a more generalized understanding and developing guidelines for tag utilization in recombinant protein studies.

The deployment of c-myc and 6x His tags in enzyme studies encapsulates a rich and complex field. While offering invaluable benefits in purification and detection, their impact on functionality must be meticulously evaluated in each unique scenario. The study of their use, benefits, drawbacks, and application in various experimental setups enriches the understanding of their role in modern biotechnology. Their continued use and exploration may pave the way for innovative applications and new discoveries in the ever-evolving field of enzyme research.

Endogenous Proteases in the *Pichia pastoris* Expression System: Challenges and Considerations

The *Pichia pastoris* expression system has been a preferred choice to produce recombinant proteins owing to its high yield and ease of scalability. However, one of the persistent challenges in this system is the presence of endogenous proteases that can adversely affect the integrity and functionality of expressed proteins (Puxbaum et al. 2015). These endogenous proteases include serine, aspartic, and metalloproteases, which differ in their substrate specificity and cleavage sites. For example, aspartic proteases are known to cleave peptide bonds involving aspartic acid residues, while serine proteases may preferentially target arginine or lysine residues (Sorensen et al. 1994; Maeda et al. 2004; Robert et al. 2009).

Proteolytic activity can result in partial or complete degradation of the expressed protein. This degradation can alter the protein's structure, stability, activity, and its functionality in enzymatic studies (Jahic et al. 2003). The variations in molecular weight, charge, and other physicochemical properties due to proteolytic cleavage can affect purification and characterization processes, leading to inconsistent results (Zhang et al. 2007; Karbalaei et al. 2020).

To mitigate the impact of these proteases, several strategies have been adopted. One common approach is the selection of specific strains with reduced protease activity or the genetic modification of strains to knock out protease genes, such as the PEP4 (encoding proteinase A), PRB1 (proteinase B) and seven YPSs genes (yapsin family members) (Wu et al. 2013). Additionally, the optimization of culture conditions, such as temperature and pH, can suppress the expression or activity of these proteases (Zhang et al. 2007). Inclusion of protease inhibitors such as PMSF in the culture medium or during purification has also proven effective in some cases (Sakai et al. 1998; Yang et al. 2009).

Cp3GT Biochemical Characterization

Cp3GT was comprehensively tested with various flavonoid subclasses, such as chalcones, flavanones, flavones, dihydroflavonols, isoflavones, and anthocyanidins. It was found to specifically glucosylate flavonols at the 3-OH position, favoring quercetin as the main substrate (Owens and McIntosh 2009). However, it also exhibited 57.9%, 40.74%, and 22% relative activity with kaempferol, myricetin, and fisetin, respectively (Owens and McIntosh 2009; Devaiah et al. 2016). Following the initial characterization of Cp3GT, over 30 site-directed mutants were generated, each undergoing biochemical characterization and *in silico* modeling. Some of these mutants displayed variations in activities and specificities, warranting further investigations. A concise summary of the most intriguing and pertinent mutant characteristics is provided below.

Cp3GT Mutational Analysis – Highlights from Recent Work

Mutational analysis stands as a powerful tool in the exploration of enzymatic function and the specific roles played by individual amino acid residues (Qu et al. 2020). The contributions of those residues to substrate specificity, binding affinity, catalytic efficiency, and overall enzymatic function can be evaluated by introducing targeted mutations into key residues of an enzyme. The study of specific Cp3GT mutants in the McIntosh Lab detailed in the following sections illustrates the value of this approach, demonstrating how subtle alterations in amino acid sequences can lead to profound changes in enzymatic behavior.

The S20G-T21S Cp3GT mutant, constructed in 2014, incorporated two specific mutations: serine at residue 20 was replaced with glycine, and threonine at residue 21 with serine

(Sathanantham 2015). This modified enzyme demonstrated broader substrate specificity, glucosylating the flavanone naringenin, and favoring the flavonol kaempferol over quercetin (Sathanantham 2015). A 16% increase in overall activity was also observed compared to the wild type (Sathanantham 2015). Docking analyses indicated that the catalytic residue HIS-22 shifted closer to the glucosylation sites in both kaempferol and quercetin, but significantly more so in the former, providing an explanation for the mutant's preference for kaempferol (Sathanantham 2015). It was hypothesized that this that this positional shift also allowed HIS-22 to come sufficiently close to the 7-OH site in naringenin, enabling its glucosylation.

A different mutation, P297F, involving the substitution of proline at residue 297 with phenylalanine, resulted in a complete loss of activity with quercetin without any broadening of substrate specificity (Adepoju 2014). Computational modeling suggested that this mutation altered a loop near the catalytic cleft, blocking the binding pocket and eliminating enzyme activity (Adepoju 2014). This mutation, selected to potentially broaden regiospecificity in Cp3GT, was based on a comparison of amino acid sequences between different flavonoid GTs (Adepoju 2014). In the same study, an additional mutant, S20L, was biochemically evaluated. The mutation of SER-20 to leucine led to a complete loss of activity (Adepoju 2014). Given the proven significance of SER-18 in VvGT1 through X-ray crystallography (Offen et al. 2006) and the biochemical findings associated with the S20L mutant, it's evident that SER-20 in Cp3GT has a crucial role in glucosylation.

Another noteworthy mutation, P145T, emerged from modeling Cp3GT with VvGT1 (Kandel 2016). In this instance, a proline residue in Cp3GT was replaced with threonine, a common constituent of functional centers in proteins, which is thought to confer less conformational rigidity. This mutant, Cp3GT-P145T, exhibited a broadened substrate specificity

for naringenin and showed increased activity with kaempferol relative to quercetin (Kandel 2016), in a manner similar to S20G-T21S. Intriguingly, *in silico* models anticipated that the HIS-22 catalytic residue would be distant from the 7-OH site, preventing naringenin glucosylation. However, biochemical analyses contradicted this prediction, emphasizing the vital role of empirical validation for computational models.

A recent study performed a mutational analysis on Cp3GT, characterizing eight mutants, including L41M, N242K, E269K, 2AA, 5AA, and N242E+E269K, all of which were active, as well as 2AA+N242K+E269K and 5AA+N242K+E269K, which were inactive (Devaiah et al. 2018). Through homology modeling, the study reveals structural similarities and variations between Cp3GT and a related enzyme from *Citrus sinensis*, with specific attention to unstructured loop regions. Mutational analysis identified key amino acid residues (including Phe17, Ser20, His22, Gln87, Glu122, Phe124, Ser150, His154, Pro191, Phe203, Leu207, Phe385, Gly386, Asp387, and Glu388) that are flexible and play a role in ligand binding (Devaiah et al. 2018). Molecular docking techniques are employed to analyze the binding mode and interaction of ligands with both wild type and mutant Cp3GT, providing valuable insights into the enzyme's substrate and regional specificity (Devaiah et al. 2018).

Mechanistic Insights from Crystallized Plant Glycosyltransferases

The comparative analysis between Cp3GT and seven crystallized plant GTs is presented in Table 1.1. The structural analysis of plant GT crystal structures has provided significant insights into their substrate specificity and key catalytic mechanisms. From *Medicago truncatula's* UGT71G1, the vital roles of His-22, Asp-121, and Glu-381 residues were established (Shao et al. 2005). These residues not only played pivotal roles in the glycosylation process but were also vital in acceptor substrate recognition and catalysis (Shao et al. 2005). *Vitis*

vinifera's VvGT1 reiterated the crucial function of the PSPG box and identified key residues like Asp374, Gln375, Thr141, and His-20 in substrate interaction (Offen et al. 2006). *Arabidopsis thaliana's* UGT72B1 was unique in its glycosylation mechanism, with the catalytic histidine residue His19 displaying an alternate interaction than typically observed (Brazier-Hicks et al. 2007). *Medicago truncatula's* UGT85H2 unveiled a notable Trp360 residue behavior, indicating possible conformational changes post sugar-donor acceptance (Li et al. 2007). The subsequent UGT78K6 crystallization revealed critical residues in the donor binding site included Asp367 and Gln368, interacting with the glucose moiety on UDP-glucose, along with conserved catalytic residues His17 and Asp114, and in the acceptor-binding site, Pro78, Asp181, and Asp367 interacted with delphinidin(Hiromoto et al. 2013; Hiromoto et al. 2015). Lastly, the 2023 crystallization of a MiCGT-VFAH mutant from *Citrus micrantha* revealed a pocket formed by the substrate-binding residues Phe138, Val139, Ala140, and His142, which enhanced interaction with flavanol substrates, thus driving its specificity (Wen et al. 2021). Collectively, these studies underscore the importance of structural analysis in understanding the determinants of enzyme specificity and potential modification avenues for desired substrate interactions.

Table 1.1 Seven Crystallized Plant Glucosyltransferases. This table lists seven plant glucosyltransferases for which crystal structures have been determined, detailing their respective substrates, structural similarity, and identity percentages.

Title – Author - Year	**Name (PDB ID)**	**Substrate**	**Similarity**	**Identity**
Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago truncatula* (Shao et al. 2005)	UGT71G1 (2ACV)	Quercetin Medicagenic Acid Hederagenin	65.6%	32.5%
Structure of a Flavonoid Glucosyltransferase Reveals the Basis	VvGT1 (2C1Z)	Cyanidin Kaempferol Quercetin	82.3%	56.7%

for Plant Natural Product Modification (Offen et al. 2006)				
Crystal Structure of *Medicago truncatula* UGT85H2 - Insights into the Structural Basis of a Multifunctional (Iso)flavonoid Glycosyltransferase (Li et al. 2007)	UGT85H2 (2PQ6)	Kaempferol Quercetin Genistein Biochanin A Isoliquiritigenin	61.7%	25.5%
Characterization and Engineering of the Bifunctional N- and O-glucosyltransferase Involved in Xenobiotic Metabolism in Plants (Brazier-Hicks et al. 2007)	UGT72B1 (2VCE)	Quercetin Kaempferol Myricetin 2,4,5-Trichlorophenol (TCP) 3,4-dichloroaniline (DCA)	58.7%	28.3%
Crystal Structures of Glycosyltransferase UGT78G1 Reveal the Molecular Basis for Glycosylation and Deglycosylation of (Iso)flavonoids (Modolo et al. 2009)	UGT78G1 (3HBJ)	Quercetin Kaempferol Myricetin Cyanidin Pelargonidin Genistein Biochanin A	78.2%	47.5%
Crystal Structure of UDP-glucose: anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea* (Hiromoto et al. 2013) Structural basis for Acceptor-substrate Recognition of UDP-glucose: Anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea* (Hiromoto et al. 2015)	UGT78K6 (4REN)	Delphinidin Malvidin Peonidin Pelargonidin Flavonols (trace activity)	74.4%	43%
Directed Evolution of a Plant Glycosyltransferase for Chemo- And Regioselective Glycosylation of Pharmaceutically Significant Flavonoids (Wen et al. 2021)	MiCGT-VFAH (7VA8)	Quercetin Kaempferol	39.7%	24.%

Combined with detailed biochemical assays, the procurement of high-resolution structural data through techniques like X-ray crystallography and cryo-electron microscopy has

been indispensable in delineating the intricate links between protein structure and function (Nakai 1983; Pandey et al. 2020). However, obtaining such precise structural data experimentally poses challenges for certain proteins (Carpenter et al. 2008). To address this, computational algorithms were formulated to predict protein's 3-dimensional conformations. Introduced four decades ago, early algorithms leaned heavily on methods such as homology modeling and molecular dynamics simulations (Guzzo 1965; Prothero 1966; Schiffer and Edmundson 1967; Kotelchuck and Scheraga 1969; Lewis et al. 1970; Chou and Fasman 1974; Garnier et al. 1978). With the advent of technology, especially the rise of artificial intelligence, the field witnessed a transformative shift towards integrating deep learning and machine learning strategies (Shi et al. 2021; F. Li et al. 2022). These innovative algorithms present a marked improvement in the accuracy of protein structure predictions. Serving as a convenient foundational platform, these models are instrumental in hypothesizing protein structure-function relationships, an aspect crucial to Cp3GT analysis.

Combined with detailed biochemical assays, the procurement of high-resolution structural data through techniques like X-ray crystallography and cryo-electron microscopy has been indispensable in delineating the intricate links between protein structure and function (Nakai 1983; Pandey et al. 2020). However, obtaining such precise structural data experimentally poses challenges for certain proteins (Carpenter et al. 2008). To address this, computational algorithms were formulated to predict protein's 3-dimensional conformations. Introduced four decades ago, early algorithms leaned heavily on methods such as homology modeling and molecular dynamics simulations (Guzzo 1965; Prothero 1966; Schiffer and Edmundson 1967; Kotelchuck and Scheraga 1969; Lewis et al. 1970; Chou and Fasman 1974; Garnier et al. 1978). With the advent of technology, especially the rise of artificial intelligence,

the field witnessed a transformative shift towards integrating deep learning and machine learning strategies (Shi et al. 2021; F. Li et al. 2022). These innovative algorithms present a marked improvement in the accuracy of protein structure predictions. Serving as a convenient foundational platform, these models are instrumental in hypothesizing protein structure-function relationships, an aspect crucial to Cp3GT analysis.

Artificial Intelligence in Predicting Protein Structures

Artificial intelligence-driven protein structure prediction tools have become increasingly sophisticated, leveraging vast datasets and intricate neural network architectures. Tools such as AlphaFold, D-I-TASSER, and RoseTTAFold have emerged as frontrunners, utilizing deep learning to predict protein structures comparable to experimental methods (Baek et al. 2021; Jumper et al. 2021; Zheng et al. 2023). These tools primarily function by analyzing patterns in known protein structures and applying this knowledge to predict the 3D conformations of novel proteins. Their reliability has been showcased in recent Critical Assessment of Structure Prediction (CASP) competitions, which offers a platform where developers from around the world challenge their protein structure prediction algorithms on undisclosed protein structures. By comparing predictions to subsequently revealed experimental data, CASP provides an objective benchmark for the performance and progression of computational methods in the realm of structural biology (Kryshtafovych et al. 2021).

Advances in Protein Structure-Function Analysis Using AI-Driven *In Silico* Modeling

AlphaFold, developed by DeepMind, operates using a deep learning framework based on a deep residual neural network (Jumper et al. 2021; Hu et al. 2023). The system is specifically trained to predict inter-residue distances and dihedral (torsion) angles from amino acid

sequences, which are critical determinants of a protein's three-dimensional structure (Saravanan and Selvaraj 2017). To refine its predictions, AlphaFold leverages multiple sequence alignments, effectively comparing the target protein sequence with a vast database of known protein sequences to identify evolutionary patterns and constraints (Cuff and Barton 2000). This evolutionary information offers vital spatial insights that guide the protein's folding pathway (Chen and Kurgan 2007). Additionally, AlphaFold's model utilizes attention mechanisms, akin to those in language translation models, enabling it to capture long-range interactions between amino acids (Chen et al. 2021). By synthesizing these predicted distances, angles, and evolutionary cues, AlphaFold computes the most probable 3D configuration of a protein (Evans et al. 2021; Jumper et al. 2021).

DI-TASSER, an evolution of the I-TASSER methodology, distinguishes itself through a multifaceted approach. From the initial protein sequence, DI-TASSER generates intricate inter-residue contact maps, distance maps, and hydrogen-bond (HB) networks (Zheng et al. 2023). This is achieved using an array of deep neural-network predictors, such as the AttentionPotential — a self-attention network built upon the Multiple Sequence Alignment (MSA) transformer — and DeepPotential (Chen et al. 2021; Y. Li et al. 2022). Concurrently, structural templates are identified via the meta-threading LOMETS3 approach (Zheng et al. 2022). Notably, this includes models derived from the AlphaFold2 program. The culmination of this process sees full-length atomic models assembled through iterative fragment assembly Monte Carlo simulations. These simulations are guided by the I-TASSER force field and the deep-learning-generated contact, distance, and HB restraints (Yang and Zhang 2015). Furthermore, biological functions of the protein in question can be inferred from these structural models using the integrated

COFACTOR methodology, which predicts ligand binding sites and generates an easily accessible Program Database File (pdb) of receptor-ligand complexes (Zhang et al. 2017).

Hypothesis 1: Impact of C-Terminal c-myc/6x Histidine Tag on Cp3GT Activity

This research was designed to test the hypothesis that a C-terminal c-myc/6x histidine tag will not influence the activity of Cp3GT, including its kinetics, substrate preference, pH curve, and interaction with metals. In assessing this hypothesis, Cp3GT will be meticulously examined under identical conditions, with and without the c-myc/6x His tags. It was hypothesized that this tailored analysis will reveal no substantial differences in enzyme function, reflecting the assertion that these specific tags have no determinable effect on Cp3GT's intrinsic enzymatic properties.

Hypothesis 2: Scale-up, Expression, and Purification of Cp3GT for Structural Analysis

This research was designed to test the hypothesis that Cp3GT can be successfully expressed and purified to a degree of homogeneity suitable for structural analysis through the employment of the methylotrophic *Pichia pastoris* expression system, and through the utilization of standard chromatographic techniques, including immobilized metal affinity chromatography, anion exchange chromatography, and size exclusion chromatography.

Hypothesis 3: Proteolytic Cleavage of Cp3GT During Purification and Analysis

Given prior observations of protein bands on western blots that do not exhibit glucosyltransferase activity, this study hypothesizes that these bands are resultant forms of Cp3GT that have undergone proteolytic degradation by endogenous proteases of *P. pastoris*

during the purification process. This degradation is posited to alter the enzyme's properties, affecting its structural integrity, stability, and functional enzyme activity in subsequent experimental applications.

CHAPTER 2. THE EFFECT OF RECOMBINANT TAGS ON *CITRUS PARADISI* FLAVONOL-SPECIFIC 3-OGLUCOSYLTRANSFERASE ACTIVITY

Aaron S. Birchfield and Cecilia A. McIntosh*

Department of Biological Sciences, P.O. Box 70703, East Tennessee State University, Johnson City, TN 37614, USA; birchfieldas@etsu.edu

* Correspondence: mcintosc@etsu.edu; Tel.: +1-(423)-439-5838

Plants (Basel)

Year: 2020

Volume: 9

Issue: 3

Page: 402

doi: 10.3390/plants9030402

https://www.mdpi.com/2223-7747/9/3/402

Abstract

Recombinant tags are used extensively in protein expression systems to allow purification through IMAC (Immobilized Metal Affinity Chromatography), identification through Western blot, and to facilitate crystal formation for structural analysis. While widely used, their role in enzyme characterization has raised concerns with respect to potential impact on activity. In this study, a flavonol-specific 3-O glucosyltransferase (Cp3GT) from grapefruit (*Citrus paradisi*) was expressed in *Pichia pastoris*, and was assayed in its untagged form and with a C-terminal c-myc/6x His tag under various conditions to determine the effect of tags. Prior characterization of pH optima for Cp3GT obtained through expression in *Escherichia coli*, containing an N-terminal thioredoxin/6x His tag, indicated an optimal pH of 7–7.5, which is indicative of a normal physiological pH and agrees with other glucosyltransferase (GT) pH optima. However, characterization of Cp3GT expressed using *P. pastoris* with a C-terminal c-myc-6x His tag showed a higher optimal pH of 8.5–9. This suggests a possible tag effect or an effect related to physiological differences between the cell expression systems. Results testing recombinant Cp3GT expressed in Pichia with and without C-terminal tags showed a possible tag effect with regard to substrate preference and interactions with metals, but no apparent effect on enzymatic kinetics or pH optima.

Keywords: glucosyltransferase; recombinant tags; His-tag; flavonoids; flavonol; grapefruit; C-terminal recombinant tags; pH optima; reaction kinetics

2.1 Introduction

Plants produce secondary metabolites that impart a wide range of beneficial effects and increase overall survivability. Flavonoids belong to the phenolic class of secondary metabolites and comprise a diverse class of over 6000 compounds that assist with plant defense, taste, smell, and coloration of flowers [1]. They play a role in insect-pollinator interactions and have been shown to facilitate the symbiotic interaction of nitrogen-fixating bacteria with legumes [2–4]. Flavonoids have antioxidant capabilities and participate in countless physiological activities. They, in part, account for the nutritional potential of many fruits and vegetables and have been shown to reduce inflammation and protect the heart [5]. Numerous flavonoids have shown medicinal potential as drugs to fight cancer [6,7].

Unraveling flavonoid biosynthesis has led to the production of novel plant variants with unique appearance and taste [8]. Engineering unique metabolic pathways requires a finely tuned collection of cell systems, enzymes, and substrates that are reconstituted in vitro to produce a desired compound [9]. Engineered pathways can be made more productive by incorporating enzyme components optimized through mutational characterization studies [10]. Isolating an enzyme for study relies widely on cell expression lines that use vector encoded tags for identification and purification. Some tags are only a few amino acids in length and are placed either at the very beginning or end of an enzyme, yet it is possible they may affect the structure of the enzyme, causing a change in activity [11]. Some vectors come with a cleavage site that allows removing tags to obtain the native enzyme, however, many enzymes are characterized with tags intact with no account given as to whether they impact activity [12–14]. It is generally assumed that tags have no impact on enzyme activity, yet research has shown instances where activity is in fact affected [15–18]. Furthermore, many enzymes are subsequently crystallized for

structure determination with tags intact. These tags are not always reported in Protein Data Base (PDB) entries, yet there is evidence that inclusion of tags during crystallization can both aid or hinder crystallization depending on the enzyme, as well as influencing expression [19–22].

Citrus paradisi flavonol-specific 3-O glucosyltransferase (Cp3GT) (Genbank Protein ID: ACS15351) was previously characterized and shown to catalyze the transfer of glucose to the flavonols quercetin and kaempferol [22]. Cp3GT also showed activity with the flavonols myricetin, fisetin, and gossypetin to a lesser extent [23,24]. Cp3GT showed exclusive activity with flavonols and is regioselective for the 3-OH position. Cp3GT was initially cloned into the pCD1 vector, a modified PET32 vector that includes N terminal thioredoxin/6x His tags and a thrombin cleavage site [23]. Concerns regarding expression of Cp3GT in *E. coli* and inclusion body formation prompted cloning into the pPicZa vector and transforming into *P. pastoris* [23]. A thrombin recognition sequence was inserted into the recombinant pPicZa vector upstream of the C terminal c-myc/6x His tags. This was necessary for cleaving the tags so that untagged and tagged Cp3GT could be compared.

Initial characterizations of pH optima for Cp3GT, obtained through expression in *E. coli*, indicated an optimal pH of 7–7.5, which is indicative of a normal physiological pH. This agrees with other flavonoid glucosyltransferase (GT) pH optima of 6.5–8 [25–27]. More recent characterizations of Cp3GT expressed using *P. pastoris*, however, showed a higher optimal pH of 8.5–9. This suggests either an effect due to the location and chemistry of the recombinant tags, or an effect related to physiological differences between the cell expression systems [24]. This research was designed to test the hypothesis that recombinant tags have an effect on Cp3GT activity with respect to pH optima, substrate preference, kinetic parameters, and interaction with various metals.

2.2 Results and Discussion

2.2.1 Effect of Tags on Optimal pH

To make a comparison regarding pH optima between tagged and untagged enzyme, Cp3GT activity was measured over a pH range of 5.5–9.5 using various buffers that overlapped in pH. Among plant GT's, pH optima tend to align with physiological pH (7–7.5) but can vary as high as 9 and as low as 5.5 when expressed in recombinant systems [26,28–30]. A shift away from physiological pH could suggest that conditions surrounding the expression and preparation of the enzyme are influencing stability and activity, especially for enzymes such as plant GT's that have structural similarity. It is unclear what conditions define this influence, yet such a shift in pH optima was observed when Cp3GT was assayed using different cell expression systems and different recombinant tags. Recombinant Cp3GT expressed in *E. coli* and the enzyme expressed in the Pichia system showed different pH optima. One hypothesis is that this could be due to the different chemistry and position of the tags.

The characterization of Cp3GT showed an optimal pH of 7–7.5 for the protein expressed in *E. coli* with N terminal thioredoxin/6x His tags [23]. The tendency for bacterial cell-expression systems to generate mis-folded recombinant proteins as a result of conformational stress has been well studied [31]. Thioredoxin is a thermally stable 12 kDa protein often used in *E. coli* expression systems to aid in solubility of the recombinant protein and has furthermore been used as a stabilizing fusion partner for driving crystallization in some instances [32–34]. However, given its size, it is possible that it could impact conformational folding of Cp3GT. Analysis of *E. coli* expressed Cp3GT with and without the N-terminal thioredoxin.6x His tags showed no impact of these tags on enzyme activity [23].

A characterization of pH optima was conducted for Cp3GT expressed in *P. pastoris* that contained C terminal c-myc/6x His tags. The optimal pH shifted to 8.5–9 in both tagged and untagged samples (Fig. 2.1) which is higher than that observed for the enzyme expressed in *E. coli*. There also was an apparent CHES (N-cyclohexyl-2-aminoethane sulfonic acid) buffer effect showing an increase in activity at pH 9 when compared with bicine. This is consistent with pH optima from a recent study reporting on mutational analysis of Cp3GT [24].

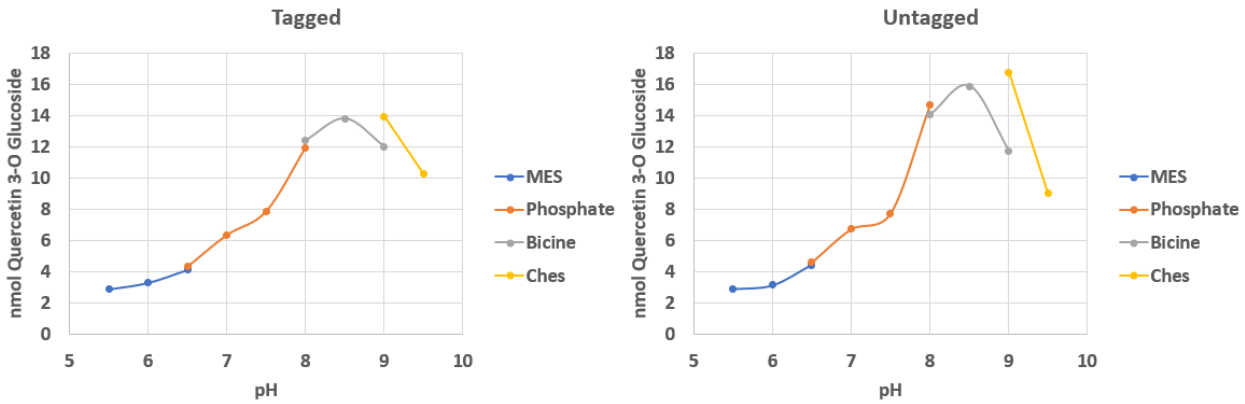

Fig. 2.1 Optimal pH for Activity in Tagged and Untagged Cp3GT from *P. pastoris*. Formation of quercetin 3-O glucoside was measured and plotted against the pH range tested. A mild buffer effect was seen with N-cyclohexyl-2-aminoethane sulfonic acid (CHES) pH 9. n = 2.

It was hypothesized that using different tags at the C terminus contributed to a higher pH optimum. However, cleaving the c-myc/6x His tag and testing untagged Cp3GT from *P. pastoris* over the same pH range showed no significant difference in activity levels between tagged and untagged Cp3GT (Fig. 2.1). Likewise, cleaving the thioredoxin tag and testing untagged Cp3GT from *E. coli* over a pH range showed no significant difference in activity levels between tagged and untagged Cp3GT [23].

These findings together indicate no tag effects exist with regards to pH optima when Cp3GT is expressed in *P. pastoris* and contains a C terminal c-myc/6x His tag. This is consistent

with data comparing tagged and untagged Cp3GT expressed in *E. coli*, wherein the presence of tags did not significantly alter pH stability [23]. It remains unclear why the optimum pH for Cp3GT activity shifted from 7–7.5 when expressed in *E. coli* to 8.5–9 when expressed in *Pichia pastoris*. While it may be something inherent to expression in prokaryotic versus eukaryotic systems, this remains to be determined. For example, it may be that Cp3GT may have been post-translationally modified in the yeast system, although this has not as yet been reported for plant GTs.

It should be noted that the overall yield of soluble Cp3GT enzyme was greater in the *Pichia* system. As previously stated, expression in the *E. coli* system resulted in the majority of Cp3GT being localized in inclusion bodies and the use of sonication to burst the cells to release soluble protein could have resulted in some denaturing of the protein. As a result, the amount of purified Cp3GT from *E. coli* needed to run kinetic reactions was over 50x greater than that used for the purified *Pichia* expressed protein (pH 7.5). The 0.5 µg Cp3GT from yeast used in our assays sustained linear velocity of activity for 10 min, compared with the 20–30 µg Cp3GT from *E. coli* that sustained linear velocity for 5 min. This suggests that a purer, cleaner enzyme is obtained when expressing Cp3GT in *P. pastoris*. The ability to obtain high concentrations of pure Cp3GT has positive implications for structure determination using X-ray crystallography.

2.2.2 Effect of Tags on Substrate Specificity of Cp3GT

Some studies have shown 6x-His tag interactions with catalytic structural features can impair substrate binding [18]. To investigate differences in substrate preference with regard to the presence or absence of recombinant tags, Cp3GT was assayed for activity with its preferred flavonol substrate quercetin, as well as two other flavonols (kaempferol and 4,-methoxy-

flavonol), and the flavanone naringenin (Fig. 2.2). This experiment tested the hypothesis that recombinant tags will have no impact on substrate preference.

In general, flavonol-specific Cp3GT shows the greatest activity with quercetin and kaempferol. It did not interact with flavonols containing methoxy groups, as seen by the lack of significant activity when assayed with 4,-methoxy-flavonol. Cp3GT was previously shown to be inactive when methoxy groups were placed on the flavonoid backbone [23,25,35]. There also was no significant activity with the flavanone naringenin which lacks a 3-OH group; this was also noted in previous studies [23,25,35]. The absence of tags appears to enhance Cp3GT activity with quercetin and kaempferol by approximately 14% and 20% respectively (Fig. 2.2).

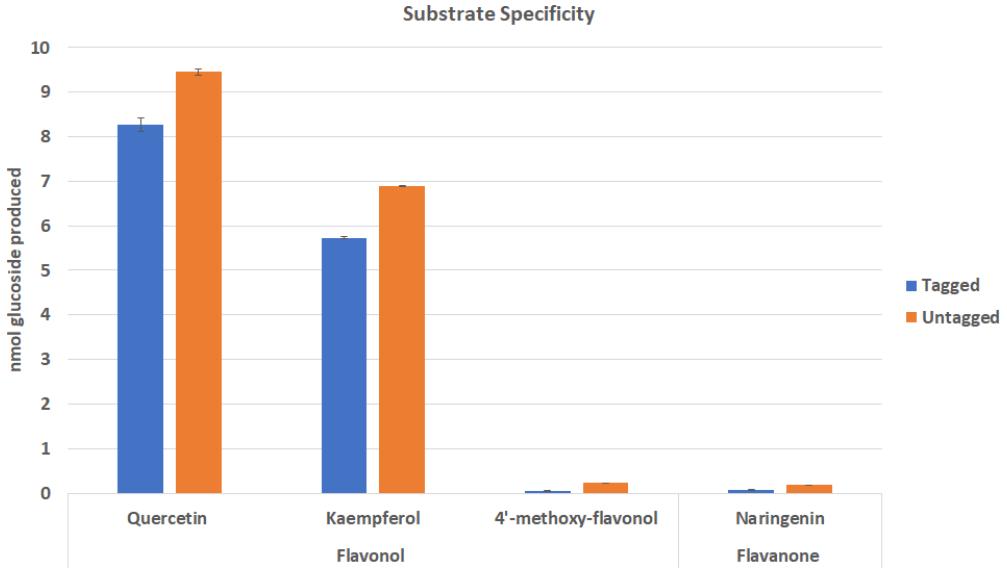

Fig. 2.2 Effect of Tags on Cp3GT Substrate Preference. Untagged Cp3GT showed greater activity with the preferred flavonol substrates quercetin and kaempferol. There was no significant activity with methoxylated flavonols or flavanones. $n = 2$.

Other studies have shown that recombinant tags can impact activity in various ways [14–17]. One such interaction occurred between the catalytic binding site of a bacterial dehydrogenase and its recombinantly expressed C-terminal 6x His-tag that resulted in loss of

activity due to alteration of the binding site [18]. In fact, the recombinant enzyme with an N-terminal tag showed 20-fold higher activity [18]. *In silico* analysis previously carried out on Cp3GT detected no interactions of this kind, however these models should be verified using X-ray crystallography [24]. With regard to substrate specificity of Cp3GT, it is possible that tags impact folding and conformation to some degree through structural interactions, and these interactions may alter substrate binding. The full scope of Cp3GT- recombinant tag interactions may be identified in future analyses using X-ray crystallography.

2.2.3 Effect of Tags on Kinetic Parameters

Kinetics assays were conducted to determine any tag effects related to K_m^{app}, V_{max}, catalytic efficiency, and specificity constant. Kinetic assays were carried out under standard conditions of testing various concentrations of one substrate while having unlimiting concentrations of the other. Cp3GT exhibits Michaelis–Menten kinetics, thus the Lineweaver–Burke linear transformation was plotted to determine K_m^{app} and V_{max} (Table 2.1). This was done for Cp3GT containing the C-terminal tags and for Cp3GT with tags removed. No statistically significant differences were detected between tagged and untagged Cp3GT (mean ± SD). The K_m^{app} for quercetin and kaempferol were not significantly different, though previous findings indicated a slightly higher K_m^{app} for quercetin [23,24]. Furthermore, catalytic efficiency for quercetin was higher than kaempferol, which suggested more overall product could be produced when quercetin is used as acceptor substrate. These K_m^{app} are similar to those found for the protein expressed from *E. coli* [22]. This would suggest that whatever is impacting the pH optima of the protein expressed in the two systems is not impacting the actual flavonol substrate kinetics.

Table 2.1**:** Kinetics on tagged and untagged Cp3GT with three substrates. Values are expressed as mean ± SD. *Average of 2 experiments (*n* = 2) **Average of 6 experiments (*n* = 2). UDPG = Uridine Diphosphoglucose.

Substrate		**K_m^{app} (μ)**	**Vmax (pKat/μg)**	**Kcat (s^{-1})**	**Kcat/K_m^{app} (μM/s)**
Quercetin*	Tagged	51.9 ± 8.29	31.7 ± 3.39	1.84 ± 0.19	0.035 ± 0.001
	Untagged	42.4 ± 3.48	26.5 ± 1.06	1.54 ± 0.06	0.036 ± 0.001
Kaempferol*	Tagged	44.1 ± 3.41	22.0 ± 0.56	1.28 ± 0.02	0.029 ± 0.001
	Untagged	53.7 ± 10.6	23.1 ± 3.00	1.34 ± 0.17	0.025 ± 0.001
UDPG**	Tagged	49.7 ± 5.32	10.9 ± 1.59	0.637 ± 0.092	0.012 ± 0.002
	Untagged	55.0 ± 4.97	12.7 ± 2.15	0.738 ± 0.125	0.012 ± 0.0004

The K_m^{app} observed for UDP-glucose (Table 2.1) was approximately 10-fold lower than previously observed in Cp3GT expressed from both *E. coli* (669 μM) and *P. pastoris* (878 μM) [23,24]. Potential issues with amounts of active enzyme from the *E. coli* system were previously discussed. With respect to the current work, the Cp3GT protein was significantly purer than that in the previously published study on yeast expressed protein [24]. It is possible that while the protein was less active, it was still able to bind UDP-glucose thus making the K_m^{app} appear significantly larger. For some enzymes that utilize UDP-glucose, K_m^{app} values can be higher for the UDP donor substrate than for their corresponding acceptor substrate, yet still these values can vary considerably [36–41]. V_{max} was noticeably lower for UDP-glucose than for acceptor substrates quercetin and kaempferol and is lower than the previously reported values of 28.17 (Cp3GT from *E. coli*) and 83.5 (Cp3GT from *P. pastoris*) [23,24]. This could be due, in part, to the increased purity and activity of Cp3GT in this current study.

UDP has been shown previously to be a competitive inhibitor of Cp3GT as well as other flavonoid GT enzymes [23,35,42,43]. Inhibition was tested with tagged and untagged protein from the yeast expression system. Results again confirmed competitive inhibition with a Ki of

61.8 μM for tagged Cp3GT and a Ki of 73.1 μM for untagged Cp3GT. These results are comparable with the UDP Ki of 69.5 μM obtained during the first characterization of Cp3GT expressed in *E. coli* indicating that the presence of tags on the protein does not change the nature of competitive inhibition by UDP [23]. This suggests that the presence of N-terminal or C-terminal tags on Cp3GT does not impact competitive binding of UDP and the Kis are similar. This, coupled with similar K_m^{app} for quercetin and kaempferol for tagged and untagged Cp3GT in this study, suggest that these tags do not significantly impact folding and orientation of the binding cleft.

2.2.4 Effect of Tags and Metals on Cp3GT Activity

Previous studies have shown that some metals can impact Cp3GT activity [23,42,44]. Therefore, nine metals were included in separate assays at concentrations of 1 and 10 mM (Table 2.2). Significant inhibition was observed for both tagged and untagged Cp3GT in reactions containing zinc, iron, calcium, copper, and manganese relative to control reactions without these compounds. A significantly greater inhibitory effect was observed for untagged samples using zinc and manganese for both 1 and 10 mM concentrations. Cp3GT activity was inhibited in samples containing both 1 and 10 mM iron with no significant difference between tagged and untagged samples. In the presence of 1 and 10 mM copper, both tagged and untagged Cp3GT were inhibited. Interestingly, magnesium significantly enhanced activity over control in the 10 mM sample, and the enhancement was greater in samples using untagged Cp3GT. KCl and NaCl also slightly enhanced Cp3GT activity.

Table 2.2 Activity of Cp3GT in the presence of various metals. Of the metals tested, $ZnCl_2$, NaCl, $MnCl_2$, $MgCl_2$ were significantly different between tagged and untagged Cp3GT. Values are expressed in % relative activity (n = 2).

	% Relative Activity			
Metal	**Tagged**		**Untagged**	
	1 mM	10 mM	1 mM	10 mM
Untreated	100	100	100	100
$ZnCl_2$	44	28	25	23
KCl	117	119	114	112
$FeSO_4$	42	16	42	5
NaCl	118	121	100	95
Na_2SO_4	92	88	92	87
$CaCl_2$	89	69	84	72
$CuSO_4$	79	3	70	3
$MnCl_2$	87	87	60	56
$MgCl_2$	104	123	103	140

Some differences were observed between tagged and untagged Cp3GT with regard to metal interactions. Specifically, untagged Cp3GT activity was enhanced in the presence of magnesium to a greater extent than tagged Cp3GT, however, in the presence of NaCl, tagged Cp3GT was enhanced to a greater extent over untagged Cp3GT. It remains unclear what specific interactions are driving this influence. It is possible that the stabilizing magnesium-UDP-G complex is more readily formed when magnesium is added to the solution. The need for magnesium in GT-catalyzed uridine diphosphate sugar transfer is well documented in biochemical systems, shown here by an increase in activity for samples containing added magnesium [45].

Consistent with previous findings, Cp3GT from *E. coli* and *P. pastoris* showed almost complete loss of activity in the presence of Fe^{2+}, Cu^{2+}, and Zn^{2+} ions [23,24]. Cp3GT from *E. coli* also showed a mild stimulatory effect in the presence of Mg^{2+} [23]. With respect to metal

interactions, no differences were previously detected between tagged and untagged Cp3GT when expressed in *E. coli*.

2.3 Conclusions and Directions for Future Research

It was hypothesized that for pH optima, substrate binding, and interactions with inhibitor/activators, Cp3GT activity was impacted by the presence of a C-terminal c-myc/6x tag. Results do not support this hypothesis as this interaction does not appear to impact overall Cp3GT kinetics nor optimal pH. It is clear that recombinant tags do not explain the pH optima difference. Despite having no impact on kinetics or optimal pH, the differences observed in substrate binding and metal interaction between tagged and untagged Cp3GT suggest that recombinant tags effect substrate preference and metal binding. This could be resolved through structure determination by X-ray crystallography. Furthermore, it is possible that either donor or acceptor substrate more readily binds to untagged Cp3GT, thus aiding efforts to co-crystallize Cp3GT with substrate. These findings encourage the use of untagged Cp3GT when possible due to an apparent increase in substrate binding and activity over Cp3GT with tags intact.

While a few plant secondary product glucosyltransferases have had crystal structures solved [46–51], there is currently no crystal structure for a plant UDP glucosyltransferase that has the novel substrate and regiospecificity exhibited by Cp3GT. Additionally, there are no reports that compare crystal structures of plant UDP glucosyltransferases with and without tags. Therefore, future work to perform high-resolution structural analysis on Cp3GT will make a significant contribution to the field due to both its novel substrate and regiospecificity. It is also possible that this analysis will elucidate recombinant tag interactions that influence activity.

It is logical to propose that addition of tags to recombinant proteins could impede upon the determination of its true biological nature, yet unless gross deficiencies are observed

experimentally, this assertion is often overlooked. Recombinant tag effects may only need to be investigated on a case by case basis, but it was indeed observed that Cp3GT binds flavonol substrates and interacts with some metals differently when a recombinant tag was affixed at the C terminus. Therefore, future work will focus on a high-resolution structural analysis of tagged Cp3GT that will elucidate both the structural basis for its novel specificity and determine how recombinant tags impact Cp3GT structure. Analysis of Cp3GT co-crystallized with acceptor and donor substrates would further elucidate both the nature of substrate binding itself, and whether recombinant tags effect substrate binding in some way. Structural analysis of Cp3GT will also include verifying recent models that predict residues responsible for Cp3GT substrate and regiospecificity. The effects of implementing additional purification steps to remove non-native artifacts is generally well tolerated in crystallizing GTs [52].

The ubiquity of recombinant tags in protein expression systems cannot be ignored and their use may almost be completely necessary. Indeed, their prevalence does not preclude thoroughly investigating the scope of recombinant tag effects. It is highly possible that determining the nature of these effects can greatly benefit the field.

2.4 Methods and Materials

2.4.1 Reagents and Materials

Reagents used were analytical grade and were obtained from the following sources: quercetin, $FeSO_4$, $CuSO_4$, and $ZnCl_2$ were from MilliporeSigma (St. Louis, MO, USA); kaempferol was from Indofine (Hillsborough, NJ, USA), UDP-glucose was from Calbiochem (Gibbstown, NJ, USA), UDP- [U-14C] glucose was from PerkinElmer (Boston, MA, USA); Midiprep plasmid extraction kit was obtained from Promega (Madison, WI, USA). Miniprep plasmid extraction kit, $MgCl_2$, $MnCl_2$, KCl, and NaCl were purchased from Fisher Scientific

(Waltham, MA, USA); $CaCl_2$ was from Acros Organics (Morris Plains, Morris, NJ, USA); Na_2SO_4 was from Merk (Kenilworth, NJ, USA). Bovine thrombin was obtained from MP Biomedicals (Solon, OH, USA).

2.4.2 Insertion of Thrombin Cleavage Site and Transformation into Yeast

A thrombin cleavage site was cloned into the pPicZa plasmid containing the Cp3GT sequence (Genbank Protein ID: ACS15351) using Agilent's QuikChange XL Site-Directed Mutagenesis Kit. Recombinant plasmid was transformed into competent *E. coli* cells and plasmid was sequenced to verify site insertion. DNA was transformed into X-33 *Pichia pastoris* of the Mut+ phenotype and expressed under methanol induction according to Invitrogen's EasySelect Pichia Expression Kit and as previously described [24,35].

2.4.3 Removal of Recombinant Tags by Thrombin Digestion

Bovine thrombin was resuspended in 0.7% NaCl to a concentration of 20 U/µL. Approximately 2 units per microgram of Cp3GT was needed to achieve complete removal of tags at 4 ◦C for 2 h as verified by SDS-Page and Western Blot (Fig.s 2.3–2.5).

2.4.4 Purification and GT Enzyme Assay

Extraction and purification of recombinant Cp3GT from yeast was carried out as previously described, using His-Pur immobilized cobalt metal affinity resin in gravity flow columns[23,35]. The bound resin was washed with 50 mL of sodium phosphate buffer containing 300 mM NaCl and 10 mM imidazole to remove non-specific, weakly bound proteins. Cp3GT was eluted using sodium phosphate buffer containing 150 mM imidazole. Eluent fractions were pooled, dialyzed, and concentrated into 50 mM NaPO4 buffer, pH 7.5 containing 14 mM BME using Amicon-15 (30 kDa) centrifugal filters. Purification was confirmed through SDS-Page and Western Blot. Enzyme activity assays were conducted as previously described using 0.5 µg of

purified protein at 30 ◦C, pH 7.5 [42]. Initial time course assays indicated reaction was linear for at least 10 min, thus all reactions were run for 10 min.

2.4.4.1 pH Optima

For enzymatic assays testing optimal pH, reactions were buffered with either 2-ethansulfonic acid (MES, pH 5.5, 6, and 6.5), potassium phosphate (pH 6.5, 7, 7.5, and 8), bicine (pH 8, 8.5, and 9), or CHES (pH 9 and 9.5). Reactions were carried out as previously described using 0.5 µg of purified Cp3GT at 30 ◦C, pH 7.5.

2.4.4.2 Substrate Specificity

Enzymatic assays testing substrate specificity were carried out as previously described using 0.5 µg of purified Cp3GT at 30 ◦C, pH 7.5. Each reaction contained 50 nanomoles of either quercetin, kaempferol, 4,-methoxy-flavonol, or naringenin.

2.4.4.3 Kinetics

Assays measuring Cp3GT kinetics were carried out using increasing concentrations of quercetin and kaempferol ranging from 0–667 µM. Kinetic assays using UDP-G were carried out using increasing concentrations ranging from 0–1333 µM. All reactions were carried out as previously described using 0.5 µg of purified Cp3GT at 30 ◦C, pH 7.5.

2.4.4.4 Metals and UDP Inhibition

Enzymatic assays testing metal interactions were carried out as previously described. Each reaction contained one of nine metals at a final concentration of either 1 or 10 mM. For assays measuring UDP inhibition, reactions contained either 0, 0.05, or 0.1 mM UDP.

2.4.5 Statistical Analysis

Calculations were performed to determine mean ± SD for replicate experiments. Sample results with overlapping SD bars were considered not significantly different.

2.5 Supplemental Materials

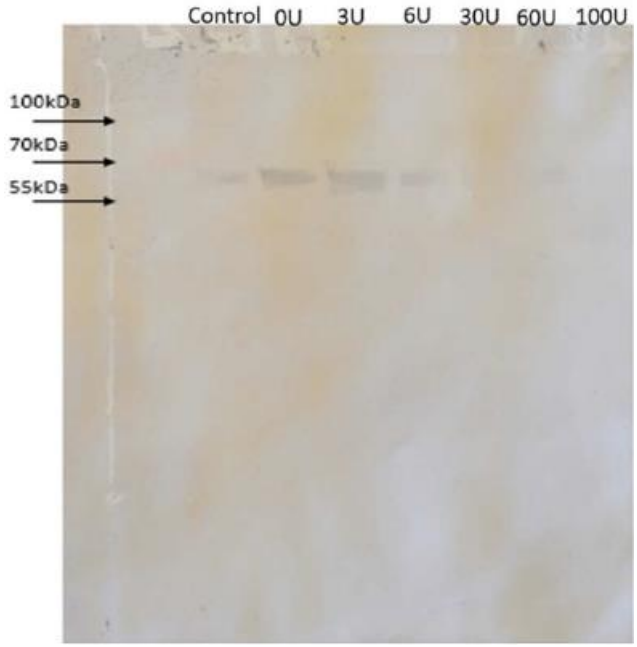

Fig. 2.3 Cp3GT Digest with Increasing Amounts of Thrombin**.** Complete removal of tags indicated by absence of bands. The lack of a band is due to the removal of the c-myc peptide that corresponds to our antibody.

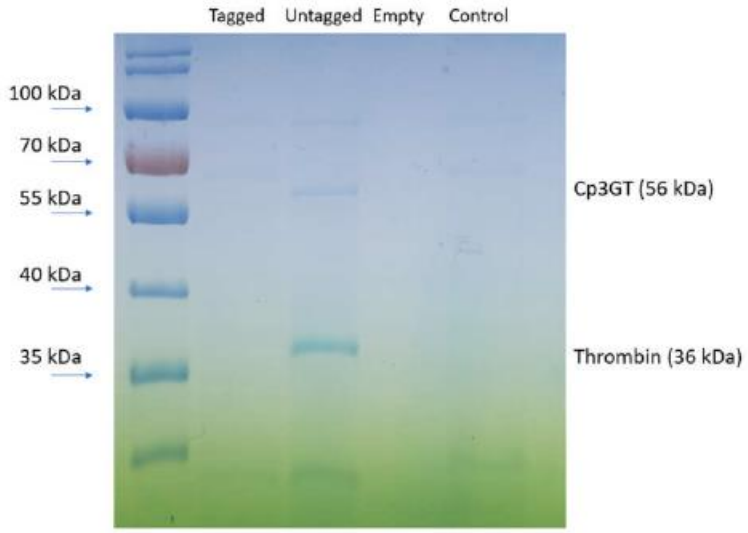

Fig. 2.4 Coomassie Stain of Cp3GT Before and After Treating with Thrombin**.** Cp3GT can be seen at 56 kDa in its native form, whereas with tags it runs slightly higher due to the increase in size from the recombinant tags. Thrombin is shown at 36kDa in untagged samples.

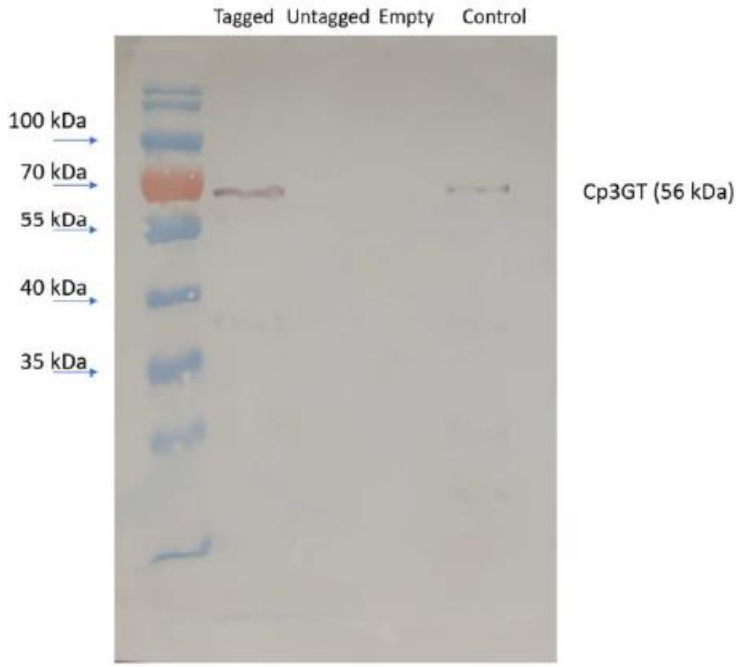

Fig. 2.5 Western Blot of Cp3GT Before and After Treating with Thrombin. Complete removal of tags indicated by the absence of a band in the untagged sample.

2.6 Author Contributions

Both authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

2.7 Funding

This research was funded by an East Tennessee State University Research Development Committee Major Grant awarded to CAM.

2.8 Conflicts of Interest

The authors declare no conflict of interest.

References

1. Mathesius, U. Flavonoid Functions in Plants and Their Interactions with Other Organisms. *Plants (Basel)* **2018**, *7*, doi:10.3390/plants7020030.

2. Agati, G.; Tattini, M. Multiple Functional Roles of Flavonoids in Photoprotection. *New Phytologist* **2010**, *186*, 786–793, doi:10.1111/J.1469-8137.2010.03269.X.

3. Liu, C.W.; Murray, J.D. The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. *Plants 2016, Vol. 5, Page 33* **2016**, *5*, 33, doi:10.3390/PLANTS5030033.

4. Simmonds, M.S.J. Importance of Flavonoids in Insect–Plant Interactions: Feeding and Oviposition. *Phytochemistry* **2001**, *56*, 245–252, doi:10.1016/S0031-9422(00)00453-2.

5. Kozłowska, A.; Szostak-Wegierek, D. Flavonoids--Food Sources and Health Benefits. *Rocz Panstw Zakl Hig* **2014**, *65*, 79–85.

6. De Masi, L.; Bontempo, P.; Rigano, D.; Stiuso, P.; Carafa, V.; Nebbioso, A.; Piacente, S.; Montoro, P.; Aversano, R.; D'Amelia, V.; et al. Comparative Phytochemical Characterization, Genetic Profile, and Antiproliferative Activity of Polyphenol-Rich Extracts from Pigmented Tubers of Different *Solanum Tuberosum* Varieties. *Molecules* **2020**, *25*, doi:10.3390/molecules25010233.

7. Dai, J.; Mumper, R.J. Plant Phenolics: Extraction, Analysis and Their Antioxidant and Anticancer Properties. *Molecules* **2010**, *15*, 7313–7352, doi:10.3390/molecules15107313.

8. Falcone Ferreyra, M.L.; Rius, S.P.; Casati, P. Flavonoids: Biosynthesis, Biological Functions, and Biotechnological Applications. *Front Plant Sci* **2012**, *3*, 222, doi:10.3389/fpls.2012.00222.

9. Zhou, K.; Qiao, K.; Edgar, S.; Stephanopoulos, G. Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products. *Nat Biotechnol* **2015**, *33*, 377–383, doi:10.1038/nbt.3095.

10. Chen, W.; Yao, J.; Meng, J.; Han, W.; Tao, Y.; Chen, Y.; Guo, Y.; Shi, G.; He, Y.; Jin, J.-M.; et al. Promiscuous Enzymatic Activity-Aided Multiple-Pathway Network Design for Metabolic Flux Rearrangement in Hydroxytyrosol Biosynthesis. *Nat Commun* **2019**, *10*, 960, doi:10.1038/s41467-019-08781-2.

11. Terpe, K. Overview of Tag Protein Fusions: From Molecular and Biochemical Fundamentals to Commercial Systems. *Appl Microbiol Biotechnol* **2003**, *60*, 523–533, doi:10.1007/s00253-002-1158-6.

12. Fitzgerald, K.; Holliger, P.; Winter, G. Improved Tumor Targeting by Disulphide Stabilized Diabodies Expressed in *Pichia Pastoris*. *Protein Eng* **1997**, *10*, 1221–1225, doi:10.1093/protein/10.10.1221.

13. Furqan, B.R.N.; Akhmaloka Heterologous Expression and Characterization of Thermostable Lipase (Lk1) in *Pichia Pastoris* GS115. *Biocatal Agric Biotechnol* **2020**, *23*, 101448, doi:10.1016/J.BCAB.2019.101448.

14. Hong, R.; Sun, Y.; Su, L.; Gu, L.; Wang, F.; Wu, J. High-Level Expression of *Humicola Insolens* Cutinase in *Pichia Pastoris* without Carbon Starvation and Its Use in Cotton Fabric Bioscouring. *J Biotechnol* **2019**, *304*, 10–15, doi:10.1016/J.JBIOTEC.2019.07.011.

15. Chen, Z.; Li, Y.; Yuan, Q. Study the Effect of His-Tag on Chondroitinase ABC I Based on Characterization of Enzyme. *Int J Biol Macromol* **2015**, *78*, 96–101, doi:10.1016/J.IJBIOMAC.2015.03.068.

16. Dickson, J.M.J.; Lee, W.J.; Shepherd, P.R.; Buchanan, C.M. Enzyme Activity Effects of N-Terminal His-Tag Attached to Catalytic Sub-Unit of Phosphoinositide-3-Kinase. *Biosci Rep* **2013**, *33*, 857–863, doi:10.1042/BSR20130075.

17. Sabaty, M.; Grosse, S.; Adryanczyk, G.; Boiry, S.; Biaso, F.; Arnoux, P.; Pignol, D. Detrimental Effect of the 6 His C-Terminal Tag on YedY Enzymatic Activity and Influence of the TAT Signal Sequence on YedY Synthesis. *BMC Biochem* **2013**, *14*, doi:10.1186/1471-2091-14-28.

18. Yeon, Y.J.; Park, H.J.; Park, H.Y.; Yoo, Y.J. Effect of His-Tag Location on the Catalytic Activity of 3-Hydroxybutyrate Dehydrogenase. *Biotechnology and Bioprocess Engineering* **2014**, *19*, 798–802, doi:https://doi.org/10.1007/s12257-014-0089-2.

19. Smyth, D.R.; Mrozkiewicz, M.K.; McGrath, W.J.; Listwan, P.; Kobe, B. Crystal Structures of Fusion Proteins with Large-Affinity Tags. *Protein Sci* **2003**, *12*, 1313–1322, doi:10.1110/PS.0243403.

20. Noirclerc-Savoye, M.; Flayhan, A.; Pereira, C.; Gallet, B.; Gans, P.; Ebel, C.; Breyton, C. Tail Proteins of Phage T5: Investigation of the Effect of the His6-Tag Position, from Expression to Crystallization. *Protein Expr Purif* **2015**, *109*, 70–78, doi:10.1016/J.PEP.2015.02.003.

21. Woestenenk, E.A.; Hammarström, M.; Van Den Berg, S.; Härd, T.; Berglund, H. His Tag Effect on Solubility of Human Proteins Produced in *Escherichia Coli*: A Comparison between Four Expression Vectors. *J Struct Funct Genomics* **2004**, *5*, 217–229, doi:10.1023/B:JSFG.0000031965.37625.0E.

22. Carson, M.; Johnson, D.H.; McDonald, H.; Brouillette, C.; DeLucas, L.J. His-Tag Impact on Structure. *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 295–301, doi:10.1107/S0907444906052024.

23. Owens, D.K.; McIntosh, C.A. Identification, Recombinant Expression, and Biochemical Characterization of a Flavonol 3-O-Glucosyltransferase Clone from *Citrus Paradisi*. *Phytochemistry* **2009**, *70*, 1382–1391, doi:10.1016/j.phytochem.2009.07.027.

24. Devaiah, S.P.; Tolliver, B.M.; Zhang, C.; Owens, D.K.; McIntosh, C.A. Mutational Analysis of Substrate Specificity in a *Citrus Paradisi* Flavonol 3-O-Glucosyltransferase. *J Plant Biochem Biotechnol* **2018**, *27*, 13–27, doi:10.1007/s13562-017-0411-0.

25. Chen, Y.Y.; Xu, Z.S.; Xiong, A.S. Identification and Characterization of DcUSAGT1, a UDP-Glucose: Sinapic Acid Glucosyltransferase from Purple Carrot Taproots. *PLoS One* **2016**, *11*, doi:10.1371/JOURNAL.PONE.0154938.

26. McIntosh, C.A.; Owens, D.K. Advances in Flavonoid Glycosyltransferase Research: Integrating Recent Findings with Long-Term Citrus Studies. *Phytochemistry Reviews* **2016**, *15*, 1075–1091, doi:10.1007/s11101-016-9460-6.

27. Haroth, S.; Feussner, K.; Kelly, A.A.; Zienkiewicz, K.; Shaikhqasem, A.; Herrfurth, C.; Feussner, I. The Glycosyltransferase UGT76E1 Significantly Contributes to 12-O-Glucopyranosyl-Jasmonic Acid Formation in Wounded *Arabidopsis Thaliana* Leaves. *J Biol Chem* **2019**, *294*, 9858–9872, doi:10.1074/JBC.RA119.007600.

28. Cui, L.; Yao, S.; Dai, X.; Yin, Q.; Liu, Y.; Jiang, X.; Wu, Y.; Qian, Y.; Pang, Y.; Gao, L.; et al. Identification of UDP-Glycosyltransferases Involved in the Biosynthesis of Astringent Taste

Compounds in Tea (*Camellia Sinensis*). *J Exp Bot* **2016**, *67*, 2285–2297, doi:10.1093/JXB/ERW053.

29. Li, Y.; Li, X.L.; Lai, C.J.S.; Wang, R.S.; Kang, L.P.; Ma, T.; Zhao, Z.H.; Gao, W.; Huang, L.Q. Functional Characterization of Three Flavonoid Glycosyltransferases from *Andrographis Paniculata*. *R Soc Open Sci* **2019**, *6*, doi:10.1098/RSOS.190150.

30. Ono, E.; Homma, Y.; Horikawa, M.; Kunikane-Doi, S.; Imai, H.; Takahashi, S.; Kawai, Y.; Ishiguro, M.; Fukui, Y.; Nakayama, T. Functional Differentiation of the Glycosyltransferases That Contribute to the Chemical Diversity of Bioactive Flavonol Glycosides in Grapevines (*Vitis Vinifera*). *Plant Cell* **2010**, *22*, 2856–2871, doi:10.1105/TPC.110.074625.

31. Gasser, B.; Saloheimo, M.; Rinas, U.; Dragosits, M.; Rodríguez-Carmona, E.; Baumann, K.; Giuliani, M.; Parrilli, E.; Branduardi, P.; Lang, C.; et al. Protein Folding and Conformational Stress in Microbial Cells Producing Recombinant Proteins: A Host Comparative Overview. *Microb Cell Fact* **2008**, *7*, doi:10.1186/1475-2859-7-11.

32. Corsini, L.; Hothorn, M.; Scheffzek, K.; Sattler, M.; Stier, G. Thioredoxin as a Fusion Tag for Carrier-Driven Crystallization. *Protein Sci* **2008**, *17*, 2070–2079, doi:10.1110/PS.037564.108.

33. Ruggiero, A.; Masullo, M.; Ruocco, M.R.; Grimaldi, P.; Lanzotti, M.A.; Arcari, P.; Zagari, A.; Vitagliano, L. Structure and Stability of a Thioredoxin Reductase from *Sulfolobus Solfataricus*: A Thermostable Protein with Two Functions. *Biochim Biophys Acta* **2009**, *1794*, 554–562, doi:10.1016/J.BBAPAP.2008.11.011.

34. Cura, V.; Gangloff, M.; Eiler, S.; Moras, D.; Ruff, M. Cleaved Thioredoxin Fusion Protein Enables the Crystallization of Poorly Soluble ERalpha in Complex with Synthetic

Ligands. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **2008**, *64*, 54–57, doi:10.1107/S1744309107066444.

35. 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.; et al. Identification, Recombinant Expression, and Biochemical Analysis of Putative Secondary Product Glucosyltransferases from *Citrus Paradisi*. *J Agric Food Chem* **2016**, *64*, 1957–1969, doi:10.1021/acs.jafc.5b05430.

36. Kogawa, K.; Kato, N.; Kazuma, K.; Noda, N.; Suzuki, M. Purification and Characterization of UDP-Glucose: Anthocyanin 3',5'-O-Glucosyltransferase from *Clitoria Ternatea*. *Planta* **2007**, *226*, 1501–1509, doi:10.1007/S00425-007-0584-1.

37. Ciesla, W.P.; Bobak, D.A. *Clostridium Difficile* Toxins A and B Are Cation-Dependent UDP-Glucose Hydrolases with Differing Catalytic Activities. *J Biol Chem* **1998**, *273*, 16021–16026, doi:10.1074/JBC.273.26.16021.

38. Ford, C.M.; Boss, P.K.; Hæj, P.B. Cloning and Characterization of *Vitis Vinifera* UDP-Glucose. Flavonoid 3- O-Glucosyltransferase, a Homologue of the Enzyme Encoded by the Maize Bronze- 1 Locus That May Primarily Serve to Glucosylate Anthocyanidins in Vivo. *Journal of Biological Chemistry* **1998**, *273*, 9224–9233, doi:10.1074/jbc.273.15.9224.

39. Yamazaki, M.; Gong, Z.; Fukuchi-Mizutani, M.; Fukui, Y.; Tanaka, Y.; Kusumi, T.; Saito, K. Molecular Cloning and Biochemical Characterization of a Novel Anthocyanin 5-O-Glucosyltransferase by MRNA Differential Display for Plant Forms Regarding Anthocyanin. *J Biol Chem* **1999**, *274*, 7405–7411, doi:10.1074/jbc.274.11.7405.

40. Taguchi, G.; Ubukata, T.; Hayashida, N.; Yamamoto, H.; Okazaki, M. Cloning and Characterization of a Glucosyltransferase That Reacts on 7-Hydroxyl Group of Flavonol and 3-

Hydroxyl Group of Coumarin from Tobacco Cells. *Arch Biochem Biophys* **2003**, *420*, 95–102, doi:10.1016/j.abb.2003.09.027.

41. Noguchi, A.; Saito, A.; Homma, Y.; Nakao, M.; Sasaki, N.; Nishino, T.; Takahashi, S.; Nakayama, T. A UDP-Glucose:Isoflavone 7-O-Glucosyltransferase from the Roots of Soybean (*Glycine Max*) Seedlings. Purification, Gene Cloning, Phylogenetics, and an Implication for an Alternative Strategy of Enzyme Catalysis. *J Biol Chem* **2007**, *282*, 23581–23590, doi:10.1074/jbc.M702651200.

42. McIntosh, C.A.; Latchinian, L.; Mansell, R.L. Flavanone-Specific 7-O-Glucosyltransferase Activity in *Citrus Paradisi* Seedlings: Purification and Characterization. *Arch Biochem Biophys* **1990**, *282*, 50–57, doi:10.1016/0003-9861(90)90085-d.

43. Owens, D.K.; McIntosh, C.A. Biosynthesis and Function of Citrus Glycosylated Flavonoids. In *Recent Advances in Phytochemistry: The Biological Activity of Phytochemicals*; Gang, D., Ed.; Springer New York, 2011; Vol. 41, pp. 67–96 ISBN 978-1-4419-7299-6.

44. McIntosh, C.A.; Mansell, R.L. Biosynthesis of Naringin in *Citrus Paradisi*: UDP-Glucosyl-Transferase Activity in Grapefruit Seedlings. *Phytochemistry* **1990**, *29*, 1533–1538, doi:10.1016/0031-9422(90)80115-W.

45. Zea, C.J.; Camci-Unal, G.; Pohl, N.L. Thermodynamics of Binding of Divalent Magnesium and Manganese to Uridine Phosphates: Implications for Diabetes-Related Hypomagnesaemia and Carbohydrate Biocatalysis. *Chem Cent J* **2008**, *2*, doi:10.1186/1752-153X-2-15.

46. Shao, H.; He, X.; Achnine, L.; Blount, J.W.; Dixon, R.A.; Wang, X. Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago Truncatula*. *Plant Cell* **2005**, *17*, 3141–3154, doi:10.1105/tpc.105.035055.

47. 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. Structure of a Flavonoid Glucosyltransferase Reveals the Basis for Plant Natural Product Modification. *EMBO Journal* **2006**, *25*, 1396–1405, doi:10.1038/sj.emboj.7600970.

48. Brazier-Hicks, M.; Offen, W.A.; Gershater, M.C.; Revett, T.J.; Lim, E.K.; Bowles, D.J.; Davies, G.J.; Edwards, R. Characterization and Engineering of the Bifunctional N- and O-Glucosyltransferase Involved in Xenobiotic Metabolism in Plants. *Proc Natl Acad Sci U S A* **2007**, *104*, 20238–20243, doi:10.1073/pnas.0706421104.

49. Li, L.; Modolo, L. V; Escamilla-Trevino, L.L.; Achnine, L.; Dixon, R.A.; Wang, X. Crystal Structure of *Medicago Truncatula* UGT85H2 - Insights into the Structural Basis of a Multifunctional (Iso)Flavonoid Glycosyltransferase. *J Mol Biol* **2007**, *370*, 951–963, doi:10.1016/j.jmb.2007.05.036.

50. Modolo, L. V; Li, L.; Pan, H.; Blount, J.W.; Dixon, R.A.; Wang, X. Crystal Structures of Glycosyltransferase UGT78G1 Reveal the Molecular Basis for Glycosylation and Deglycosylation of (Iso)Flavonoids. *J Mol Biol* **2009**, *392*, 1292–1302, doi:10.1016/j.jmb.2009.08.017.

51. 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-Glucosyltransferase from *Clitoria Ternatea*. *Protein Science* **2015**, *24*, 395–407, doi:10.1002/pro.2630.

52. Zong, G.; Li, J.; Gao, Y.; Fei, S.; Liu, X.; Wang, X.; Shen, Y. Overexpression, Purification, Biochemical and Structural Characterization of Rhamnosyltransferase UGT89C1 from *Arabidopsis Thaliana*. *Protein Expr Purif* **2019**, *156*, 44–49, doi:10.1016/j.pep.2018.12.007.

CHAPTER 3. EXPRESSION AND PURIFICATION OF CP3GT: STRUCTURAL ANALYSIS AND MODELING OF A KEY PLANT FLAVONOL-3-O GLUCOSYLTRANSFERASE

Aaron S. Birchfield and Cecilia A. McIntosh*

Department of Biological Sciences, P.O. Box 70703, East Tennessee State University, Johnson City, TN 37614, USA; birchfieldas@etsu.edu

* Correspondence: mcintosc@etsu.edu; Tel.: +1-(423)-439-5838

In preparation for submission

Abstract

Glycosyltransferases (GTs) are pivotal enzymes in the biosynthesis of various biological molecules. This study focuses on the scale-up, expression, and purification of a plant flavonol-specific 3-O glucosyltransferase (Cp3GT), a key enzyme from *Citrus paradisi*, for structural analysis and modeling. Challenges associated with recombinant protein production in Pichia pastoris, such as proteolytic degradation, were addressed through optimization of culture conditions and purification processes. The purification strategy employed affinity, anion exchange, and size exclusion chromatography, leading to the production of Cp3GT with greater than 95% homogeneity. Advanced in silico modeling, using tools like D-I-TASSER and COFACTOR integrated with the AlphaFold2 pipeline, provided insights into the structural dynamics of Cp3GT and its ligand-binding sites, offering predictions for enzyme-substrate interactions. These models were compared to experimentally derived structures, enhancing our understanding of the enzyme's functional mechanisms. The findings present a comprehensive approach to produce highly pure Cp3GT, suitable for crystallographic studies, and shed light on the structural basis of flavonol specificity in plant GTs, with significant implications for synthetic biology and enzyme engineering in pharmaceutical applications.

Keywords: Glycosyltransferases (GTs), flavonols, recombinant protein production, *Pichia pastoris*, protein purification, structural analysis, *in silico* modeling, enzyme-substrate interactions, synthetic biology

3.1 Introduction

3.1.1 Role and Significance of Glycosyltransferases

Glycosyltransferases (GTs) are enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, playing vital roles in the biosynthesis of carbohydrates, lipids, proteins, and secondary metabolites. Ubiquitous across the plant and animal kingdoms, their functions are diverse and fundamental to various biological processes. In plants, GTs are particularly essential for the glucosylation of secondary metabolites, including flavonoids. Flavonoid glycosides serve crucial roles for plants, offering protection against herbivores, pathogens, and environmental stressors, as well as providing protection from ultra-violet light, and imparting unique taste and color profiles to edibles [1–9]. Furthermore, these glycosides have garnered significant attention in the human health domain for their roles in modulating inflammatory pathways, neutralizing reactive oxygen species, and inhibiting tumor cell proliferation [10–15].

3.1.2 GTs in Synthetic Biology and Enzyme Engineering

The rapidly advancing field of synthetic biology has cast a spotlight on the instrumental role of GTs, particularly in the realm of enzyme engineering and the tailored synthesis of complex molecules. Many GTs exhibit exceptional regio-, stereo-, and substrate specificities, making them invaluable tools for forming specific glycosidic bonds [16–19]. Through rational mutagenesis and directed evolution strategies, GT's have been engineered to facilitate the efficient synthesis of a diverse array of complex carbohydrates, glycoproteins, antibiotics, and herbal extracts with medicinal value [20–26]. The potential implications of these bioengineered

molecules are profound, particularly in the pharmaceutical industry where they play pivotal roles in drug discovery and therapeutic developments [27–29].

3.1.3 Characteristics of Flavonol Specific 3-O Glucosyltransferase

A flavonol specific 3-O glucosyltransferase (Cp3GT) was identified in *Citrus paradisi* and shown through biochemical characterization to preferentially glucosylate the flavonols quercetin, kaempferol, myricetin, and fisetin at the 3-OH position [30,31]. Mutational and *in situ* analyses suggested essential residues that play crucial roles in substrate docking and 3-O glucosylation [32,33] Sequence alignments further highlighted that Cp3GT shares a 56.7% sequence identity with a UGT from *Vitis vinifera* (VvGT1) and a 43% sequence identity with a UGT from *Clitoria ternatea* (UGT78K6). While VvGT1 can glucosylate both flavonols and anthocyanidins, UGT78K6 is specialized for the glucosylation of anthocyanidins alone. In contrast, Cp3GT exclusively catalyzes the glucosylation of flavonols [34–36].

A notable point of convergence among these glucosyltransferases is the conservation of the N-terminal histidine residue—specified as His20 in VvGT1, His22 in Cp3GT, and His17 in UGT78K6—critical for the 3-O glucosylation process [33,34,36]. The presence of this conserved catalytic residue in enzymes with diverse substrate specificities suggests that other structural elements or residues must play determinant roles in substrate recognition and specificity. To decipher the mechanistic underpinnings of these specificities, an in-depth structural and functional analysis of the binding pocket and the residues therein is warranted. Understanding the spatial configuration and properties of these residues can provide insights into substrate orientation and docking dynamics.

3.1.4 Structural Insights: Past to Present

The most recent Cp3GT *in silico* analysis, conducted in 2018, relied on EasyModeller for structure prediction and AutoDock Vina for docking [33]. Since then, significant advances in structure prediction methodologies have been made. Notably, the emergence of deep learning and machine learning-driven, artificial intelligence tools such as I-TASSER and AlphaFold have revolutionized the field [37,38]. Unlike conventional modeling tools such as EasyModeller, I-TASSER and AlphaFold employ advanced deep learning algorithms that provide a more dynamic understanding of protein folding [37,38]. These algorithms can accurately predict intricate local and global structural patterns in proteins and could outperform traditional techniques that primarily depend on template-based modeling.

Given the evolution of these newer, more advanced tools, there's a compelling need to revisit Cp3GT structural models and docking simulations. By comparing these with earlier models, analogous GTs with resolved crystal structures, and correlating them with biochemical insights from GT activity and mutational assays, a consolidated understanding of Cp3GT's structure can be achieved, potentially unearthing novel insights.

3.1.5 Challenges in Recombinant Protein Expression Using *Pichia pastoris*

The use of the *Pichia pastoris* system for producing recombinant proteins is popular due to its yield and scalability. However, its endogenous proteases, including serine, aspartic, and metalloproteases, pose challenges by potentially degrading expressed proteins and altering their structural integrity [39–42]. Such degradation can affect a protein's stability, activity, and its behavior during purification, leading to variability in enzymatic studies [43–45]. To counteract protease action, strains with reduced protease activity or genetically modified to eliminate

specific protease genes, such as PEP4, PRB1, and YPSs, are utilized [46]. Moreover, adjusting culture conditions and integrating protease inhibitors such as phenylmethylsulfonyl fluoride (PMSF) have shown efficacy in minimizing protein degradation [30,31,43,47,48].

3.1.6 Research Objectives and Hypotheses

This research is designed to test the hypothesis that recombinant Cp3GT expression, using the methanol-inducible *Pichia pastoris* system, can be scaled up to produce sufficient quantities of protein to purify for in-depth analysis. Furthermore, Cp3GT can be purified to ≥95% homogeneity utilizing standard chromatographic techniques, specifically affinity, anion exchange, and size-exclusion. This research also delves into the potential degradation of recombinant Cp3GT during purification by endogenous *Pichia pastoris* proteases. As such, it is designed to test the hypothesis that proteolytic degradation may result in a truncated protein lacking residues critical for GT catalysis or the C-terminal c-myc/6x-His recombinant tags necessary for purification and identification. Lastly, utilizing advanced *in silico* modeling and docking techniques, an *in silico* structural analysis of Cp3GT, focusing on identifying essential residues for flavonol binding and its subsequent 3-O glucosylation was conducted.

3.2 Methods

3.2.1 Cloning, Verification, and Transformation of Cp3GT into *Pichia pastoris*

The pPicZα plasmid, previously cloned to house the Cp3GT sequence, was initially transformed into *Escherichia coli* cells for propagation [49]. After cultivation, plasmid DNA was isolated and sequenced, confirming the wild-type Cp3GT gene's presence. Using site-directed mutagenesis, a thrombin cleavage sequence (Leu Val Pro Arg Gly Ser) was integrated upstream

of the C-terminal c-myc/6x His recombinant tags, allowing for potential post-purification tag removal (Fig. 3.20).

This modified plasmid was subsequently introduced into competent *Pichia pastoris* cells via electroporation as previously described [49]. To confirm the Cp3GT gene's successful integration into the *Pichia* genome, colony PCR was performed using 5' and 3' primers for the alcohol oxidase gene (AOX). DNA sequencing on PCR-positive colonies confirmed the correct integration of the Cp3GT gene with the thrombin cleavage sequence [49].

3.2.2 *Pichia pastoris* Growth and Expression

The growth curve and expression profile of transformed *P. pastoris* were determined by inducing with methanol at 28°C, as described previously [49], with samples taken at 4, 8, 21, 25, 29, 33, 45, and 72-hour intervals. Samples at each time point were analyzed for absorbance at an optical density (OD) of 600nm. At each time point, 2mL of culture was centrifuged at 1500 x g for 5 minutes at 4°C, resuspended in 200µL cell lysis buffer (50mM NaPO4, 1mM PMSF, 5mM βME, 5% glycerol, pH 7.4), and lysed using glass beads and vortexing at high speed. The lysed samples were microcentrifuged at max speed for 10 minutes at 4°C. The supernatant was transferred to a new tube and the crude protein concentration was estimated using absorbance at an OD of 280nm. For each time point, 100µg of crude protein was analyzed using SDS-PAGE and western blot.

3.2.3 Expression Scale-Up and Lysis

The following parameters were experimentally determined to be the most efficient methods for scaling up culture, expression, and purification. Transformed *P. pastoris* clones with recombinant Cp3GT were streaked on yeast peptone dextrose (YPD) agar and incubated at 28°C

for 3 days. Single colonies from these clones were inoculated into four respective 5mL buffered minimal glycerol yeast media (BMGY) aliquots in glass tubes, followed by incubation at 28°C with shaking at 250 rpm for 24 hours. From each of the four cultures, 1mL was used to inoculate 100mL BMGY aliquots in 250mL baffled flasks. These aliquots were then incubated at 28°C, shaking at 250 rpm, for another 24 hours. From each flask, 100mL of cell culture was collected, centrifuged at 1500 x g, and then resuspended in 1L aliquots of buffered minimal methanol (0.5%) yeast media (BMMY) to achieve a final OD_{600} of 1. These resuspended cells were placed into 2.8L baffled flasks and incubated for 24-hours at 28°C, with shaking at 250 rpm. Cells were then harvested by centrifugation at 1500 x g at 4°C. The combined 4L cell pellet was then resuspended in 100mL of cell lysis buffer and lysed using a ThermoSpectronic French Pressure Cell at 20,000 PSI for 5 cycles at 4°C. The lysate was centrifuged at 20,000 x g for 20 minutes at 4°C. The supernatant was transferred to a new vessel and PMSF was added to a final concentration of 1mM.

3.2.4 Cp3GT Cobalt-Affinity Chromatography

Prior to scaling up, the cell lysate was transitioned into Equilibration Buffer (50mM NaPO4, 300mM NaCl, 10mM Imidazole, 5mM βME pH 7.4) using PD-10 prepacked columns. These columns dictated a maximum volume of 2.5mL for desalting per column. Upon scaling up, this 2.5mL limitation introduced challenges, slowing the purification process, and elevating the risk of potential Cp3GT degradation. The primary motive for transitioning into the Equilibration Buffer was its composition, particularly the presence of a small amount of imidazole and the high NaCl concentration. These components effectively reduced non-specific binding. This reduction became especially crucial as the increased cell quantity during scale-up exacerbated non-specific interactions.

To address these challenges, instead of conventional desalting, the cell lysate supernatant, post-centrifugation, was transferred to a new tube and diluted 10-fold with the Equilibration Buffer. This dilution served as an alternative to direct buffer exchange through desalting. It effectively transitioned the protein into the desired buffer conditions and, based on experimental data, enhanced the purification process overall.

The lysate was then evenly distributed across 12, 3.25mL His-Pur Cobalt Resin (Thermo Scientific 89964) aliquots in gravity flow columns, previously equilibrated in Equilibration Buffer. After applying the lysate, each column was washed with 200mL of wash buffer (50mM $NaPO_4$, 300mM NaCl, 20mM Imidazole, pH 7.4) in 50mL intervals. A small sample was collected after every 50mL interval for analysis. Cp3GT was eluted in 10 mL of elution buffer (50mM $NaPO_4$, 250mM Imidazole, pH 7.4) in 1mL fractions. Crude lysate, flowthrough, wash, and eluate were analyzed using SDS-PAGE and western blot.

3.2.5 Cp3GT Anion Exchange and Size Exclusion Chromatography

The cobalt-affinity eluate fractions containing Cp3GT were collected. A small sample was dialyzed to remove imidazole and promptly tested for GT activity, followed by visualization using SDS-PAGE and western blot. The remaining eluate was placed into an Amicon Ultra-15 centrifugal concentrator (Millipore UFC9030) and concentrated according to the manufacturer's instructions. Buffer A (25mM Bicine, 14mM βME pH 8.5) was added to the centrifugal concentrator to a final volume of 15mL. The eluate was again concentrated. This process was repeated two more times to ensure the cobalt-affinity elution buffer was completely exchanged into Buffer A.

The concentrated, desalted His-Pur eluate was loaded into a 10mL Superloop (Cytiva 18111381) and injected onto a MonoQ 5/50 GL (Cytiva 17516601) prepacked column using a GE ÄKTA Purifier Fast Protein Liquid Chromatography (FPLC) machine at a flow rate of 0.5mL/min. The flowthrough was collected in 5mL volumes. Cp3GT was eluted using a 20% gradient of Buffer B (25mM Bicine, 1M NaCl, 5mM βME) for 48 minutes in 0.5mL volumes at a flow rate of 0.5mL/min. The flowthrough and all peaks showing absorbance were assayed for activity and analyzed using SDS-PAGE, western blot, and for low-protein samples, silver stain.

The MonoQ fractions containing Cp3GT were combined and then concentrated to a final volume of 150µL in Buffer A. After concentration, the sample was loaded into a 100µL sample loop. This was then injected onto a Superdex 75 10/300 GL (Cytiva 17-5174-01) prepacked column, which had been previously equilibrated with MonoQ Buffer A. The process utilized FPLC at a flow rate of 0.5mL/min. Resulting peaks from the process were analyzed using SDS-Page and western blot. For samples with low protein content, silver stain was used.

3.2.6 Cp3GT Glucosyltransferase Activity Assay

Glucosyltransferase activity was assayed by measuring the incorporation of a radiolabeled ^{14}C-Glucose as previously described [50,51]. Briefly, 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, 5-15µL of purified enzyme (added upon starting), and enough 50mM phosphate buffer (pH 7.5 containing 14mM βME) to make a total reaction volume of 75uL [50]. Reactions were run for 5-45 minutes at 37°C and stopped with 10µL of 6M HCl. Glucosylated product was extracted by adding 250µL ethyl acetate and the reaction mix was vortexed for 20 seconds, followed by microcentrifugation at max speed for 10 seconds. From the top layer 100µL was taken and added to 2.5mL CytoScint Liquid Scintillation

Cocktail (MP Biomedicals 882453) and counted for 2 minutes in a Beckman Coulter LS 6500 Scintillation Counter. Glucosylated product amount in nanomoles was determined by multiplying the counts per minute (cpm) by 2.5 to correct for volume and then divided by 400 (400cpm/nmol UDP-G).

To quickly assess GT activity in column fractions, a screening assay was implemented as previously described [50] with the following exceptions: 10µL UDP-14C glucose (containing 20,000cpm) taken directly from source vial with no other UDP-G added and 2-20µL of the column fraction was used. Reactions were conducted for 2-5 minutes.

3.2.7 Model Generation and Ligand Binding Site Prediction

Cp3GT models were generated using the Distance-Guided Protein Structure Prediction (D-I-TASSER) server (https://zhanggroup.org/D-I-TASSER/) [37]. Options were selected to predict protein function based on structural models, utilize the large IMG/JGI metagenomic database for MSA construction, and implement the integrated AlphaFold2 pipeline. Models were ranked based on their TM-scores, and the model with the highest-ranking score was selected for structural analysis.

All ligand binding site predictions were generated using the D-I-TASSER-integrated COFACTOR algorithm (https://zhanggroup.org/COFACTOR/) [52,53]. The highest-ranking template that included a flavonol ligand, preferably quercetin or kaempferol, was chosen for analysis. Structural alignments were generated using UCSF Chimera (https://www.cgl.ucsf.edu/chimera/) [54]. Models were analyzed using Biovia Discovery Studio [55] and Maestro [56].

3.3 Results and Discussion

3.3.1 Growth and Expression Analysis

During induction, *P. pastoris* exhibited exponential (or logarithmic) growth for the initial 20 hours, after which it entered the stationary phase and growth plateaued (Fig. 3.1 A). Cp3GT expression was detected at 4hrs and continued to increase until 25hrs, after which no increase in expression was detected (Fig. 3.1 B and C). The band corresponding to Cp3GT expression was estimated to be 65.6kDa. A band approximating 56.6 kDa emerged faintly at 20 hours, intensifying until its peak around 45 hours. Due to concerns about Cp3GT degradation, subsequent cultures were limited to a 20-hour growth duration.

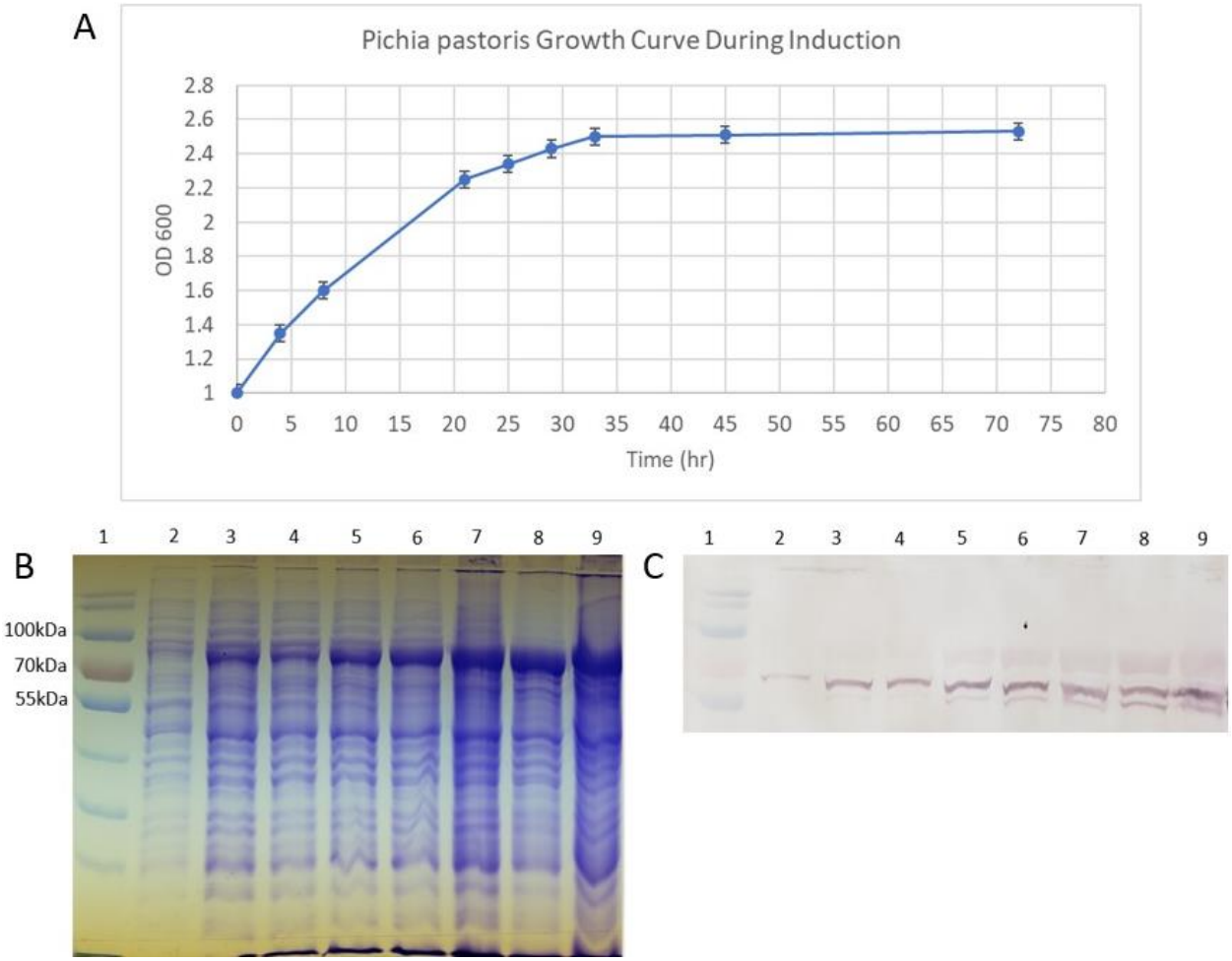

Fig. 3.1 Growth and Expression Analysis of *P. pastoris* Transformed with Recombinant Cp3GT. A. Growth curve during induction. B. Cp3GT Expression Profile, Coomassie stained, and C. western blot loaded with 1. Molecular Weight Marker (MWM), 2. 4hr, 3. 8hr, 4. 21hr, 5. 25hr, 6. 29hr, 7. 33hr, 8. 45hr, 9. 72hr.

The exponential growth phase during the first 20 hours signifies a robust metabolic state for the yeast, providing an ideal cellular environment for recombinant protein expression. Notably, Cp3GT's expression peaking at around 25 hours aligns well with the transition of the yeast cells from the exponential to stationary phase. This might suggest an optimal window for recombinant protein synthesis before the cells divert resources in response to changing growth conditions. The presence of a 56.6 kDa band, appearing later in the expression timeline, could be indicative of potential proteolytic degradation, which warranted further investigation. The decision to restrict culture growth to 20 hours for subsequent experiments seems prudent given this potential degradation, as it capitalizes on the optimal expression window while reducing the chances of unwanted protein modifications or breakdown.

3.3.2 Cp3GT Cobalt-Affinity Chromatography Analysis

During scale-up experiments, it was determined that a maximum of 250mL of lysed culture could be applied to 2.5mL affinity resin before Cp3GT was detected in the flowthrough fraction. Therefore, a total of 40mL affinity resin was used to purify protein from cells from 4L of culture. Resin volumes of more than 5mL increased non-specific binding and reduced flow rates, therefore the 40mL of resin was distributed across 12 columns in 3.25mL aliquots. Coomassie stain and western blot analyses showed Cp3GT in the crude lysate but not in the flowthrough or wash fractions (Fig. 3.2 A and C). No bands were detected after washing with 150mL wash buffer (Fig. 3.2 A). Cp3GT eluted in fractions 4-7 taking 0.5mL fractions (Fig. 3.2

B and D). Despite exhaustive washing until no bands were visible in the wash fractions, the eluate still displayed multiple bands, most prominently a 100kDa protein of unknown identity (Fig. 3.2 B). Interestingly, the intensity of the 100kDa protein in the wash fractions diminished with increased washing, implying its partial removal (Fig. 3.2 A). Its co-elution with Cp3GT in the eluate hints at an unexplained protein-protein interaction or the presence of regions rich in histidine residues.

As seen in Fig. 3.2 B and D, an additional band in the eluate was detected with a similar difference in size to Cp3GT as seen in Fig. 3.1. The presence of this band in the cobalt affinity eluate, combined with its detection on the western blot, implies the existence of a c-myc/6x His tag, strongly hinting at it being a Cp3GT degradation product.

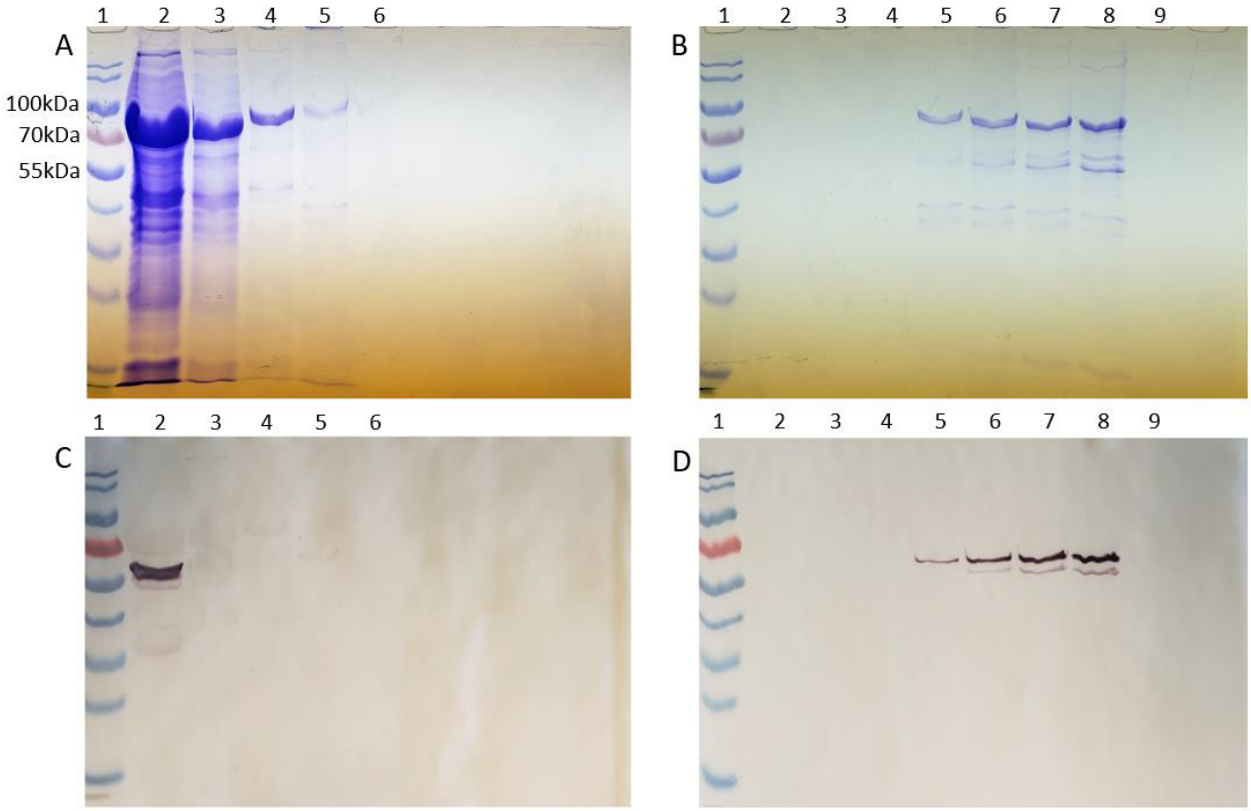

Fig. 3.2 Cobalt-affinity Chromatography Purification Profile. A. Coomassie stain and C. western blot loaded with 1. MWM, 2. Crude lysate, 3. Flowthrough, 4. Wash 1, 5. Wash 2, 6. Wash 3. B. Coomassie stain and D. western blot loaded with 1. MWM 2. Elution 1, 3. Elution 2, 4. Elution 3, 5. Elution 4, 6. Elution 5, 7. Elution 6, 8. Elution 7, 9. Elution 8. Wash fractions were collected every 50mL. Elution fractions were collected every 0.5mL.

The observed upper limit in the volume of lysed culture that could be applied to the affinity resin without detecting Cp3GT in the flowthrough sheds light on the importance of optimizing scale-up conditions for maximal protein recovery. Distributing the affinity resin across multiple columns was a logical adaptation to prevent non-specific binding and maintain effective flow rates. The co-elution of Cp3GT with the 100kDa protein raises questions about potential interactions between the two proteins or shared characteristics, such as histidine-rich regions that facilitate affinity to the resin. Regardless, the cobalt-affinity purification method applied in this study was an effective first step. To achieve the level of purity necessary for high-resolution structural analyses, supplementary chromatographic steps are frequently required [34–36,57].

The detection of the suspected Cp3GT degradation product in the cobalt affinity eluate and its appearance in the western blot is indicative that the product contains the C-myc tag. The presence of the c-myc/6x His tag implies that degradation is occurring at the N-terminus because the C-terminal tags remain intact and detectable. This suggests that whatever proteolytic activity is at play, it's specifically targeting the N-terminal region without compromising the C-terminus. This selective degradation can potentially impact the functionality and stability of Cp3GT. To maintain the integrity of recombinant proteins throughout the purification process, it's imperative to pinpoint the exact causes of such degradation. The specificity of this N-terminal degradation

might also hint at cellular proteases or external factors that cleave proteins at motifs or regions. The consistent and timely use of PMSF during cell handling and lysis reduced degradation during purification. Ensuring that Cp3GT remains intact throughout the purification process is critical not only for achieving optimal yields but also for guaranteeing the accuracy and reliability of downstream analyses.

3.3.3 Anion Exchange Chromatography

Considering Cp3GT's calculated pI of 6.38, it was hypothesized that anion exchange chromatography could enhance Cp3GT's purification. Given MonoQ's efficiency and its proven track record in purifying plant GTs, it was chosen as the preferred anionic resin [50,58,59]. After loading the pooled and dialyzed affinity eluate on MonoQ, the flowthrough was collected in 5mL fractions. Cp3GT was initially eluted using a 100% gradient of Buffer B over 30 mL collecting 1mL fractions. Four peaks labeled 1-4 were detected across fractions 8-11 respectively between approximately 130mM and 250mM NaCl (Fig. 3.3 A). The trailing end of peak 4 was captured in fraction 12 and was also tested (Fig. 3.3). Coomassie stain showed no discernible bands in peaks 1-3, but two bands approximately 9kDa apart appeared in peak 4, with most of the protein in fraction 11 (Fig. 3.3 B). Western blot showed a single band in peak 3 that measured 66kDa, while peak 4 contained two bands that measured 65 and 56kDa (Fig. 3.3 C).

It was hypothesized that degradation during Cp3GT's purification resulted in the loss of approximately 9kDa at the N terminus, equivalent to around 80 amino acids, which would change its theoretical pI from 6.38 to 5.92. Such a change might affect the protein's anionic potential, influencing its binding affinity to the column and causing it to elute at higher NaCl concentrations (Fig. 3.3). It is also possible that the combination of high cell density from the scale-up and PMSF's instability in aqueous solutions was leading to insufficient inhibition of

endogenous proteases. To address this, subsequent purifications introduced PMSF not only during cell lysis but also post-centrifugation of the lysed cells. By adding PMSF to the supernatant just before loading it onto the affinity column, the emergence of the second band was markedly reduced.

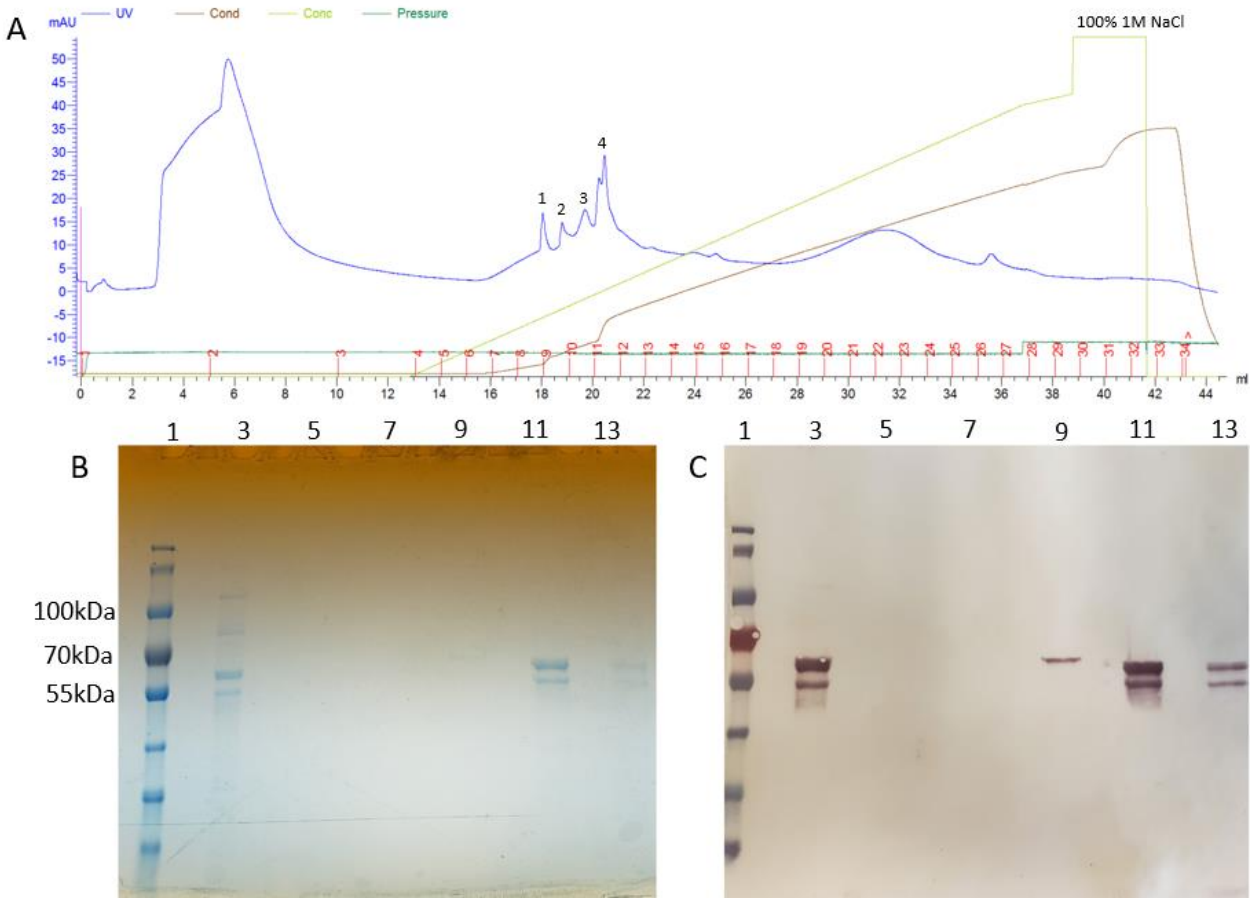

Fig. 3.3 MonoQ Anion Exchange Chromatography of Cobalt-Affinity Purified Cp3GT Eluate #1. A. MonoQ chromatogram showing 4 peaks of interest labeled 1-4 eluted between 130-250mM NaCl, eluted using a 100% gradient of Buffer B over 30mL collecting 1mL fractions. The y-axis is milli absorbance units (mAU) and the x axis is mL buffer (black). Fraction numbers are denoted as red dashes along the x-axis. B. Coomassie stain and C. western blot loaded with 1.

MWM, 3. Pooled cobalt affinity eluate, 5. Peak 1, 7. Peak 2, 9. Peak 3, 11. Peak 4 (Fraction 11), 13. Peak 4 (Fraction 12). All other lanes contained loading dye.

The employment of anion exchange chromatography using MonoQ demonstrated mixed outcomes in the pursuit of Cp3GT purification. The observation of different elution peaks underpins the intricate and dynamic nature of protein purification. As discussed previously, the detection of Cp3GT at varying sizes, as evidenced by the western blot, raises an inherent question about the protein's structural integrity during the purification process. The 9kDa difference, estimated to correspond to approximately 80 amino acids, indicates a degradation event. Given the revised theoretical pI of 5.92 for the truncated protein, one can infer that the change in pI due to N-terminal truncation might have implications for the protein's interaction with the anionic column, leading to a shift in its elution profile. This is notable by an increased presence of this degradation band eluting at slightly higher salt concentrations.

Furthermore, the possibility of inadequate protease inhibition, especially with PMSF's instability in aqueous solutions, becomes a pressing concern. The practice of adding PMSF post-centrifugation was performed as a response to these challenges. Its apparent efficacy in curbing the emergence of the second band may point to the broader necessity for dynamic adjustments in protein purification workflows, especially when handling labile or degradation-prone proteins.

3.3.3.1 Anion Exchange Chromatography: Refining the Gradient and Verifying Cp3GT Integrity. In a subsequent purification, both the crude and the affinity purified eluate showed two bands at 65 and 56kDa on a western blot (Fig. 3.6). However, the 65kDa band showed greater intensity, and the 56kDa band showed less intensity than in the previous purification (Fig. 3.3 B and C). The 56kDa degradation product was not present in the MonoQ fractions analyzed using

SDS-PAGE (Fig. 3.6). Furthermore, the approximately 100kDa band seen previously appeared strongly in the affinity eluate (Fig. 3.6).

The affinity eluate was again dialyzed and purified using MonoQ. A more shallow gradient of 20% Buffer B over 50mL was employed to better distinguish any peaks, and fractions were collected in 0.5mL increments (Fig. 3.4). Having previously determined that Cp3GT did not elute past 250mM NaCl (Fig. 3.3), the three peaks at fractions 14 (72mM NaCl), 16 (92mM NaCl), and 18 (104mM NaCl) were subjected to a kinetic assay for activity (Fig. 3.5) and analyzed using SDS-PAGE and western blot (Fig. 3.6).

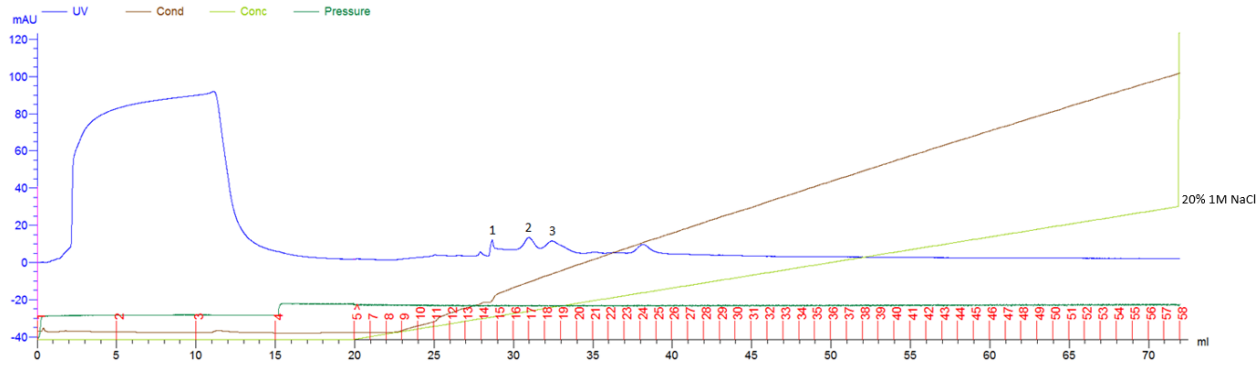

Fig. 3.4 MonoQ Anion Exchange Chromatography of Cobalt-Affinity Purified Cp3GT Eluate #2. Three peaks of interested labeled 1-3 eluted at 72, 92, and 104mM NaCl, respectively, using a 20% gradient of Buffer B over 50mL.

Peak 1 showed no GT activity (Fig. 3.5) and showed a single, 100kDa band on Coomassie stain (Fig. 3.6 A), and a single, faint, 65kDa band on western blot (Fig. 3.6 B). Peak 2 showed moderate GT activity (Fig. 3.5) and showed both a 100kDa band and a 65kDa band on Coomassie stain (Fig. 3.6 A), and a single 65kDa band on western blot (Fig. 3.6B). Peak 3 had the highest GT activity (Fig. 3.5) and again showed a 100kDa band and a 65kDa band on Coomassie (Fig. 3.6 A), and a single 65kDa band on western blot with greater intensity than

peaks 1 and 2 (Fig. 3.6)). GT kinetic assay and western blot confirmed that the 65kDa protein is Cp3GT.

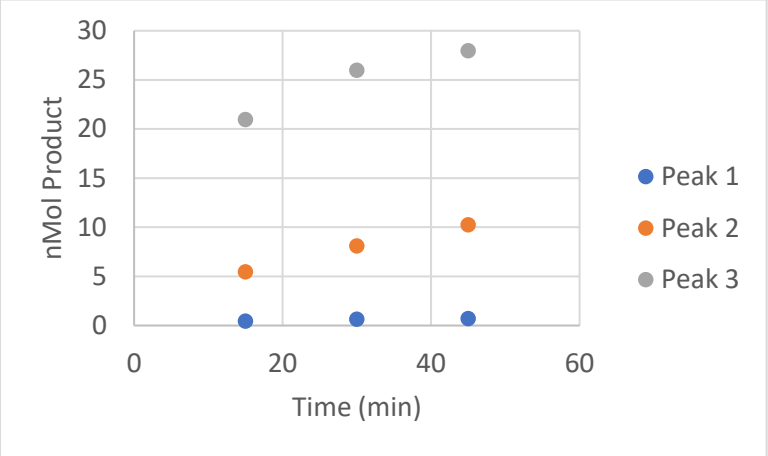

Fig. 3.5 MonoQ Peaks 1-3 GT Activity Assay Time Course. Assay conducted using 10uL of each sample for 45 minutes.

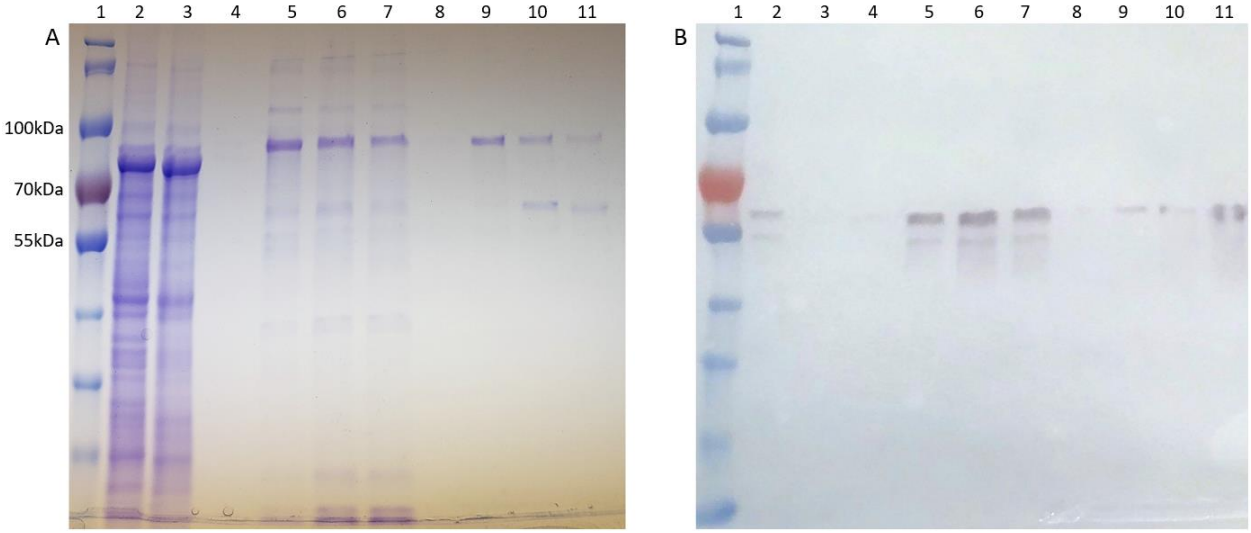

Fig. 3.6 Cp3GT Cobalt-affinity and MonoQ Purification Profile #2. A. Coomassie stain and B. western blot loaded with 1. MWM, 2. Crude lysate, 3. Cobalt-affinity flowthrough, 4. Affinity

elution fraction 1, 5. Affinity elution fraction 2, 6. Affinity elution fraction 3, 7. Affinity elution fraction 4, 8. Loading dye, 9. MonoQ peak 1, 10. MonoQ peak 2, 11. MonoQ peak 3. Affinity elution fractions were collected every 1mL.

The optimization of the MonoQ gradient revealed critical insights into the Cp3GT purification process. The appearance of two distinct bands in both the crude and affinity-purified eluate reaffirms the tendency for Cp3GT degradation, yet the modified intensity of the 65kDa band suggests improved preservation of the enzyme's integrity. The absence of the 56kDa degradation product in MonoQ fractions, coupled with the strong presence of the approximately 100kDa band, prompts further investigation into the factors contributing to Cp3GT stability post-affinity purification. It is imperative to understand the nature of this 100kDa protein, as it consistently appears alongside Cp3GT during purification.

Employing a more shallow MonoQ gradient notably enhanced Cp3GT peak resolution, underscoring the importance of gradient adjustments in complex protein purification processes. The differential GT activity across the peaks, combined with the consistent presence of the 65kDa band, corroborates the identity of the Cp3GT. Furthermore, the peak with the highest GT activity showcasing a more intense 65kDa band on the western blot further solidifies the assertion that this band corresponds to the intact Cp3GT. The presence of the 100kDa protein across all peaks, however, remains a perplexing observation, even as it appears most strongly in peak 1. Understanding its interaction with Cp3GT, whether it's merely a co-eluting protein, a Cp3GT binding partner, or a binding competitor, is paramount for future purification and functional studies.

3.3.3.2 Extending Gradients Improve Anion Exchange Chromatography Purification. The observation of Cp3GT co-eluting in two peaks with the unidentified 100kDa protein during

anion exchange chromatography prompted the hypothesis that a more shallow NaCl gradient might separate them. To test this, an additional purification was performed using an elution gradient of 20% buffer B over 60mL. Two strong peaks labeled Peak A and Peak B emerged at 72 and 83mM NaCl respectively and were analyzed for GT activity (Fig. 3.8). Peak A exhibited roughly 50% more GT activity than Peak B (Fig. 3.8). While Peak A displayed a confirmed 65kDa band corresponding to Cp3GT, Peak B interestingly revealed a slightly smaller 64kDa band (Fig. 3.9 A: Lane 10). Despite its reduced size, the 64kDa band is likely still Cp3GT, given its GT activity (Fig. 3.8) and its identification through western blot (Fig. 3.9 B). Because the positive western blot indicates the presence of C-terminal tags, it is possible that minor N-terminal degradation may have occurred in Peak B protein resulting in the loss of approximately 9 amino acids. It is also possible that Peak B activity (Fig. 3.8) is due to some residual presence of Peak A due to peak overlap (Fig. 3.7).

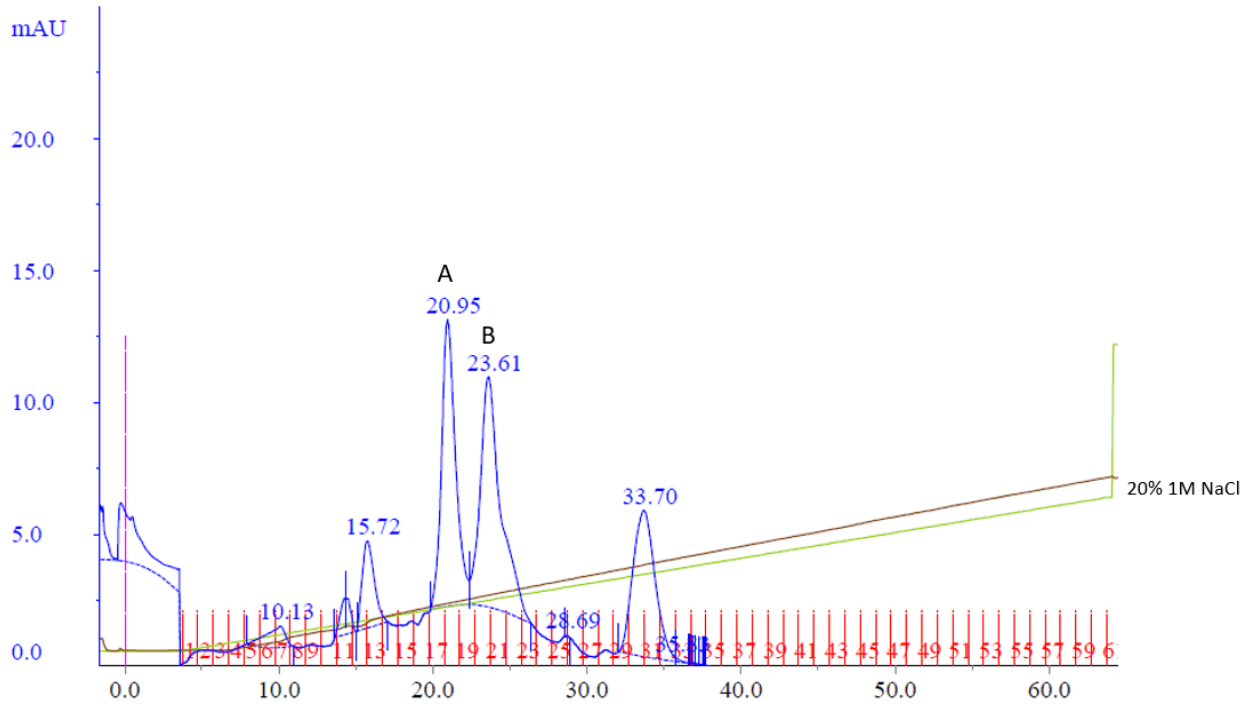

Fig. 3.7 MonoQ Anion Exchange Chromatography of Cobalt-Affinity Purified Cp3GT Eluate #3. Two peaks of interested labeled A and B eluted at 72 and 83mM NaCl respectively using a 20% gradient of Buffer B over 60mL.

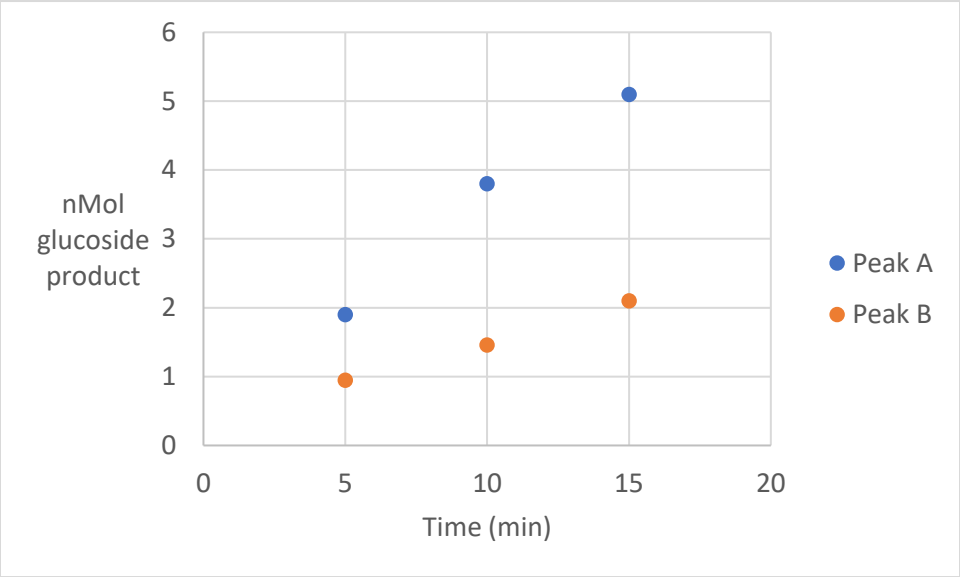

Fig. 3.8 MonoQ Peaks A and B GT Activity Assay Time Course. Assay conducted using 5uL of sample for 15 minutes.

The Coomassie stain showed a trend across the two peaks, where the 100kDa protein contaminant eluted by itself in fraction 17, and then co-eluted with Cp3GT until fraction 20, lessening in intensity as it went (Fig. 3.9 A). In fraction 21, the Cp3GT eluted as a single band without the 100kDa contaminant (Fig. 3.9). The GT kinetic assay and western blot again confirmed that the protein in peak A is Cp3GT. With regards to the 56kDa degradation product, both the crude and the pooled affinity eluate showed Cp3GT and 56kDa bands on Coomassie and western blot (Fig. 3.9) as previously described. However, as in Fig. 3.6, the degradation product

band did not appear anywhere in the MonoQ elution. Thus, introducing additional PMSF was effective in slowing degradation.

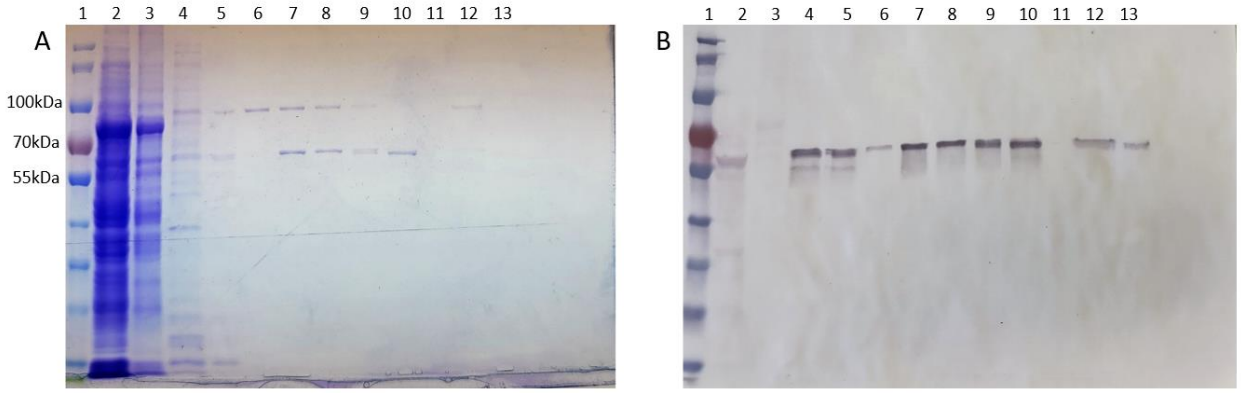

Fig. 3.9 Cp3GT Cobalt-affinity and MonoQ Anion Exchange Purification Profile #3: A. Coomassie stain and B. western blot loaded with 1. MWM, 2. Crude lysate, 3. Cobalt-affinity flowthrough, 4. Cobalt-affinity eluate pooled 1, 5. Cobalt-affinity eluate pooled 2, 6. Peak A (fraction 17), 7. Peak A (fraction 18), 8. Fraction 19, 9. Peak B (fraction 20), 10. Peak B (fraction 21), 11. Loading dye, 12. Cp3GT control, 13. Cp3GT control.

The methodological refinement to employ a shallower NaCl gradient provided clearer resolution between Cp3GT and the 100kDa contaminant, as evidenced by the differential elution patterns observed. The emergence of two peaks, Peak A and Peak B, both exhibiting GT activity, further illustrates the complex nature of Cp3GT purification. The appearance of a 64kDa band in Peak B, juxtaposed with GT activity in that sample and presence in the western blot, suggests its identity as a slightly degraded Cp3GT variant and/or the presence of some protein overlap from Peak A. The hypothesis of minor N-terminal degradation leading to the loss of approximately 9 amino acids in the Peak B protein underscores the challenge of Cp3GT stability during purification. Concurrently, the possibility of Peak B activity being influenced by residual Peak A due to overlap cannot be dismissed and warrants further investigation.

While the observed trend in Cp3GT elution, particularly the individual elution of the 100kDa protein in fraction 17, followed by its co-elution with Cp3GT and eventual disappearance, is notable, the primary objective of this research was to purify Cp3GT away from this protein. This elution pattern could provide valuable insight into potential interactions or affinities between the two proteins, however, unraveling the nature of their relationship may be reserved for future research endeavors, keeping the current focus here on Cp3GT purification.

The addition of supplementary PMSF during the MonoQ chromatography step has proven to be a successful strategy in mitigating degradation, as evidenced by the absence of the 56kDa product during MonoQ chromatography. This successful intervention underscores the precision of this approach in enhancing Cp3GT's stability during critical purification junctures. While the degradation product was observed in the earlier stages, the corrective measures applied in subsequent steps have demonstrated a capability to refine the process and maintain the integrity of Cp3GT.

3.3.4 Cell Handling and Storage Modifications Reduce Cp3GT Degradation and Further Refine Anion Exchange Chromatography

It was hypothesized that the previous approach of storing induced recombinant cells whole at -80°C could subject Cp3GT to potential degradation or denaturation during freeze-thaw cycles, thereby increasing the exposure time to intracellular proteases. To test this hypothesis, the cell handling procedure was modified. Cells were immediately lysed post-induction and centrifuged to separate cellular debris. The resulting cell lysate supernatant was swiftly treated with 1mM PMSF and 20% glycerol before storage at -80°C. This adjusted strategy aimed to minimize degradation risks by reducing Cp3GT's exposure to proteolytic agents and promptly inhibiting proteases with PMSF.

Cp3GT purification proceeded as previously described using affinity and MonoQ chromatography. The pooled affinity eluate was separated on the MonoQ column using a 15% gradient of Buffer B over 37mL and 0.5mL fractions were collected. Two strong peaks emerged at 78mM NaCl and 90mM NaCl, eluting across fractions 34-40 (Peak A) and 41-48 (Peak B), with the shallow gradient employed (Fig. 3.10).

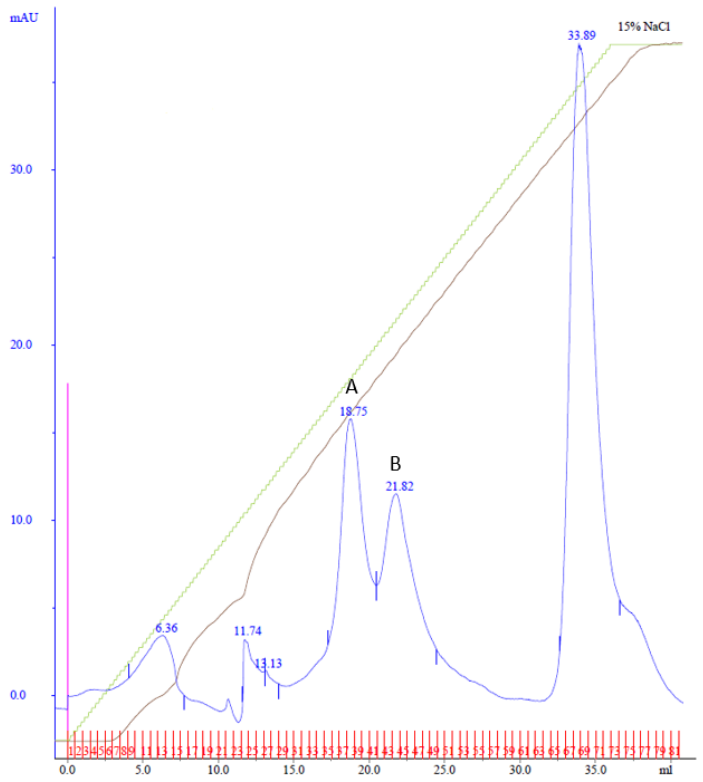

Fig. 3.10 MonoQ Anion Exchange Chromatography of Cobalt-Affinity Purified Cp3GT Eluate #4. Two peaks of interest labeled A and B eluted at 78 and 90mM NaCl respectively using a 15% gradient over 37mL.

These fractions were tested for GT activity using a screening assay designed to identify the presence or absence of enzymatic activity, and then analyzed using SDS-PAGE (Fig. 3.11). GT activity was detected in peaks A and B. As seen in previous purifications, a 100kDa band

eluted immediately preceding Cp3GT elution and proceeded to co-elute with Cp3GT across both peaks (Fig. 3.11). The cell lysate and pooled affinity eluate showed Cp3GT presence as confirmed by SDS-PAGE, but the 56kDa degradation product was not detected (Fig. 3.11).

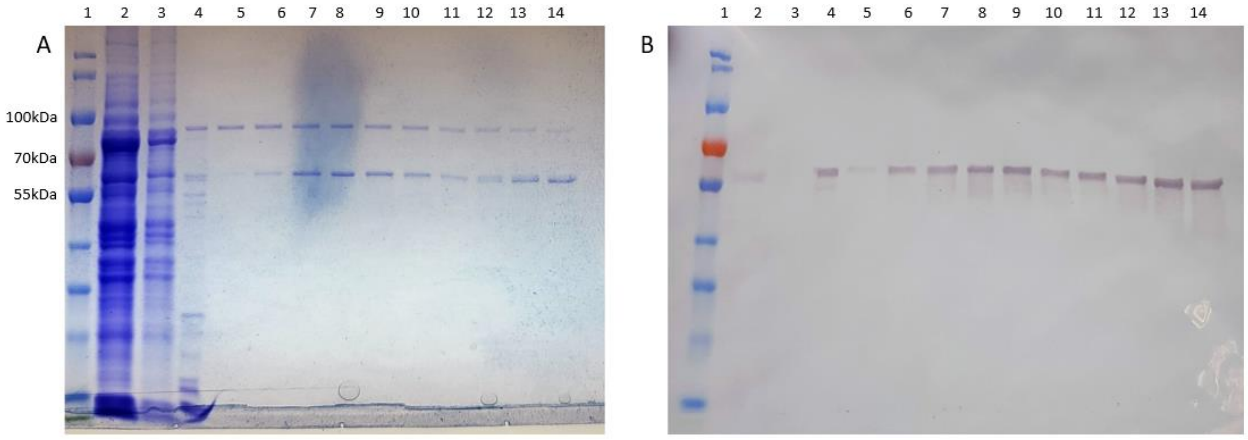

Fig. 3.11 Cp3GT Cobalt-affinity and MonoQ Anion Exchange Purification Profile #4. A. Coomassie stain and B. western blot loaded with 1. MWM, 2. Crude lysate, 3. Affinity flowthrough, 4. Pooled affinity eluate, 5-10. Peak A fractions 35-40, 11-14. Peak B fractions 41-44.

The adjustment in cell handling showed that by promptly lysing cells post-induction, segregating cellular debris, and rapidly administering PMSF and glycerol, the strategy aimed to minimize Cp3GT's vulnerability to intracellular proteases. Such modifications potentially confer dual advantages. Firstly, they might protect Cp3GT from proteolytic degradation, enhancing yields and integrity. Secondly, the expedited addition of protease inhibitors could further shield the protein from premature degradation. One of the most significant outcomes of this revised approach is the absence of the 56kDa degradation product in all stages of purification. Thus, this is the process used for all further purification.

The chromatographic findings following these adjustments are noteworthy. The emergence of two distinct peaks, designated as Peak A and B, at specific NaCl concentrations, may indicate the presence of two differentially charged species or forms of Cp3GT under the chromatographic conditions employed. Although the 100kDa band has been a consistent observation in previous stages, its implications and potential association with Cp3GT remain to be fully elucidated, highlighting a nuanced challenge in the purification process.

These results identified several considerations for optimizing purification processes, notably the avoidance of cell freezing prior to lysis, the early incorporation of PMSF, and the implementation of a gradual gradient during MonoQ chromatography. Additionally, gel filtration was investigated as a potential concluding step in the purification protocol.

3.3.5 Cp3GT Size Exclusion Chromatography Analysis

It was hypothesized that Cp3GT could be effectively separated from the 100kDa contaminant protein using size exclusion chromatography. To test this hypothesis, all MonoQ elution fractions containing Cp3GT were pooled, concentrated, and applied to a Superdex 75 size exclusion column. The aim of this strategy was to purify Cp3GT to apparent homogeneity, setting the stage for future crystallographic and other structural analyses.

The resulting chromatogram revealed three distinct peaks labeled SEC Peak 1-3 (Fig. 3.12). Coomassie staining faintly detected Cp3GT in peak 1, however, it was more visible by silver stain and western blot (Fig. 3.13). Notably, despite their stronger absorbance, no proteins were detected by Coomassie in peaks 2 and 3, however there were faint bands detected using western blot (Fig. 3.13). The 100kDa contaminant was absent, suggesting its elution prior to Cp3GT. Fractions before the first peak, showing low absorbance, were not analyzed.

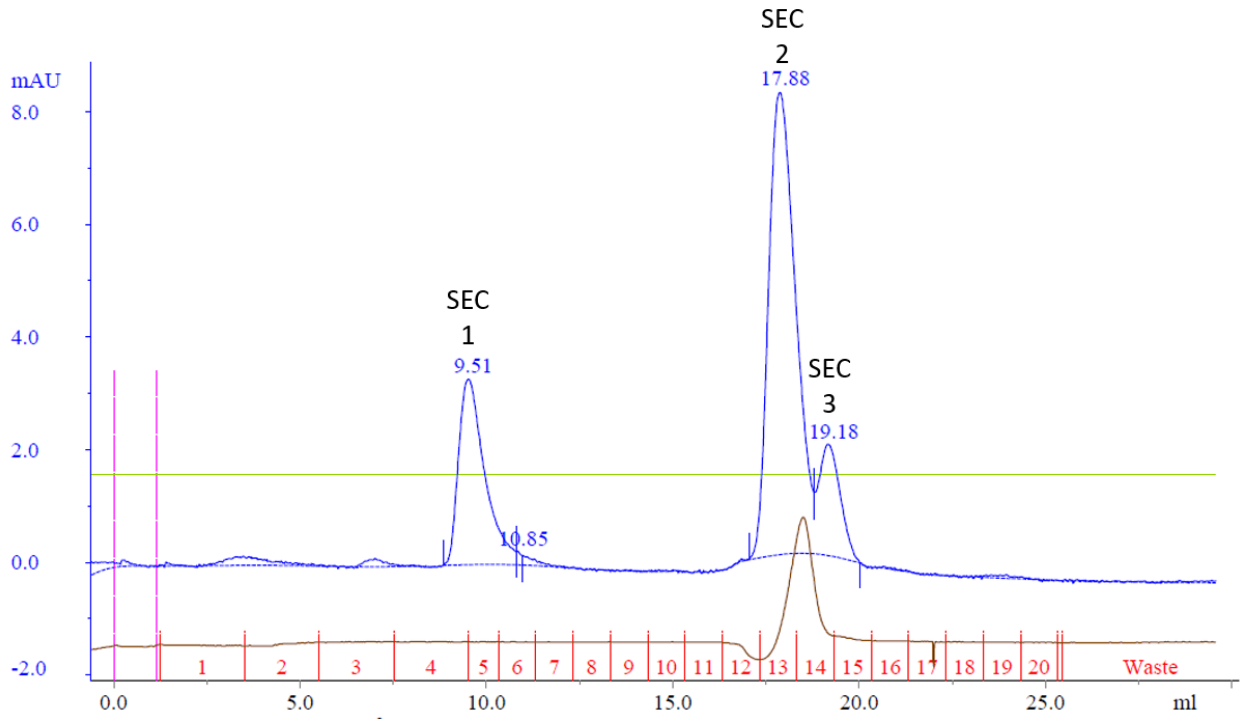

Fig. 3.12 Size Exclusion Chromatography of MonoQ Purified Cp3GT: Three peaks of interest labeled SEC 1-3 were analyzed for separation of Cp3GT from 100kDa protein.

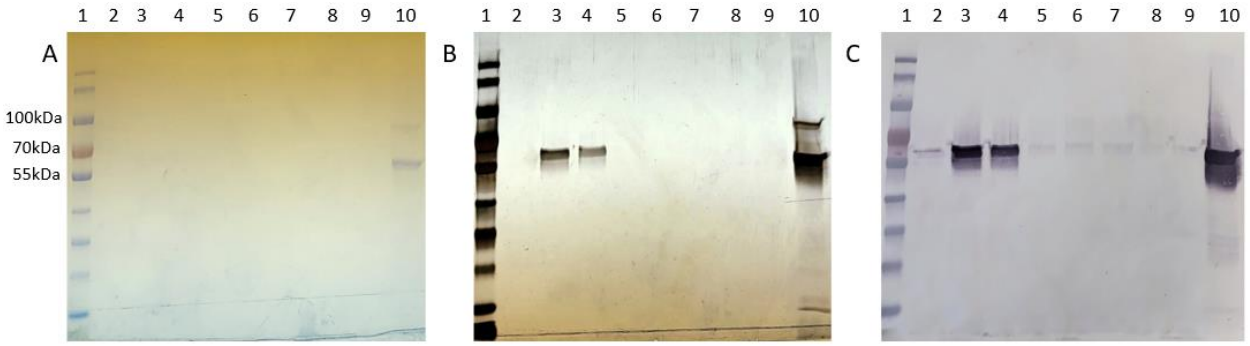

Fig. 3.13 Cp3GT Size Exclusion Chromatography Profile. A. Coomassie stain, B. silver stain, and C. western blot of Superdex 75 purification loaded with 1. MWM, 2-4. SEC Peak 1 in fractions 4-6, 5. Peak 2 (fraction 12), 6. Peak 2 (fraction 13), 7. Peak 2/3 (fraction 14), 8. Peak 3 (fraction 15), 9. Loading dye, 10. Pooled MonoQ

Size exclusion chromatography, given its reliance on molecular size for separation, was an apt choice to address the persistent issue of coelution involving the 100kDa protein. By employing a Superdex 75 size exclusion column, it was anticipated that Cp3GT could separate from the 100kDa contaminant protein. The observed absence of this 100kDa band in the resulting fractions corroborated the strength of this strategy.

The chromatogram obtained, with its three discernible peaks, points toward a complexity of proteins or protein states in the sample. The detection disparity between Coomassie and silver staining, as well as western blotting, indicates differences in Cp3GT concentration in the elution fractions. Notably, the detection of pure Cp3GT in the SEC Peak 1 is notable. This outcome, combined with the absence of the 100kDa contaminant, underscores the significance of size exclusion chromatography in the purification toolkit.

Cp3GT was successfully expressed in *P. pastoris*, and the procedure was fine-tuned to consistently yield recombinant Cp3GT devoid of degradation. Through meticulous development and assessment, Cp3GT was efficiently purified to apparent homogeneity utilizing a composite chromatographic approach that harnessed the strengths of affinity, anion exchange, and size exclusion chromatography. This methodical purification strategy ensures the reproducible production of intact, homogenous Cp3GT.

Achieving highly pure and intact recombinant Cp3GT is particularly significant in the context of advanced structural analyses. Direct experimental methods, such as X-ray crystallography and cryogenic electron microscopy, demand elevated concentrations of unadulterated protein to render meaningful outcomes. With a consistent source of pure Cp3GT now available, the stage is set for employing these techniques to delve deeper into the protein's structural dynamics.

3.3.6 Cp3GT, Cp3GTΔ10, and Cp3GTΔ80 Structure Generation and Ligand Binding Site Prediction

Recent advancements in protein structural prediction tools that incorporate deep learning and machine learning algorithms offer an avenue for more intricate understanding of protein-ligand interactions. Given this backdrop, it was hypothesized that an in-depth examination of Cp3GT structure and the interactions between Cp3GT residues and its flavonol substrates could be achieved through updated structural prediction models. D-I TASSER was selected for this analysis due to its notable accuracy in CASP15, its integration with the AlphaFold 2 pipeline, and its capability to incorporate ligands and generate potential docking poses using COFACTOR.

The persistent presence of degradation products during the purification process motivated a comparison of the intact Cp3GT, equipped with C-terminal c-myc/6x His recombinant tags and a thrombin-cleavage sequence [31,49], to a truncated variant that was devoid of the initial 80 amino acids on the N-terminus (Cp3GTΔ80) through *in silico* modeling. Additionally, a variant of Cp3GT with a truncation of the first 10 amino acids on the N-terminus (Cp3GTΔ10) was introduced into the analysis due to its demonstrably altered activity (Fig. 3.8). The decision to study the 80 amino acid truncation was driven by an observable size discrepancy of approximately 9kDa, corresponding to about 80 amino acids, as evidenced in Coomassie and western blot analyses (Fig. 3.1-3.3). The rationale behind examining the 10 amino acid truncation stems from the size difference of roughly 1-1.5kDa (Fig. 3.10), or equivalently around 10 amino acids.

In this endeavor, the primary aim was to elucidate new insights into Cp3GT structural characteristics that give rise to activity as well as flavonol specificity, and assess how degradation events might impact Cp3GT structure and activity using the most recent,

technologically advanced *in silico* tools. Additionally, this research aimed to validate the accuracy of these structural models and ligand binding site predictions by comparing them with an anthocyanidin/flavonol specific GT (VvGT1) with 59% homology with Cp3GT [34], cross checking them with previous docking experiments that used Autodock [32,33], and evaluating the models in the context of previous biochemical analyses of Cp3GT with mutated residues.

The highest-ranking models generated by D-I-TASSER were used for analysis, with estimated TM-Scores for Cp3GT, Cp3GTΔ10, and Cp3GTΔ80 being 0.63, 0.64, and 0.61 respectively. These scores were considered acceptable for analysis; however, the inclusion of C-terminal recombinant tags in the modeling resulted in lower scores because the reference PDB crystal structures used as templates did not report such tags. Indeed, when Cp3GT was modeled without tags, the TM-Scores for Cp3GT, Cp3GTΔ10, and Cp3GTΔ80 were 0.65, 0.65, and 0.66 respectively.

The D-I-TASSER pipeline subjected each model to the COFACTOR algorithm and produced the scoring parameters seen in Table 3.1. The analysis of this data provided insights into the binding characteristics of the Cp3GT variants. All three protein models, Cp3GT, Cp3GTΔ10, and Cp3GTΔ80, displayed remarkably consistent C-scores, hovering around the 0.34 mark. This uniformity suggested a comparable level of confidence in the ligand-binding predictions, despite the N-terminal truncations.

Each prediction's foundation on the structure of VvGT1, as evidenced by the PDB analog, was noteworthy. Given VvGT1's established role in flavonoid glucosylation and its solved crystal structure, its selection as a reference underscored its relevance to the current study [34]. Furthermore, the TM-scores for all models surpassed the 0.8 threshold, indicative of a close structural relationship with the VvGT1 reference. Complementing this, the RMSD values were

uniformly proximate to 1.4, suggesting consistent and accurate predicted atomic positions across the board.

The sequence similarity, with IDEN values around 0.6, implied a moderate to high resemblance between the proteins in question and VvGT1. Such congruence might have influenced the server's choice of VvGT1 as a template and impacted the predicted ligand binding sites of the proteins. An equally significant observation was the coverage value. Boasting Fig.s close to 0.87, it was evident that much of the protein structure found alignment with the reference, suggesting strong modeling accuracy.

Turning to the predicted ligands, the focus on quercetin and kaempferol aligned perfectly with the emphasis on acceptor flavonol substrates. Their presence in the predictions further accentuated the functional significance of these molecules for Cp3GT. The BS-Score, a measure reflecting the local similarity between template and predicted binding sites, offered yet another layer of depth. All models showcased scores above 1, suggesting a significant match between the predicted and template binding sites, instilling greater confidence in the binding predictions provided.

Table 3.1 COFACTOR Scoring Parameters for Cp3GT and Two N-Terminal Truncated Versions as Compared to VvGT1

Name	PDB Analog	C Score	TM-Score	RMSD	IDEN	Coverage	BS-Score	Ligand
Cp3GT	2c9zA (VvGT1)	0.34	0.842	1.41	0.596	0.869	1.75	Quercetin
Cp3GTΔ10	2c9zA (VvGT1)	0.34	0.854	1.38	0.599	0.878	1.72	Quercetin
Cp3GTΔ80	2c1zA (VvGT1)	0.33	0.836	1.45	0.598	0.863	1.47	Kaempferol

3.3.7 Structure Differences and Similarities of Cp3GT with Cp3GTΔ10 and Cp3GTΔ80

The comparative analysis between intact Cp3GT and its truncated variants, Cp3GTΔ10 and Cp3GTΔ80, aimed to elucidate structural alterations potentially impacting GT activity. The superimposition of Cp3GT and Cp3GTΔ10 manifested minimal structural deviations, not immediately elucidating the observed decrement in GT activity. While the subtle disparities between these structures did not directly illuminate the diminished GT activity, a conceivable explanation might involve the overlap of Peak A residual in Peak B, warranting further investigation into whether the truncation allows for folding that retains some level of activity, albeit reduced. The detailed discussion on Cp3GTΔ10 will be further elaborated in subsequent sections, maintaining the focus herein on Cp3GTΔ80.

In contrast, the structural comparison between Cp3GT and Cp3GTΔ80 painted a different picture. Clear and significant structural alterations were evident that would have implications for the enzyme's activity. Notably, the absence of the alpha helix within the catalytic cleft of Cp3GTΔ80, containing the critical residues HID-22 and SER-20, underlines the potential for this truncation to result in a totally inactive protein.

While docking residues GLN-87, HID-154, GLU-192, and PHE-203 downstream of the 80 amino acid deletion were conserved in their orientation in both models, the positioning of the c-myc/6x His tag in Cp3GTΔ80 spanned the catalytic cleft was in stark contrast to its peripheral location in Cp3GT (Fig. 3.14). Prior studies have asserted that the presence of recombinant tags on Cp3GT doesn't alter its GT activity or substrate specificity [31]. The peripheral placement of these tags in Cp3GT aligns with this established understanding. However, in the Cp3GTΔ80 model, the inward shift of the recombinant tags might introduce unforeseen steric challenges to substrate docking.

Importantly, the conclusion that can be drawn from this data is twofold: while the positional shift of the recombinant tags in Cp3GTΔ80 might introduce potential obstacles, the more definitive deduction is that Cp3GTΔ80 certainly lacks the ability to catalyze flavonol 3-O glucosylation, primarily due to the absence of HID-22. This amino acid is a universally conserved catalytic residue, crucial for maintaining the functional integrity of GTs.

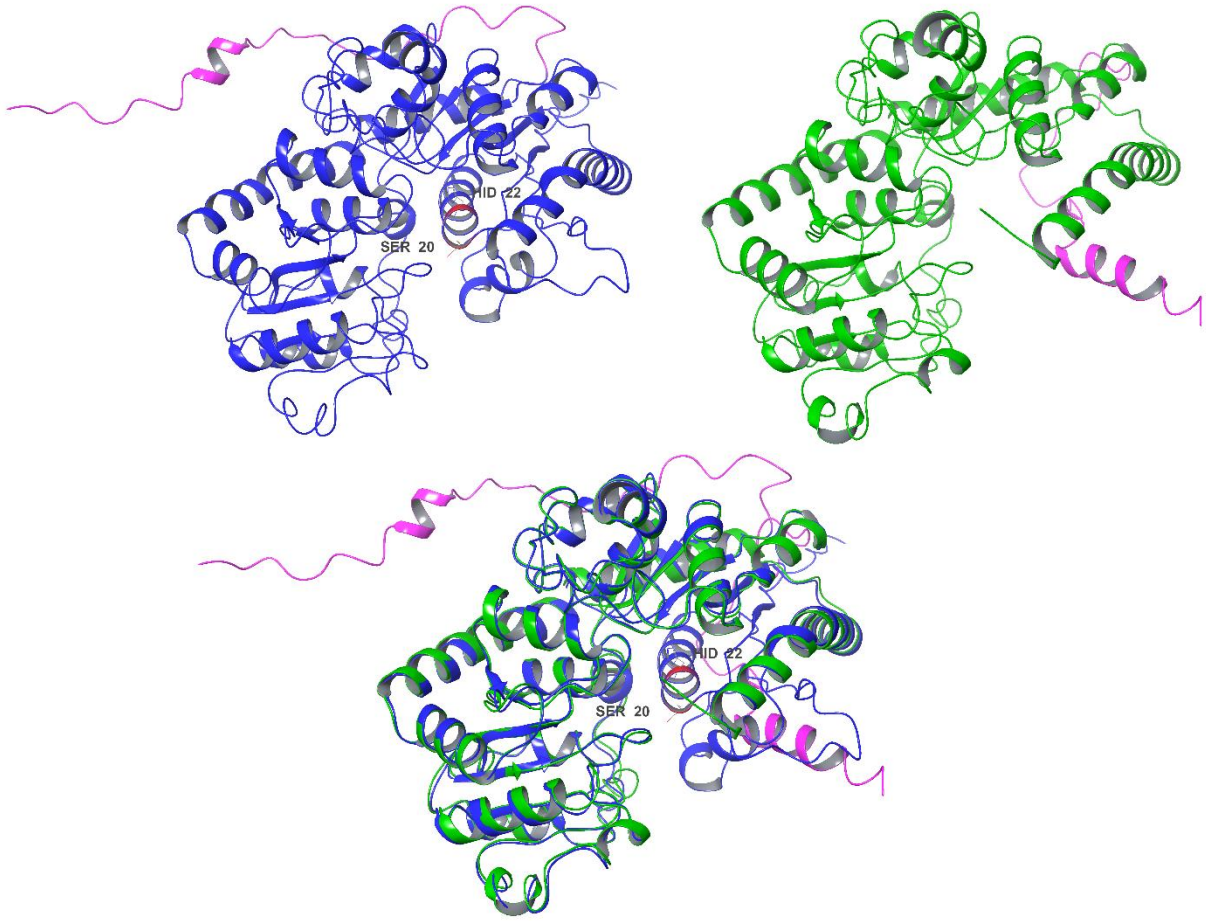

Fig. 3.14 3D Structural Comparison of Cp3GT with an N-Terminal Truncated Variant Cp3GTΔ80. Cp3GT (top left) in blue and Cp3GTΔ80 (top right) in green showing C-myc tag/6xHis in pink. Black arrows point to the alpha helix in the catalytic cleft containing catalytic residues HID-22 and SER-20 (red). The alpha helix is notably absent in Cp3GTΔ80.

Superimposition of Cp3GT with Cp3GTΔ80 (bottom center) highlights structural differences between the two.

3.3.8 Distinct Binding Dynamics in the Cp3GTΔ80 Kaempferol Model

The ligand-binding analysis for Cp3GTΔ80 was conducted using kaempferol as the substrate. This choice was driven by COFACTOR's prediction, which selected kaempferol as the binding template for Cp3GTΔ80, whereas it favored quercetin for Cp3GT and Cp3GTΔ10. While quercetin remains Cp3GT's preferred substrate, the enzyme's capability to glucosylate kaempferol ensures the relevance of this analysis. It's intriguing that an 80 amino acid deletion in Cp3GTΔ80 could alter COFACTOR's substrate prediction.

In the Cp3GTΔ80-kaempferol binding model, several stark differences compared to Cp3GT were observed. The omission of the catalytic residue HIS-22 in Cp3GTΔ80 is notable (Fig. 3.15). Considering the prior evidence from both this study and other literature, it is strongly suggested that the absence of HIS-22 would render Cp3GTΔ80 inactive. The absence of critical docking residues, SER-20, PHE-17, and PHE-19, further compounds this argument (Fig. 3.15). These residues, previously established as essential for ligand binding in Cp3GT, question the functional capacity of Cp3GTΔ80.

However, the model indicates the possibility of substrate docking in Cp3GTΔ80. Phenylalanine residues are predicted to be close enough to the A and C rings of kaempferol to establish pi bonds (Fig. 3.15). Moreover, HID-74, equivalent to HID-154 in Cp3GT, may form a pi bond with the B ring (Fig. 3.15 2D) or a hydrogen bond with the B ring hydroxyl (Fig. 3.15 3D). Coupled with the prediction that GLN-7 forms a hydrogen bond with the 7-OH of kaempferol, analogous to GLN-87 in Cp3GT, this suggests that Cp3GTΔ80 might accommodate

kaempferol docking. Yet, due to the absence of key residues, enzymatic glucosylation is improbable. In conclusion, while the Cp3GTΔ80 model shows potential for substrate docking, critical structural alterations hinder its enzymatic functionality. Direct experimental validation is needed to corroborate these *in silico* predictions.

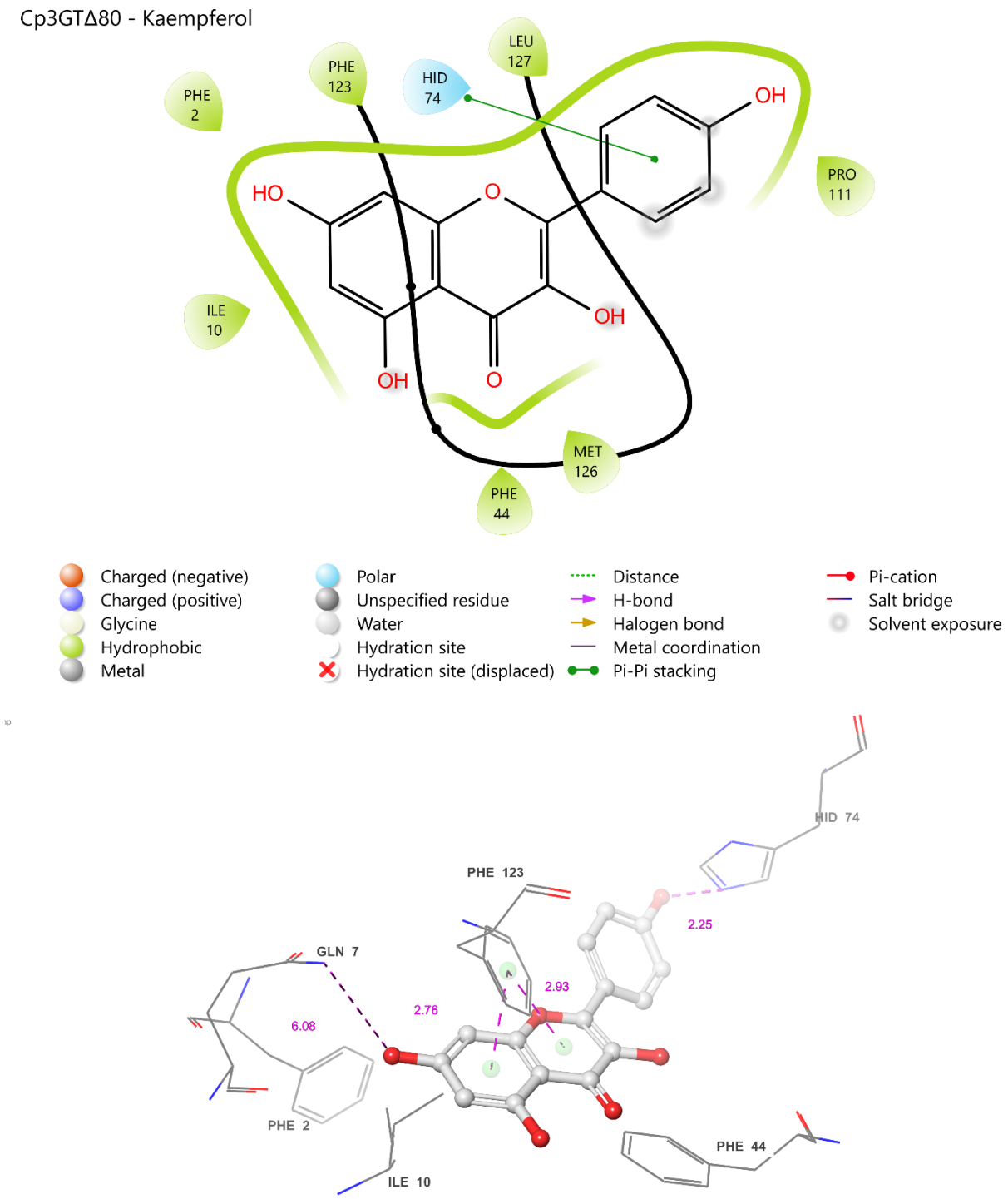

Fig. 3.15 2D and 3D models of Cp3GTΔ80 Predicted Ligand Binding Sites with Kaempferol. All residues shown are positionally equivalent to Cp3GT shown in Fig. 3.17 minus the 80 removed residues.

3.3.9 Establishing the Flavonol Glucosylation Mechanism from VvGT1 as a Comparative Template for Cp3GT Variants

The COFACTOR-generated ligand-binding sites of Cp3GT, Cp3GTΔ10 (see below), and Cp3GTΔ80 in interaction with the flavonols quercetin or kaempferol were extensively examined, and the specifics of their residue-ligand interactions were meticulously analyzed. One key observation was COFACTOR's consistent selection of VvGT1 as the benchmark template for ligand binding with quercetin and kaempferol. Interestingly, flavonols are not the preferred substrate for VvGT1, having previously been shown to preferentially glucosylate the anthocyanidin cyanidin with 100 times higher efficacy than with quercetin [34,60]. However, co-crystallization was only achieved with quercetin and kaempferol, due to inherent challenges in stabilizing anthocyanidins during the crystallization process. This obstacle was later overcome when UGT78K6, an anthocyanidin 3-O GT with 43% sequence identity with Cp3GT, was co-crystallized with petunidin [36]. This divergence between the substrate preferences of VvGT1, UGT78K6, and Cp3GT is of significant interest. Specifically, VvGT1 can accommodate both anthocyanidins and flavonols, UGT78K6 glucosylates anthocyanidins with only trace activity with flavonols, while Cp3GT exclusively glucosylates flavonols [30–33,35,36,49].

Given the availability of crystal structures for VvGT1 co-crystallized with quercetin and kaempferol, combined with VvGT1's 59% identity with Cp3GT (the highest among crystallized GTs), it's logical that VvGT1 with flavonol substrates would serve as the primary template for structural and ligand binding site predictions. As such, understanding the intricacies of VvGT1's interactions with flavonols is paramount to understanding Cp3GT flavonol specificity. The following description is based on crystallographic analysis of VvGT1 crystallized with both

quercetin and kaempferol by Offen et. al 2005 [34]. Mechanistic analysis is the same for both flavonols except where noted.

In some models, histidine residues are designated as 'HIS', which doesn't specify which nitrogen atom carries the hydrogen. In others, they are labeled 'HID', signifying that the hydrogen is on the delta nitrogen (attached to the carbon involved in the double bond) of the imidazole ring. It is anticipated that the nitrogen without the hydrogen would be involved in hydrogen bonding with flavonol hydroxyl groups. Notably, in the VvGT1 model, the placement of the double bond on the HID-20 imidazole ring is inaccurate and couldn't be rectified using existing software tools. For the sake of clarity in predicted Cp3GT binding, the nitrogen atom closest to the hydroxyl group was assumed to be the hydrogen-bonding participant, irrespective of whether it's labeled as HIS or HID. However, considering the variability in ligand prediction models, the *in vivo* conditions under which the enzyme functions, and the imidazole ring's torsional flexibility, it's conceivable that either nitrogen atom could engage in hydrogen bonding with hydroxyl groups in a physiological setting.

The VvGT1 crystal structure indicates that the N-terminal HID-20 deprotonates the kaempferol 3-hydroxyl, priming it for nucleophilic attack by the glucose moiety (Fig. 3.16) [34]. The protonation of HID-20 is stabilized by ASP 119. Concurrently, SER-18 forms a hydrogen bond with the C-ring carbonyl. Complementing this, HID-150 bonds with the B-ring hydroxyl, GLN-84 with the 7-OH group, and pi-bonding interactions occur between PHE-121/372 rings and the B/A ring respectively (Fig.3.16). The torsional freedom of the B ring plays an important role in substrate orientation in the binding pocket and is oriented at an angle of 34.3°. This intricate network establishes the optimal orientation for glucosylation, positioning the catalytic HID-22 approximately 3.83 Å from the 3-OH catalytic site, a crucial arrangement for

deprotonation (Fig. 3.16). When quercetin is the acceptor substrate, the mechanism differs only in that the 3' hydroxyl on the B ring (not present in kaempferol) interacts with the main chain carbonyl of a GLU-189 through a water-mediated hydrogen bond. Additionally, the quercetin B ring is oriented at a 25.2° angle.

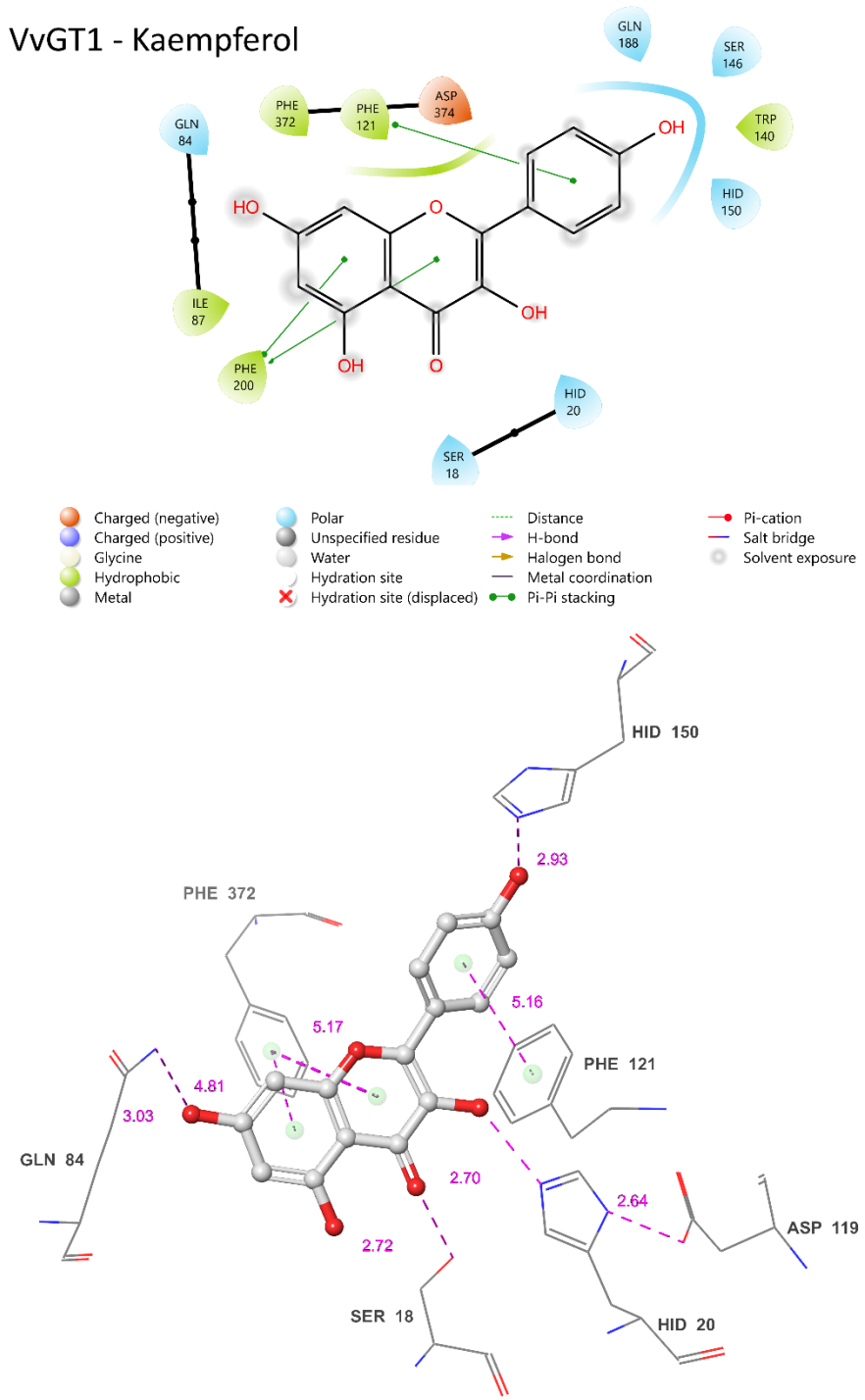

Fig. 3.16 Visual Representation of VvGT1 Residue-Substrate Interaction with Kaempferol. 2D (top) and 3D (bottom) schematic of VvGT1 docked with kaempferol derived from crystal structure (PDB ID: 2c1Z). Distances shown as dashed lines with Å in pink.

3.3.10 Comparative Analysis of Cp3GT and VvGT1 Ligand Binding Dynamics: Insights and Implications

The COFACTOR predicted ligand binding sites for Cp3GT docked with kaempferol underscores several positional parallels with the VvGT1 mechanism. In Cp3GT, HID-22 is positioned 3.01 Å away from the 3-OH catalytic site (Fig. 3.17), a slightly greater distance than the 2.70 Å observed for the analogous residue in VvGT1 (Fig. 3.16). Considering histidine's conserved catalytic function across various GTs, and given its inherent flexibility, it is highly likely that even if the model indicates a greater distance, HID-22 retains a similar functional role as observed in VvGT1.

Although ASP-119 is anticipated to stabilize HID-22, facilitating its deprotonation of the 3-OH site, it's positioned 5.73 Å away (Fig. 3.17), contrasting the closer 2.64 Å distance seen between VvGT1's ASP-119 and HID-20 (Fig. 3.16). While this difference might be due to the limitations or specificities of the modeling approach, it's important to consider that ASP-119 and HID-22 could still interact in a similar manner to that observed in VvGT1. Their actual spatial relationship and functional interactions *in vivo* will require further experimental validation to clarify.

SER-20 is acknowledged as a key residue for flavonol 3-O glucosylation in Cp3GT, a significance reinforced by the observed loss of activity in a mutant where SER-20 was replaced by leucine [61]. However, in the current analysis, its predicted distance of 6.74 Å from the C ring carbonyl (Fig. 3.17) stands in contrast to the 2.72 Å separating SER-18 and the same carbonyl in VvGT1 (Fig. 3.16). This discrepancy implies a lack of hydrogen bond formation in Cp3GT, which contradicts previous docking models that positioned SER-20 much closer to the C ring carbonyl, but notably, did not propose a mechanism for SER-20's interaction with the flavonol

[33]. In VvGT1, SER-18, analogous to SER-20 in Cp3GT, plays a crucial role by hydrogen bonding to the C ring carbonyl of flavonol quercetin, stabilizing its position within the binding pocket. The divergence between these observations and current predictions underlines the challenges and nuances inherent in different modeling approaches and emphasizes the need for experimental validation [33].

Positionally equivalent residues, GLN-87 and HIS-154, in Cp3GT are predicted to engage in hydrogen bonds with the 7-OH and B ring hydroxyl respectively (Fig. 3.17). These predictions align well with interactions observed in VvGT1, underscoring possible evolutionary conserved substrate binding strategies. However, the 6.38 Å distance between Cp3GT's PHE-124 and the A ring (Fig. 3.17) is notable. While this distance suggests that PHE-124 may not be directly involved in pi bonding as in VvGT1, one must consider that subtle differences in protein conformation could bring it closer during the catalytic cycle. Interestingly, although PHE-203 isn't a direct positional counterpart in the sequences, its proximity of 4.38 Å from the A ring hints at a compensatory pi bonding mechanism like VvGT1. Similarly, PHE-17, being 4.42 Å from the C ring, is poised to form a pi bond, mirroring VvGT1's PHE-372 (Fig. 3.17). The observation that both Cp3GT and VvGT1 have an identical 34.3° torsional angle for the B ring suggests the importance of this alignment for substrate interaction.

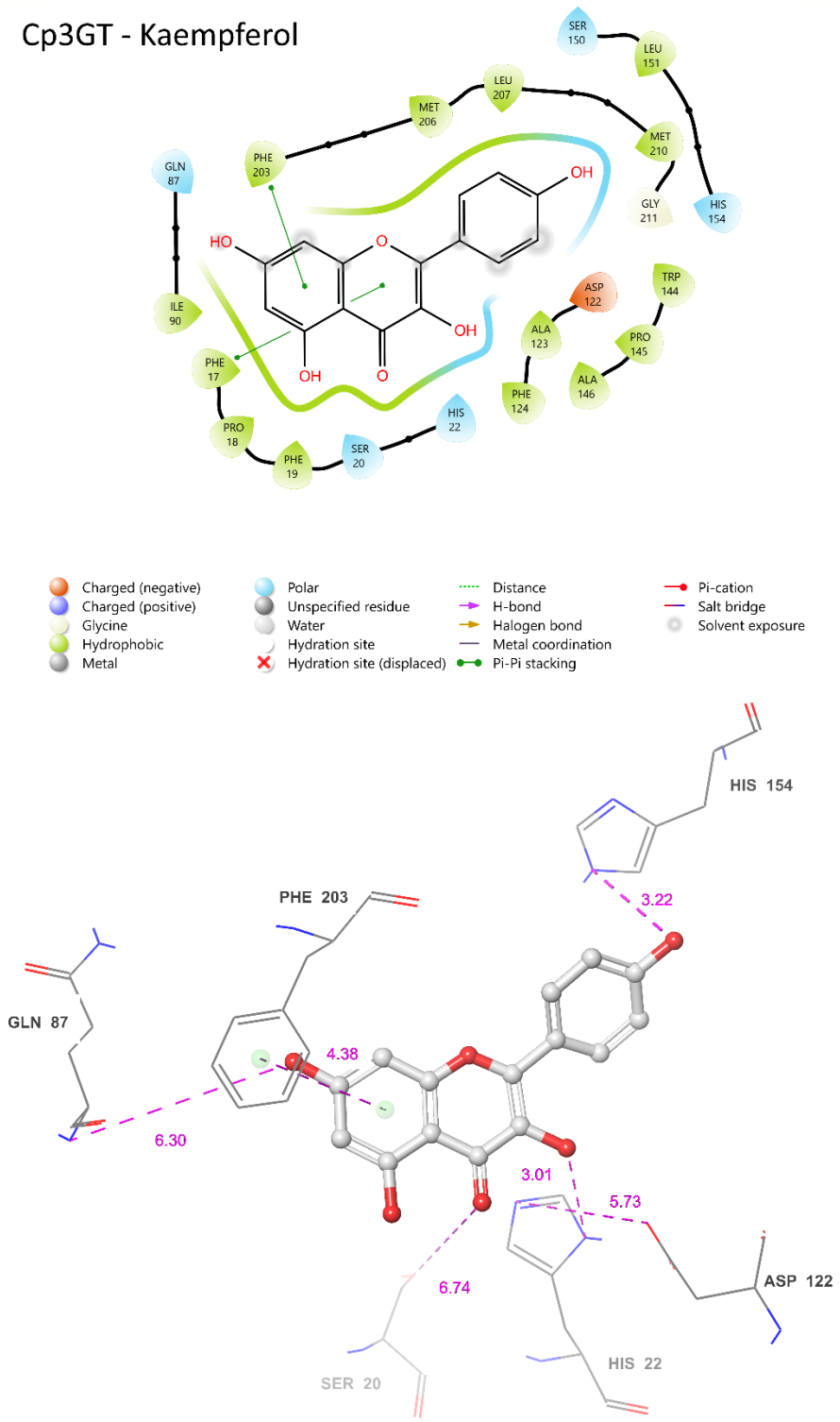

Fig. 3.17 Visual Representation of Cp3GT Residue-Substrate Interaction with Kaempferol. 2D (top) and 3D (bottom) schematic of Cp3GT docked with kaempferol derived from COFACTOR.

3.3.11 Cp3GT's Ligand Interactions with Quercetin and Kaempferol: Conserved Features and Functional Divergences

A ligand binding site prediction was conducted for Cp3GT with quercetin given that it is the preferred substrate for Cp3GT [31–33,49]. This was reinforced by COFACTOR's selection of quercetin as the binding template for both Cp3GT and Cp3GTΔ10. However, for Cp3GTΔ80, COFACTOR identified kaempferol as the binding template. To provide a comprehensive understanding, it was deemed essential to present models of Cp3GT interacting with both substrates. Although quercetin stands as the preferred substrate for Cp3GT, the enzyme readily glucosylates kaempferol [30–33,49].

Common features between Cp3GT when bound to quercetin and its kaempferol-bound state encompass residues such as HID-22, GLN-87, ASP-122, HID-154, and PHE-203 (Fig. 3.18). These residues maintain similar positions and proximity to the substrate across both interactions. Their conservation across different ligand-binding scenarios suggests a vital role in the enzyme's substrate binding or catalytic activities. For instance, HID-22's 3.26 Å distance from the 3-OH bound to quercetin (Fig. 3.18) compared with 3.01 Å distance bound to kaempferol (Fig. 3.17) reinforces the role of histidine as the catalytic Brønsted base, and supports biochemical data showing Cp3GT's ability to glucosylate both flavonols [30–33,49]. It further highlights the conserved nature of this residue's role across diverse glycosyltransferases.

In both models, ASP-122's role in stabilizing HID-22 is evident. The distance between these residues is 4.79Å when Cp3GT is bound to quercetin (Fig. 3.18) and 5.73Å with kaempferol (Fig. 3.17). Given the known function of aspartic acid in supporting the charge of histidine during GT catalysis [16,34,36], it's anticipated that this interaction is expected to remain functionally consistent, irrespective of whether the substrate is kaempferol or quercetin.

Furthermore, the distance between ASP-122 and HID-22 could exhibit variability *in vivo* due to the torsional flexibility of the histidine imidazole ring and the dynamic nature of these residues.

There are distinct variations as well. The PHE-17 pi bond with the C ring was not predicted to occur (Fig. 3.18). The consistent prediction of phenylalanine rings pi bonding with flavonol rings in other models highlights the significance of this interaction in substrate docking (Fig. 3.16 and 3.17, see also [36]. Absence or alteration of these interactions could affect substrate docking and may even play a role in substrate specificity. A disrupted or weakened pi bond interaction could potentially compromise the overall docking stability, suggesting that, while subtle, these interactions might play crucial roles in the enzyme's catalytic efficiency.

Equally important is the observation regarding SER-20. Its location in the current model is even more remote than the already significant 6.74 Å distance seen when Cp3GT is bound to kaempferol (Fig. 3.17). This increasingly distant positioning renders a direct interaction between SER-20 and the C ring carbonyl highly improbable (Fig. 3.18). This raises interesting questions about the flexibility and movement of the enzyme during the catalysis process. Given the evidence from biochemical studies showing a drastic loss of Cp3GT activity when SER-20 was mutated to leucine [61], the functional implications of such positional discrepancies between the modeled and actual states is of great interest. The contradiction between the current observation and previous docking models, which posited SER-20 in a much closer proximity to the flavonol [33], underlines the inherent challenges of *in silico* predictions. It serves as a pertinent reminder of the need to correlate computational models with empirical evidence and, where discrepancies arise, to further refine modeling approaches.

A potential pi bond interaction between TRP-144 and the B ring is observed (Fig. 3.18). The presence of such an interaction brings to light the versatile nature of tryptophan residues in

protein-ligand interactions. Tryptophan, with its unique indole ring, often plays pivotal roles in protein function, be it in ligand binding or in facilitating protein conformational changes. The observed interaction with TRP-144 might provide additional stabilization to the flavonoid within the active site, potentially ensuring efficient catalysis.

3.3.12 Structural Conservation and Divergences in Cp3GTΔ10: Implications for Binding Affinity and Enzymatic Activity

Models of Cp3GTΔ10 exhibited minimal deviation from Cp3GT in terms of catalytic and docking residues (Fig. 3.19). This observation suggests that the deletion in Cp3GTΔ10 might not significantly alter the core structure of the protein, preserving its substrate-binding pocket intact. This structural conservation, even after deletions or mutations, emphasizes the evolutionary resilience of enzymatic architectures, designed to maintain functionality over various alterations.

However, a notable divergence was observed concerning the aromatic residues PHE-17 and PHE-19. In the Cp3GT model, both residues were posited to be proximal to the A and C rings, with PHE-17 even speculated to form a pi bond with the C ring in the kaempferol-bound state (Fig. 3.16). Conversely, in Cp3GTΔ10, neither PHE-17 nor PHE-19 are sufficiently close to establish such interactions with the A and C rings (Fig. 3.19). This potential lack of pi-pi interactions in Cp3GTΔ10 could influence binding affinity and GT activity. While individual pi-pi interactions may not be particularly strong, cumulatively they play a pivotal role in stabilizing protein-ligand complexes.

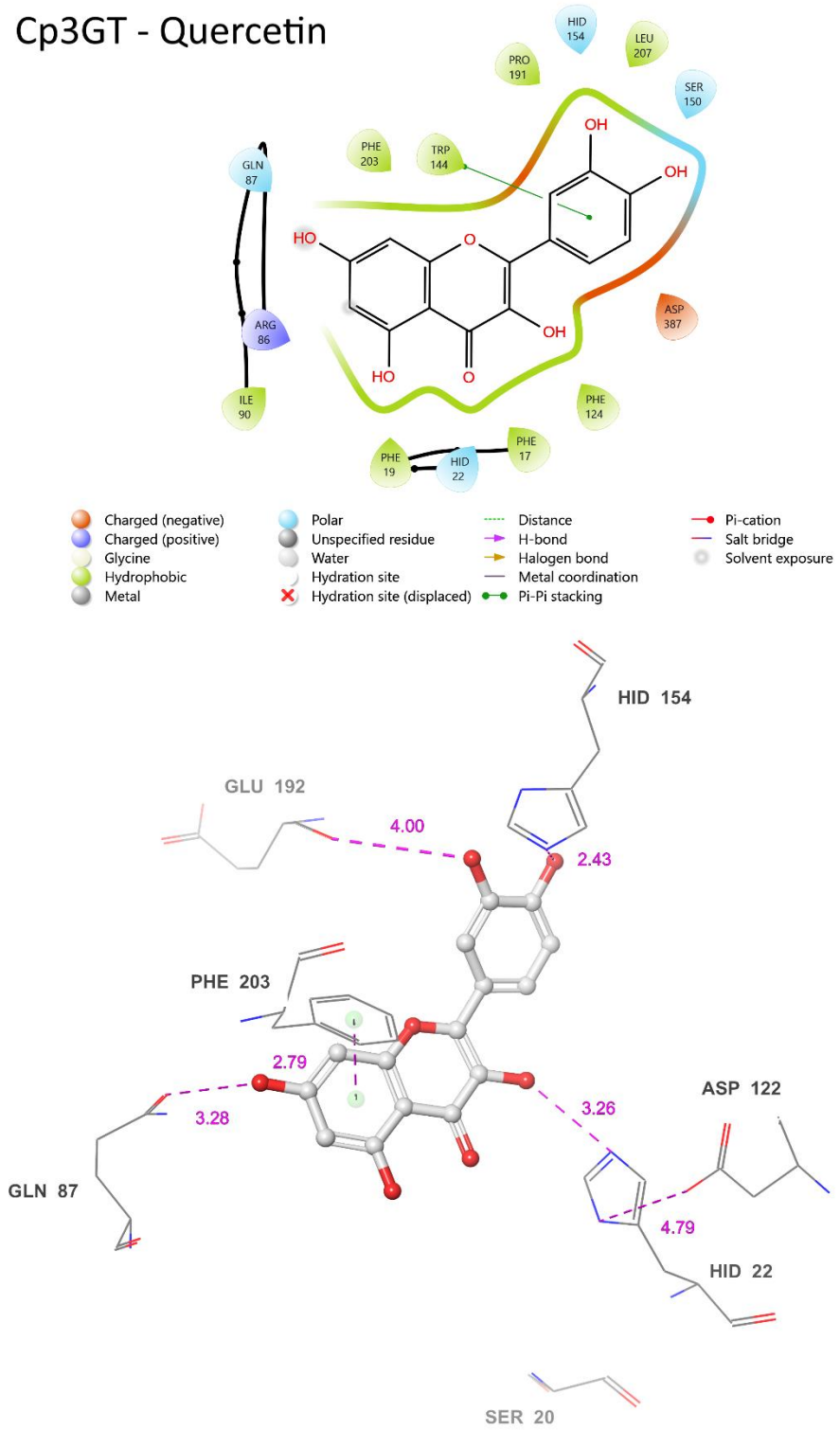

Fig. 3.18 Visual Representation of Cp3GT Residue-Substrate Interaction with Quercetin. 2D (top) and 3D (bottom) schematic of Cp3GT docked with quercetin derived from COFACTOR.

Interestingly, experimental data showed reduced GT activity for Cp3GTΔ10 (Fig. 3.8). Altered folding within the catalytic cleft could attenuate the enzyme's activity by impacting substrate binding and/or catalytic efficiency. It is also plausible that carry-over between peaks during the MonoQ chromatographic purification step resulted in sample heterogeneity, complicating the accurate assessment of its enzymatic efficiency. Regardless, the reduced activity in conjunction with the observed structural differences raises pertinent questions about Cp3GTΔ10's enzymatic behavior. Future research endeavors could focus on comprehensive enzyme kinetics studies to decipher the mechanistic underpinnings of Cp3GTΔ10's altered activity. This may involve employing steady-state kinetics to precisely determine key enzymatic parameters such as the Michaelis constant (K_m), maximum reaction velocity (V_{max}), and catalytic efficiency, which will elucidate the enzyme's affinity for its substrate and its maximal catalytic capability under defined conditions. Furthermore, assessing initial velocity (V_0) under various substrate concentrations will provide insights into the enzyme's activity and potential inhibition under initial reaction conditions, thereby providing a comprehensive understanding of the catalytic discrepancies observed in Cp3GTΔ10.

3.3.13 Deciphering Cp3GT's Flavonol Specificity: Insights and Future Perspectives

In conclusion, this study provides a detailed *in silico* exploration of Cp3GT's ligand-binding interactions with flavonols quercetin and kaempferol and extends this analysis to its truncated versions, Cp3GTΔ10 and Cp3GTΔ80. These evaluations are set against the backdrop of VvGT1, an anthocyanidin/flavonol GT with a resolved crystal structure and a well-established mechanism for flavonol glucosylation. Additionally, they are contextualized within the realm of existing biochemical data that elucidate the pivotal roles of Cp3GT's functional catalytic and docking residues and are complemented by insights derived from recent docking analyses.

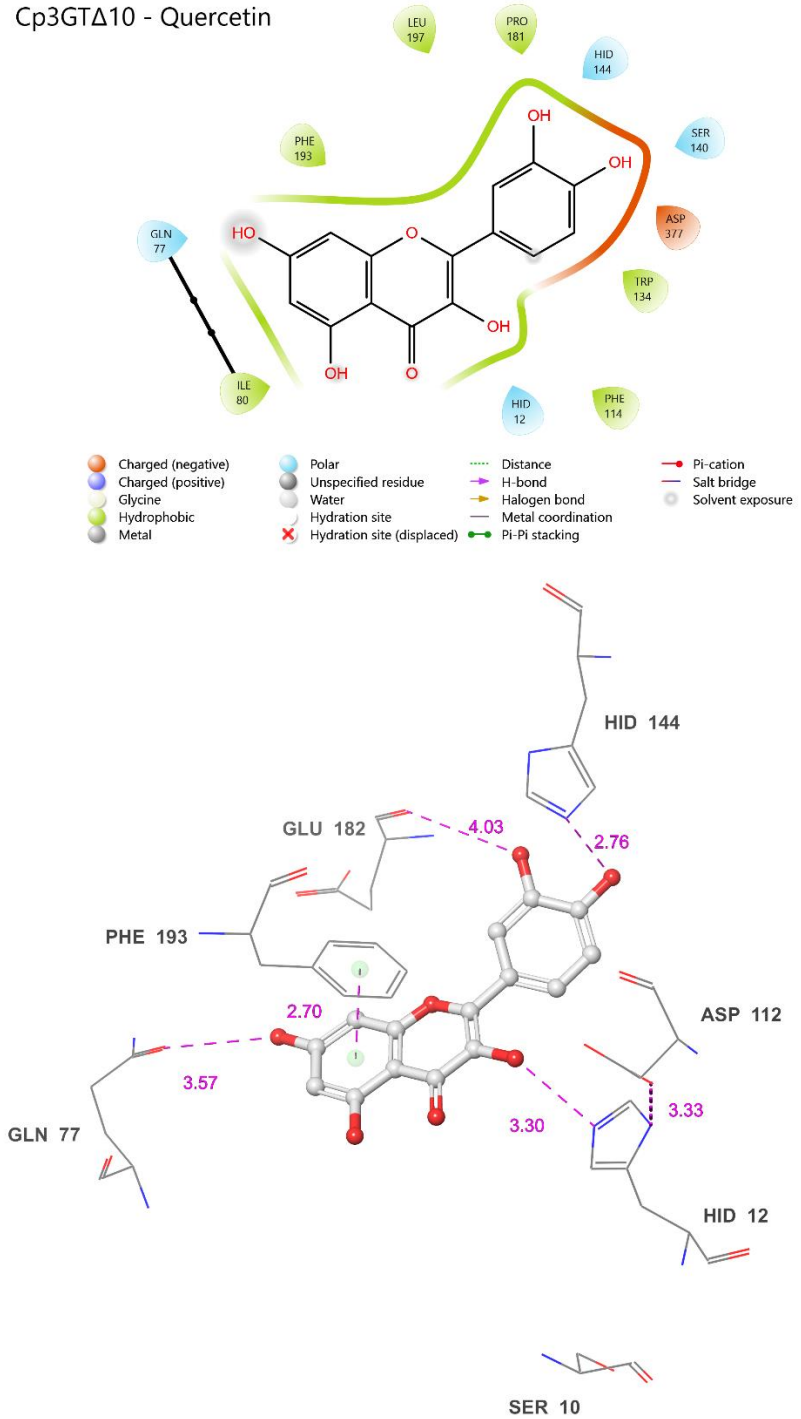

Fig. 3.19 Visual Representation of Cp3GTΔ10 Residue-Substrate Interaction with Quercetin. 2D (top) and 3D (bottom) schematic of Cp3GTΔ10 docked with quercetin derived from

COFACTOR. All residues shown are positionally equivalent to Cp3GT shown in Fig. 3.18 minus the 10 removed residues.

The integration of modern structural modeling tools, specifically D-I-TASSER, with rapid ligand binding predictors like COFACTOR, presents an invaluable avenue for probing enzyme-ligand interactions. While COFACTOR does not use rigorous docking algorithms to determine the most energetically favorable residue-ligand interaction, its predictions aligned quite harmoniously with data procured from crystal structures co-crystallized with substrates and a previous Cp3GT binding analysis that used AutoDock. Despite minor discrepancies, the synergistic use of D-I-TASSER and COFACTOR offers a promising avenue to deepen the understanding of Cp3GT's operational nuances.

Yet, a significant challenge persists – the lack of a resolved crystal structure for a GT with an intrinsic flavonol specificity. This gap renders a comprehensive understanding of Cp3GT's unique ability to selectively glucosylate flavonols at the 3-OH position elusive. Nevertheless, the meticulous expression and purification of Cp3GT showcased in this research lays the foundation to pursue structural evaluation via x-ray crystallography in the future. Achieving a crystal structure of Cp3GT co-crystallized with a flavonol substrate will undeniably decipher the mystery behind flavonol 3-O specificity. Such a revelation would not only serve as a comparative benchmark against extant *in silico* evaluations but also refine the confidence in prevailing structure and ligand prediction methodologies.

The ramifications of this prospective discovery extend future research possibilities. For example a more detailed understanding of flavonol 3-O specificity could galvanize enzyme engineering pursuits, enabling directed modification of these biological entities for specialized

tasks. The gravity and potential ripple effects of such an endeavor, both in the realms of science and practical application, underscore its importance.

3.4 Supplemental Materials

Fig. 3.20: Linear Schematic of Cp3GT Protein Construct. The 468 amico acid sequence of Cp3GT is shown in orange. Thrombin cleavage sequence is shown in red downstream of Cp3GT and separated by a small spacer sequence shown in blue. C-myc and 6x His tag are shown in green separated by another small spacer sequence shown in blue.

3.5 Author Contributions

Both authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

3.6 Funding

This research was funded by an East Tennessee State University Research Development Committee Major Grant awarded to CAM and ETSU School of Graduate and Continuing Studies Grant to ASB.

3.7 Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kröner, A.; Marnet, N.; Andrivon, D.; Val, F. Nicotiflorin, Rutin and Chlorogenic Acid: Phenylpropanoids Involved Differently in Quantitative Resistance of Potato Tubers to Biotrophic and Necrotrophic Pathogens. *Plant Physiology and Biochemistry* **2012**, *57*, 23–31, doi:10.1016/J.PLAPHY.2012.05.006.

2. Hijaz, F.M.; Manthey, J.A.; Folimonova, S.Y.; Davis, C.L.; Jones, S.E.; Reyes-De-Corcuera, J.I. An HPLC-MS Characterization of the Changes in Sweet Orange Leaf Metabolite Profile Following Infection by the Bacterial Pathogen Candidatus *Liberibacter Asiaticus* . *PLoS One* **2013**, *8*, doi:10.1371/JOURNAL.PONE.0079485.

3. Tattini, M.; Gravano, E.; Pinelli, P.; Mulinacci, N.; Romani, A. Flavonoids Accumulate in Leaves and Glandular Trichomes of *Phillyrea Latifolia* Exposed to Excess Solar Radiation. *New Phytol* **2000**, *148*, 69–77, doi:10.1046/j.1469-8137.2000.00743.x.

4. Simlat, M.; Stobiecki, M.; Szklarczyk, M. Accumulation of Selected Phenolics and Expression of PAL Genes in Carrots Differing in Their Susceptibility to Carrot Fly (*Psila Rosae* F.). *Euphytica* **2013**, *190*, 253–266, doi:10.1007/S10681-012-0836-4.

5. Fini, A.; Guidi, L.; Ferrini, F.; Brunetti, C.; Di Ferdinando, M.; Biricolti, S.; Pollastri, S.; Calamai, L.; Tattini, M. Drought Stress Has Contrasting Effects on Antioxidant Enzymes Activity and Phenylpropanoid Biosynthesis in *Fraxinus Ornus* Leaves: An Excess Light Stress Affair? *J Plant Physiol* **2012**, *169*, 929–939, doi:10.1016/J.JPLPH.2012.02.014.

6. Yin, R.; Han, K.; Heller, W.; Albert, A.; Dobrev, P.I.; Zažímalová, E.; Schäffner, A.R. Kaempferol 3-O-rhamnoside-7-O-rhamnoside Is an Endogenous Flavonol Inhibitor of Polar

Auxin Transport in *Arabidopsis* Shoots. *New Phytologist* **2013**, *201*, 466–475, doi:10.1111/nph.12558.

7. Li, J.; Ou-Lee, T.M.; Raba, R.; Amundson, R.G.; Last, R.L. *Arabidopsis* Flavonoid Mutants Are Hypersensitive to UV-B Irradiation. *Plant Cell* **1993**, *5*, 171–179, doi:10.1105/TPC.5.2.171.

8. Abad-García, B.; Garmón-Lobato, S.; Sánchez-Ilárduya, M.B.; Berrueta, L.A.; Gallo, B.; Vicente, F.; Alonso-Salces, R.M. Polyphenolic Contents in Citrus Fruit Juices: Authenticity Assessment. *European Food Research and Technology* **2014**, *238*, 803–818, doi:10.1007/S00217-014-2160-9/METRICS.

9. Denish, P.R.; Fenger, J.A.; Powers, R.; Sigurdson, G.T.; Grisanti, L.; Guggenheim, K.G.; Laporte, S.; Li, J.; Kondo, T.; Magistrato, A.; et al. Discovery of a Natural Cyan Blue: A Unique Food-Sourced Anthocyanin Could Replace Synthetic Brilliant Blue. *Sci Adv* **2021**, *7*, doi:10.1126/SCIADV.ABE7871/SUPPL_FILE/ABE7871_SM.PDF.

10. Bashllari, R.; Molonia, M.S.; Muscarà, C.; Speciale, A.; Wilde, P.J.; Saija, A.; Cimino, F. Cyanidin-3-O-Glucoside Protects Intestinal Epithelial Cells from Palmitate-Induced Lipotoxicity. *Arch Physiol Biochem* **2023**, *129*, 379–386, doi:10.1080/13813455.2020.1828480.

11. Sukprasansap, M.; Chanvorachote, P.; Tencomnao, T. Cyanidin-3-Glucoside Activates Nrf2-Antioxidant Response Element and Protects against Glutamate-Induced Oxidative and Endoplasmic Reticulum Stress in HT22 Hippocampal Neuronal Cells. *BMC Complement Med Ther* **2020**, *20*, 1–12, doi:https://doi.org/10.1186/s12906-020-2819-7.

12. Devika, M.S.; Vysakh, A.; Vijeesh, V.; Jisha, N.; Ruby, P.M. Flavonoid Glycoside Diosmin Induces Apoptosis and Cell Cycle Arrest in DLD-1 Human Colon Cancer Cell Line.

Journal of Biologically Active Products from Nature **2022**, *12*, 232–242, doi:10.1080/22311866.2022.2081608.

13. Huang, J.; Zhou, Y.; Zhong, X.; Su, F.; Xu, L. Effects of Vitexin, a Natural Flavonoid Glycoside, on the Proliferation, Invasion, and Apoptosis of Human U251 Glioblastoma Cells. *Oxid Med Cell Longev* **2022**, *2022*, doi:10.1155/2022/3129155.

14. Meng, N.; Chen, K.; Wang, Y.; Hou, J.; Chu, W.; Xie, S.; Yang, F.; Sun, C. Dihydrohomoplantagin and Homoplantaginin, Major Flavonoid Glycosides from *Salvia Plebeia* R. Br. Inhibit OxLDL-Induced Endothelial Cell Injury and Restrict Atherosclerosis via Activating Nrf2 Anti-Oxidation Signal Pathway. *Molecules* **2022**, *27*, 1990, doi:10.3390/MOLECULES27061990/S1.

15. Yu, L.; Zhang, S.D.; Zhao, X.L.; Ni, H.Y.; Song, X.R.; Wang, W.; Yao, L.P.; Zhao, X.H.; Fu, Y.J. Cyanidin-3-Glucoside Protects Liver from Oxidative Damage through AMPK/Nrf2 Mediated Signaling Pathway *in Vivo* and *in Vitro*. *J Funct Foods* **2020**, *73*, 104148, doi:10.1016/J.JFF.2020.104148.

16. Lairson, L.; Henrissat, B.; Davies, G.; Withers, S. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu Rev Biochem* **2008**, *77*, 521–555, doi:10.1146/annurev.biochem.76.061005.092322.

17. Zhang, M.; Yi, Y.; Gao, B.H.; Su, H.F.; Bao, Y.O.; Shi, X.M.; Wang, H.D.; Li, F.D.; Ye, M.; Qiao, X. Functional Characterization and Protein Engineering of a Triterpene 3-/6-/2′-O-Glycosyltransferase Reveal a Conserved Residue Critical for the Regiospecificity. *Angewandte Chemie International Edition* **2022**, *61*, e202113587, doi:10.1002/ANIE.202113587.

18. Kubo, A.; Arai, Y.; Nagashima, S.; Yoshikawa, T. Alteration of Sugar Donor Specificities of Plant Glycosyltransferases by a Single Point Mutation. *Arch Biochem Biophys* **2004**, *429*, 198–203, doi:10.1016/J.ABB.2004.06.021.

19. Nguyen, H.M.; Putterill, J.; Dare, A.P.; Plunkett, B.J.; Cooney, J.; Peng, Y.; Souleyre, E.J.F.; Albert, N.W.; Espley, R. V.; Günther, C.S. Two Genes, ANS and UFGT2, from *Vaccinium* Spp. Are Key Steps for Modulating Anthocyanin Production. *Front Plant Sci* **2023**, *14*, 1082246, doi:https://doi.org/10.3389/fpls.2023.1082246.

20. Nidetzky, B.; Gutmann, A.; Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *ACS Catal* **2018**, *8*, 6283–6300, doi:https://doi.org/10.1021/acscatal.8b00710.

21. Huang, Y.; Li, D.; Wang, J.; Cai, Y.; Dai, Z.; Jiang, D.; Liu, C. GuUGT, a Glycosyltransferase from *Glycyrrhiza Uralensis*, Exhibits Glycyrrhetinic Acid 3- and 30-O-Glycosylation. *R Soc Open Sci* **2019**, *6*, doi:10.1098/RSOS.191121.

22. Wei, W.; Wang, P.; Wei, Y.; Liu, Q.; Yang, C.; Zhao, G.; Yue, J.; Yan, X.; Zhou, Z. Characterization of *Panax Ginseng* UDP-Glycosyltransferases Catalyzing Protopanaxatriol and Biosyntheses of Bioactive Ginsenosides F1 and Rh1 in Metabolically Engineered Yeasts. *Mol Plant* **2015**, *8*, 1412–1424, doi:10.1016/J.MOLP.2015.05.010.

23. Grimsey, E.; Collis, D.W.P.; Mikut, R.; Hilpert, K. The Effect of Lipidation and Glycosylation on Short Cationic Antimicrobial Peptides. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2020**, *1862*, 183195, doi:10.1016/J.BBAMEM.2020.183195.

24. Song, Q.; Wu, Z.; Fan, Y.; Song, W.; Zhang, P.; Wang, L.; Wang, F.; Xu, Y.; Wang, P.G.; Cheng, J. Production of Homogeneous Glycoprotein with Multisite Modifications by an

Engineered N-Glycosyltransferase Mutant. *Journal of Biological Chemistry* **2017**, *292*, 8856–8863, doi:10.1074/jbc.M117.777383.

25. Ooi, K.E.; Zhang, X.W.; Kuo, C.Y.; Liu, Y.J.; Yu, C.C. Chemoenzymatic Synthesis of Asymmetrically Branched Human Milk Oligosaccharide Lacto-N-Hexaose. *Front Chem* **2022**, *10*, doi:10.3389/fchem.2022.905105.

26. Li, R.; Yu, H.; Muthana, S.M.; Freedberg, D.I.; Chen, X. Size-Controlled Chemoenzymatic Synthesis of Homogeneous Oligosaccharides of *Neisseria Meningitidis* W Capsular Polysaccharide. *ACS Catal* **2020**, *10*, 2791–2798, doi:10.1021/ACSCATAL.9B05597/ASSET/IMAGES/LARGE/CS9B05597_0001.JPEG.

27. Wang, Z.; Dordick, J.S.; Linhardt, R.J. *Escherichia Coli* K5 Heparosan Fermentation and Improvement by Genetic Engineering. *Bioeng Bugs* **2011**, *2*, 63, doi:10.4161/BBUG.2.1.14201.

28. Chang, T.S.; Wang, T.Y.; Chiang, C.M.; Lin, Y.J.; Chen, H.L.; Wu, Y.W.; Ting, H.J.; Wu, J.Y. Biotransformation of Celastrol to a Novel, Well-Soluble, Low-Toxic and Anti-Oxidative Celastrol-29-O-β-Glucoside by *Bacillus* Glycosyltransferases. *J Biosci Bioeng* **2021**, *131*, 176–182, doi:10.1016/J.JBIOSC.2020.09.017.

29. Xu, Y.; Masuko, S.; Takieddin, M.; Xu, H.; Liu, R.; Jing, J.; Mousa, S.A.; Linhardt, R.J.; Liu, J. Chemoenzymatic Synthesis of Homogeneous Ultralow Molecular Weight Heparins. *Science* **2011**, *334*, 498–501, doi:10.1126/SCIENCE.1207478.

30. Owens, D.K.; McIntosh, C.A. Identification, Recombinant Expression, and Biochemical Characterization of a Flavonol 3-O-Glucosyltransferase Clone from *Citrus Paradisi*. *Phytochemistry* **2009**, *70*, 1382–1391, doi:10.1016/j.phytochem.2009.07.027.

31. Birchfield, A.S.; McIntosh, C.A. The Effect of Recombinant Tags on *Citrus Paradisi* Flavonol-Specific 3-O Glucosyltransferase Activity. *Plants* **2020**, *9*, 402, doi:10.3390/plants9030402.

32. 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.; et al. Identification, Recombinant Expression, and Biochemical Analysis of Putative Secondary Product Glucosyltransferases from *Citrus Paradisi*. *J Agric Food Chem* **2016**, *64*, 1957–1969, doi:10.1021/acs.jafc.5b05430.

33. Devaiah, S.P.; Tolliver, B.M.; Zhang, C.; Owens, D.K.; McIntosh, C.A. Mutational Analysis of Substrate Specificity in a *Citrus Paradisi* Flavonol 3-O-Glucosyltransferase. *J Plant Biochem Biotechnol* **2018**, *27*, 13–27, doi:10.1007/s13562-017-0411-0.

34. 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. Structure of a Flavonoid Glucosyltransferase Reveals the Basis for Plant Natural Product Modification. *EMBO Journal* **2006**, *25*, 1396–1405, doi:10.1038/sj.emboj.7600970.

35. 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*. *J Synchrotron Radiat* **2013**, *20*, 894–898, doi:10.1107/S0909049513020712.

36. 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-Glucosyltransferase from *Clitoria Ternatea*. *Protein Science* **2015**, *24*, 395–407, doi:10.1002/pro.2630.

37. Zheng, W.; Wuyun, Q.; Freddolino, P.L.; Zhang, Y. Integrating Deep Learning, Threading Alignments, and a Multi-MSA Strategy for High-Quality Protein Monomer and Complex Structure Prediction in CASP15. *Proteins* **2023**, doi:10.1002/prot.26585.

38. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589, doi:10.1038/s41586-021-03819-2.

39. Puxbaum, V.; Mattanovich, D.; Gasser, B. Quo Vadis? The Challenges of Recombinant Protein Folding and Secretion in *Pichia Pastoris*. *Appl Microbiol Biotechnol* **2015**, *99*, 2925–2938, doi:10.1007/S00253-015-6470-Z/METRICS.

40. Robert, F.; Bierau, H.; Rossi, M.; Agugiaro, D.; Soranzo, T.; Broly, H.; Mitchell-Logean, C. Degradation of an Fc-Fusion Recombinant Protein by Host Cell Proteases: Identification of a CHO Cathepsin D Protease. *Biotechnol Bioeng* **2009**, *104*, 1132–1141, doi:10.1002/BIT.22494.

41. Maeda, H.; Chatani, E.; Koyama, T.; Sugiura, M.; Izumi, H.; Hayashi, R. Indiscriminate Glycosylation of Procarboxypeptidase Y Expressed in *Pichia Pastoris*. *Carbohydr Res* **2004**, *339*, 1041–1045, doi:10.1016/j.carres.2004.02.003.

42. Sorensen, S.O.; Van Den Hazel, H.B.; Kielland-Brandt, M.C.; Winther, J.R. PH-Dependent Processing of Yeast Procarboxypeptidase Y by Proteinase A *in Vivo* and *in Vitro*. *Eur J Biochem* **1994**, *220*, 19–27, doi:10.1111/J.1432-1033.1994.TB18594.X.

43. Zhang, Y.; Liu, R.; Wu, X. The Proteolytic Systems and Heterologous Proteins Degradation in the Methylotrophic Yeast *Pichia Pastoris*. *Ann Microbiol* **2007**, *57*, 553–560, doi:10.1007/BF03175354.

44. Karbalaei, M.; Rezaee, S.A.; Farsiani, H. *Pichia Pastoris*: A Highly Successful Expression System for Optimal Synthesis of Heterologous Proteins. *J Cell Physiol* **2020**, *235*, 5867–5881, doi:10.1002/JCP.29583.

45. Jahic, M.; Gustavsson, M.; Jansen, A.K.; Martinelle, M.; Enfors, S.O. Analysis and Control of Proteolysis of a Fusion Protein in *Pichia Pastoris* Fed-Batch Processes. *J Biotechnol* **2003**, *102*, 45–53, doi:10.1016/S0168-1656(03)00003-8.

46. Wu, M.; Shen, Q.; Yang, Y.; Zhang, S.; Qu, W.; Chen, J.; Sun, H.; Chen, S. Disruption of YPS1 and PEP4 Genes Reduces Proteolytic Degradation of Secreted HSA/PTH in *Pichia Pastoris* GS115. *J Ind Microbiol Biotechnol* **2013**, *40*, 589–599, doi:10.1007/S10295-013-1264-8.

47. Sakai, Y.; Koller, A.; Rangell, L.K.; Keller, G.A.; Subramani, S. Peroxisome Degradation by Microautophagy in *Pichia Pastoris*: Identification of Specific Steps and Morphological Intermediates. *Journal of Cell Biology* **1998**, *141*, 625–636, doi:10.1083/JCB.141.3.625.

48. Yang, D.; Peng, M.; Yang, H.; Yang, Q.; Xu, J. Expression, Purification and Characterization of *Gloydius Shedaoensis* Venom Gloshedobin as Hsp70 Fusion Protein in *Pichia Pastoris*. *Protein Expr Purif* **2009**, *66*, 138–142, doi:10.1016/J.PEP.2009.03.003.

49. Birchfield, A. Preparation of a Flavonol Specific Glucosyltransferase Found in Grapefruit and Site-Directed Mutants for Protein Crystallization, East Tennessee State University: Johnson City, 2019.

50. McIntosh, C.A.; Latchinian, L.; Mansell, R.L. Flavanone-Specific 7-O-Glucosyltransferase Activity in *Citrus Paradisi* Seedlings: Purification and Characterization. *Arch Biochem Biophys* **1990**, *282*, 50–57, doi:10.1016/0003-9861(90)90085-d.

51. McIntosh, C.A.; Mansell, R.L. Biosynthesis of Naringin in *Citrus Paradisi*: UDP-Glucosyl-Transferase Activity in Grapefruit Seedlings. *Phytochemistry* **1990**, *29*, 1533–1538, doi:10.1016/0031-9422(90)80115-W.

52. Zhang, C.; Zhang, X.; Freddolino, P.L.; Zhang, Y. BioLiP2: An Updated Structure Database for Biologically Relevant Ligand-Protein Interactions. *Nucleic Acids Res* **2023**, doi:10.1093/nar/gkad630.

53. Yang, J.; Roy, A.; Zhang, Y. BioLiP: A Semi-Manually Curated Database for Biologically Relevant Ligand-Protein Interactions. *Nucleic Acids Res* **2013**, *41*, D1096-103, doi:10.1093/nar/gks966.

54. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera--a Visualization System for Exploratory Research and Analysis. *J Comput Chem* **2004**, *25*, 1605–1612, doi:10.1002/jcc.20084.

55. Dassault Systèmes *BIOVIA Discovery Studio V21.2.0.20298*; Dassault Systèmes: San Diego, 2021;

56. Schrödinger Release 2023-3 *Maestro*; Schrödinger, LLC: New York, 2023;

57. Shao, H.; He, X.; Achnine, L.; Blount, J.W.; Dixon, R.A.; Wang, X. Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago Truncatula*. *Plant Cell* **2005**, *17*, 3141–3154, doi:10.1105/tpc.105.035055.

58. Sue, M.; Ishihara, A.; Iwamura, H. Occurrence and Characterization of a UDP-Glucose:Hydroxamic Acid Glucosyltransferase Isolated from Wheat (*Triticum Aestivum*) Seedlings. *Z Naturforsch C J Biosci* **2000**, *55*, 701–707, doi:10.1515/znc-2000-9-1006.

59. Nagatomo, Y.; Usui, S.; Ito, T.; Kato, A.; Shimosaka, M.; Taguchi, G. Purification, Molecular Cloning and Functional Characterization of Flavonoid *C* -Glucosyltransferases from *Fagopyrum Esculentum* M. (Buckwheat) Cotyledon. *The Plant Journal* **2014**, *80*, 437–448, doi:10.1111/tpj.12645.

60. Ford, C.M.; Boss, P.K.; Hæj, P.B. Cloning and Characterization of *Vitis Vinifera* UDP-Glucose. Flavonoid 3- O-Glucosyltransferase, a Homologue of the Enzyme Encoded by the Maize Bronze- 1 Locus That May Primarily Serve to Glucosylate Anthocyanidins in Vivo. *Journal of Biological Chemistry* **1998**, *273*, 9224–9233, doi:10.1074/jbc.273.15.9224.

61. Adepoju, O. Using Site-Directed Mutagenesis to Determine Impact of Amino Acid Substitution on Substrate and Regiospecificity of Grapefruit Flavonol Specific 3-O-Glucosyltransferase, East Tennessee State University: Johnson City, 2014.

CHAPTER 4. METABOLIC ENGINEERING AND SYNTHETIC BIOLOGY OF PLANT NATURAL PRODUCTS – A MINIREVIEW

Aaron S. Birchfield and Cecilia A. McIntosh*

Department of Biological Sciences, P.O. Box 70703, East Tennessee State University, Johnson City, TN 37614, USA; birchfieldas@etsu.edu

* Correspondence: mcintosc@etsu.edu; Tel.: +1-(423)-439-5838

Current Plant Biology

Year: 2020

Volume: 24

Page: 100163

ISSN 2214-6628

https://doi.org/10.1016/j.cpb.2020.100163

https://www.sciencedirect.com/science/article/pii/S221466282030044X

Abstract

Plant natural products include a diverse array of compounds that play important roles in plant metabolism and physiology. After elucidation of biosynthetic pathways and regulatory factors, it has become possible to metabolically engineer new capabilities in planta as well as successfully engineer whole pathways into microbial systems. Microbial expression systems for producing valuable plant compounds have evolved to incorporate polyculture and co-culture consortiums for carrying out robust biosynthesis strategies. This review focuses on four classes of plant secondary metabolites and the recent advances in generating useful compounds in microbial expression platforms and in plant metabolic engineering. They are the flavonoids, alkaloids, betalains, and glucosinolates.

Keywords: Metabolic engineering; synthetic biology; flavonoids; alkaloids; betalains; glucosinolates

4.1 Introduction

Plant secondary metabolites comprise a diverse set of compounds that have been shown to serve many roles in plants. Secondary metabolites are indirectly involved in many different stages of plant development and impart an expansive set of traits that increase survivability. These traits include increasing the likelihood of plant-pollinator inter- actions, facilitating root nodule formation from nitrogen-fixating bacteria, imparting distinctive taste and color, and protecting from UV light [1–6]. Additionally, plant metabolites impart antimicrobial properties and can prevent bacterial infections[7]. An extensive library of natural products are consumed as part of the human diet that convey such benefits as acting as medicines for treating cancer, natural dyes, dietary supplements, and more [8–11]. Plant secondary metabolism and its many complexities have captivated the interests of researchers for decades. There is significant commercial interest in many of these metabolites and there exists a strong incentive to capitalize on their production [12–15]. This can be done either through metabolic engineering of plants for value-added or enhanced production of desirable compounds or reduction of undesirable compounds. Additionally, use of microbes as production "factories" using synthetic biology approaches has application in the pharmaceutical and nutraceutical industries.

Engineering plant secondary metabolism begins with elucidating the biosynthetic pathways that lead to a desired product as well as specific regulation of individual enzymes. For metabolic engineering of plants, elucidating regulatory factors and impact on changes on the metabolic and physiological networks are important to consider.

Expressing metabolites in microbial systems requires a robust balance of enzymes, cofactors, ATP and other metabolites. This is, perhaps, more amenable for those pathways with fewer enzymes, however more complex pathways have also been successfully incorporated into

microbial systems [16]. Unraveling the biosynthetic pathways for secondary plant metabolites, combined with advances in engineering through metabolic reconstitution in microbial systems, have paved the way for significant advances in the relatively new field of synthetic biology. Numerous publications have demonstrated efficient or proof of concept production of several classes of plant natural products in microbial systems [17–25]. Dramatic advances in metabolic pathway engineering using microbial production systems has already proven to be a useful tool for producing plant secondary metabolites [26–31]. This minireview provides information on advances in flavonoid, alkaloid, betalain, and glucosinolate bioengineering.

4.2 Technology

Engineering reliable biosynthesis of plant secondary metabolites currently benefits from many new and developing tools designed to inform pathway construction and optimize pathway development. These include a host of ever-expanding -omics databases, computational pathway simulations paired with custom enzyme designs, and directed evolution approaches that seek to maximize output of the desired product [32]. Advances in microbial engineering have allowed for a diverse range of plant compounds to be synthesized. This list includes pharmaceutical precursors, dyes, dietary supplements, cosmetics, fuels, and more [33–37].

One of the most important questions to consider when developing a microbial platform for biosynthesizing plant compounds is the choice of microbial host organism. There exist a number of specimens from the fungal and bacterial kingdoms useful as secondary metabolite expressions hosts, notably *Escherichia coli*, *Saccharomyces cerevisiae*, and more recently, *Corynebacterium glutamicum and Yarrowia lipolytica* [21,27,38–44]. These microbes are relatively easy to grow and there exists a substantial body of research detailing diverse genetic manipulation and metabolic modifications of these organisms [45 and references therein]. *E. coli*

has been described as having inexhaustible engineering potential while *S. cerevisiae* is a go-to yeast for its ability to incorporate complex biosynthesis pathways [46,47].

Various approaches have arisen for tackling difficulties in reconstituting multiple pathways into one microbial expression system. The ability to culture multiple strains in one bioreactor to carry out a complex biosynthesis strategy is a powerful tool recently employed to produce a range of complex plant compounds [46,47,48,49 and references therein]. These approaches allow greater diversification of pro- duct biosynthesis while simultaneously reducing the metabolic load on a cell line. Recent advances have used different strains of *E. coli* in a balanced ratio so that batch fermentation of all strains produces the desired product. Using different strains of the same organisms in "polycultures" allows streamlining growth factors, antibiotics, and downstream processing [16].

Other techniques have used microbial consortia that combine different microbes in the same culture [25]. Elucidating growth conditions that support multiple microbes can occur in different ways but is usually determined experimentally. This may include assessing inoculation ratios from seed cultures, monitoring growth using bio- sensors, or supplementing the less prolific organism with additional nutrition to balance growth [50]. The most successful growth strategies employ all of the above [25,48,50,51].

Microbial consortia can be constructed artificially, semi-synthetically, or synthetically [48]. The former of the three involves using microbes that already live symbiotically in nature with no genetic manipulations, whereas the latter involves genetically modifying all microbes used in the consortia. Semi-synthetic construction combines a naturally occurring microbe with a genetically modified one. The type of relationship between the microbial hosts being cultured may be synergistic, commensal, or mutualistic[48]. A synergistic relationship divides the

metabolic load between the consortia for a combined output that is more efficient than using one microbe alone. Commensal interactions benefit one microbe without harming or benefiting the other, whereas a mutualistic one benefits both microbes. For a full description of these constructions and interactions see Bhatia et al. and Roell et al. [48,50].

The interaction types described above are by no means comprehensive as novel consortia platforms are currently being explored. For example, a 4th mode of interaction has been identified, designated here as "synergistic-sacrificial," wherein a division of labor between two strains enhances production of the desired product but requires programmed cell death of one strain [51]. This involved the use of a genetic circuit in one *E. coli* strain that combined expression of a glucosidase with lytic genes from a T4 phage. Upon reaching maximal expression of the glucosidase, the lytic genes were triggered causing the first strain to lyse, thus releasing the glucosidase enzyme [51]. Glucosidase accumulation in the culture medium catalyzed the conversion of cellobiose, a common intermediary found in the industrial processing of biomass, into glucose. This was then used as a carbon source by a second *E. coli* strain to drive the production of isopropanol (IPA). This consortium platform exhibited cellobiose conversion rates 4-fold faster (6 h vs. 24 h) and IPA production amounts approximately 3-fold higher (5.8 ± 0.5 mM vs. 16−19 mM) than a single microbe engineered to conduct both cellobiose conversion and IPA production [51]. This example highlights the potential for microbial consortium platforms to convert biomass waste products into energy while simultaneously driving production of useful compounds. These platforms perform better than single microbe systems and have the potential to be used in engineering plant natural products.

In a recent example, a microbial consortium was developed to produce oxygenated taxanes, a class of chemotherapeutic diterpenes that cannot be feasibly produced in one microbe

[25,52]. This was accomplished by dividing taxadiene synthesis and oxygenation between engineered strains of *E. coli* and *S. cerevisiae*. The *E. coli* strain was engineered to produce taxadiene and metabolize xylose (supplied as food) to acetate [25]. *S. cerevisiae* was engineered to use acetate as a fuel source and oxygenate taxadiene [25]. Conditions for growth of recombinant *S. cerevisiae* and *E. coli* were determined individually before attempting co-culture. Their ability to carry out their respective functions was also determined beforehand experimentally. The recombinant *S. cerevisiae* was treated with stock taxadiene to evaluate its oxygenation potential and experiments were carried out to determine the yeast's ability to grow using acetate as the sole food source. Similarly, the engineered *E. coli* strain was verified to produce taxadiene and acetate at the necessary levels [25].

Successful co-culture was achieved by inoculating seed cultures of *S. cerevisiae* and *E. coli* into a defined medium (13.3 g/L KH_2PO_4, 4 g/L $(NH_4)2(HPO_4)$, 1.7 g/L citric acid, 0.0084 g/L EDTA, 0.0025 g/L $CoCl_2$, 0.015 g/L $MnCl_2$, 0.0015 g/L $CuCl_2$, 0.003 g/L H_3BO_3, 0.0025 g/L Na_2MoO_4, 0.008 g/L $Zn(CH_3COO)_2$), 0.06 g/L Fe(III) citrate, 0.0045 g/L thiamine, 1.3 g/L $MgSO_4$, pH 7.0) supplemented with yeast extract (40 g/L) and glucose or xylose (20 g/L). The yeast to *E. coli* seed culture inoculation ratio was determined experimentally. Fermentation was carried out in a bioreactor wherein dissolved oxygen levels, pH, and temperature were all optimally maintained at 30–40 %, 7.0, and 20−30 °C respectively. A recent review on co-culturing different microbes expounded on the biotechnological potential of this method and provided examples of its utility for chemical biosynthesis and other fields [48].

Modifying natural product biosynthetic pathways can also occur *in planta*. Generating a knock-out line or introducing a new gene into a plant not normally carrying out a particular reaction has proven to be quite useful in confirming the function of a gene or protein [2,21,53–

55]. New techniques have also emerged that make use of "nanomaterial-based delivery systems" that can target specific organelles without having to fully integrate new DNA into the host plant genome [56].

Examples of transgenic plants abound and their production has evolved to serve different functions [57,58]. Recently, one approach involved creating a transgenic morning glory that co-expresses AmAS1 and Am40CGT, which drives accumulation of aureusidin glucosides in order to produce flower variants with unique color [59]. These com- pounds belong to the aurone subclass of flavonoids and are partially responsible for yellow coloration in flower petals [60]. Additional studies have sought to genetically engineer accumulation of plant secondary metabolites known for their health benefits in staple crops, such as corn or wheat. These approaches have shown great promise and are the subject of several recent reviews [58,61,62]. While some successes have been made using transgenic plants as a source of pharmaceutical or nutraceutical compounds, accumulating specific secondary metabolites in microbial systems has proven to be a more desirable approach in some cases [63,64].

4.3 Flavonoids

Flavonoids belong to the phenolic class of plant secondary metabolites and serve many functions in both plants and animals. There exist thousands of identified flavonoid compounds that are divided into 9 subclasses based on the oxidation level and conformation of the middle ring (Fig. 4.1). They serve many roles in plants including modulating various defensive traits, such as UV light protection in leaves or protection from herbivores [65,66]. Flavonoids have been shown to act as signaling compounds by assisting in the formation of symbiotic colonization of Rhizobia species, a critical relationship in plants that rely on Rhizobia for nitrogen fixation [2,67]. Additionally, flavonoids have been shown to modulate plant-insect

interactions and to play a role in feeding, as well as stimulating or discouraging oviposition behavior [68].

The ubiquity of flavonoids in edible plants ensures that for herbivorous/omnivorous mammals, they are a routine part of dietary metabolism. Studies suggest they impart many physiological benefits such as reducing inflammation and increasing radical scavenging capabilities. Indeed, numerous flavonoid extracts are marketed as dietary supplements [69].

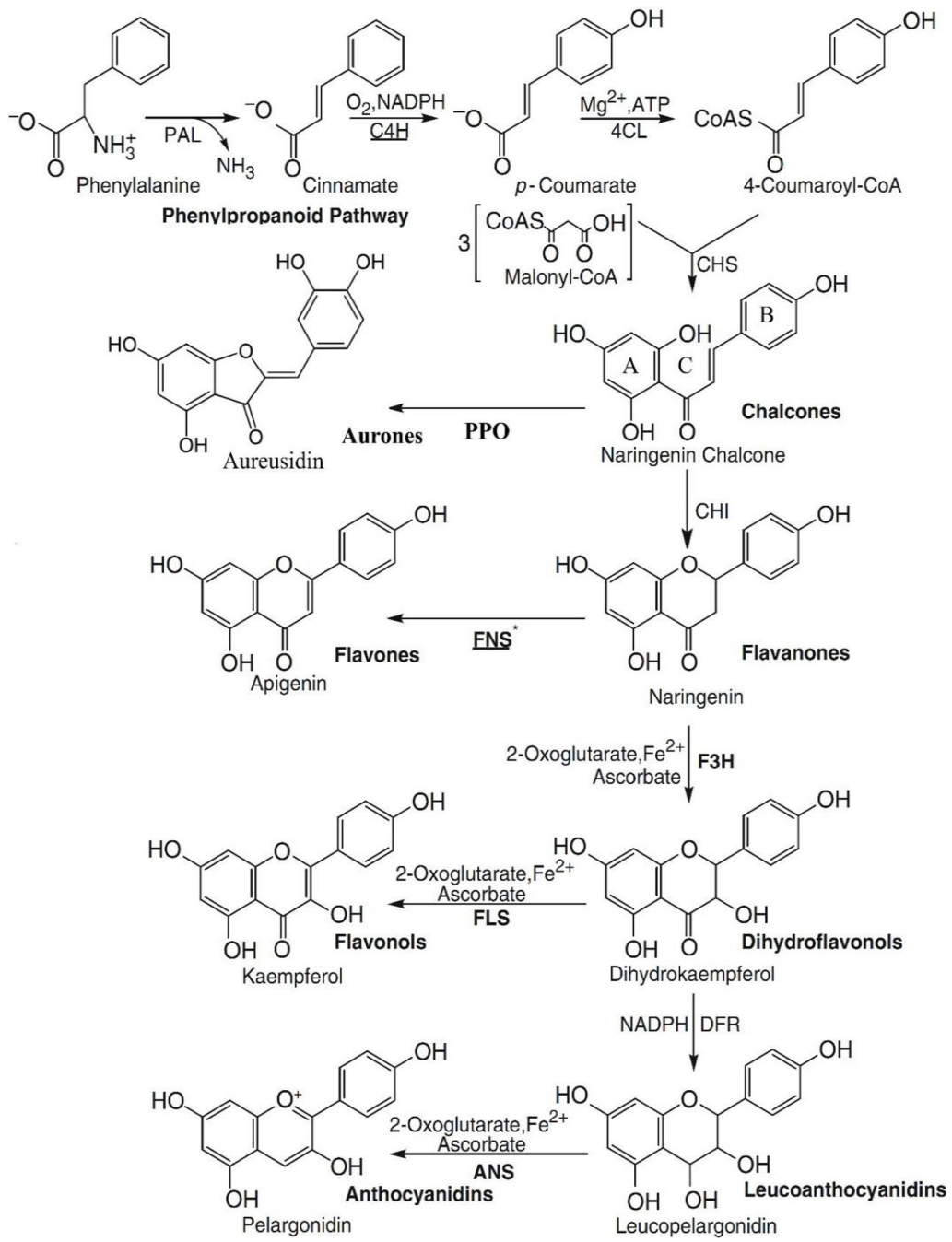

Fig. 4.1 Biosynthesis of Flavonoid Classes Through the Phenylpropanoid Pathway (adapted from [70]). Enzymes shown are phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), plant polyphenol oxidase (PPO), chalcone isomerase (CHI), flavone synthase (FNS), flavanone 3β-hydroxylase

(F3H), flavonol synthase (FLS), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS).

Many flavonoids are associated with health benefits in humans and some show the potential to inhibit angiogenesis and upregulate apoptosis in cancerous cells [71–75]. A recent investigation into the anti-cancerous potential of flavonoids extracted from blueberries highlights the importance of these compounds. The study showed that the principal flavonoids present in *Vaccinium myrtillus* (quercetin, kaempferol, and gentisic acid) had significant apoptotic effects on colorectal cancer cells [76].

The extensive use and application of plant flavonoids has driven refinement of recombinant means to produce these compounds. The core pathways for flavonoid biosynthesis have been elucidated and their chemical/biological potential has been thoroughly reviewed [70,77–79]. While all terrestrial plants synthesize flavonoids, the class of flavonoid, specific final products, and amount produced can vary significantly [26]. Furthermore, derivatives of flavonoids that do not occur in nature have been synthesized and are being investigated for their pharmaceutical potential [80].

Flavonoid biosynthesis begins with the production of naringenin chalcone from 4-coumaroyl CoA and 3 malonyl CoA, which is catalyzed by chalcone synthase (CHS) [81]. The chalcone backbone subsequently gives rise to a series of other reactions resulting in the synthesis of other flavonoid subclasses (Fig. 4.1). Chalcones can be converted to aurones by plant polyphenol oxidase (PPO) or to flavanones by chalcone isomerase (CHI). Flavanones can be converted to flavones by flavone synthase (FNS) or to dihydroflavonols by flavanone 3−OH transferase. In turn, dihydroflavonols can be converted to flavonols by flavonol synthase (FLS) or to leucocyanins by dihydroflavonol reductase(DFR). Antho- cyanins are synthesized from

leucocyanidins by a two-step process catalyzed by anthocyanin synthase (ANS) followed by glycosylation. Reconstituting these pathways requires encoding all genetic information relevant to the synthesis of the desired flavonoids into the microbial host genome. Furthermore, expression of plant genes necessary for biosynthesis must be fine-tuned to avoid undesirable product pathways and upregulate genes that lead to adequate precursor synthesis [44 and references therein]. Modifications to the final product may also be needed. These include glucosylation, prenylation, methylation, sulfation, and more. Flavonoids have been shown to be more biologically active when glycosylated making this a desirable modification for many synthesis platforms [82,83]. These steps have been successfully carried out in both *E. coli* and *S. cerevisiae* [84,85].

A recent endeavor to produce flavonols and anthocyanidin-3-O-glucosides made use of *E. coli* polycultures, a mix of *E. coli* strains each designed to carry out a specific biosynthesis step [16]. This system began by generating *E. coli* cultures with plasmids that overexpressed phenylpropanoic acids, specifically p-coumaric acid and caffeic acid. The first strain used xylose, glucose, and glycerol to produce a precursor pool of phenylpropanoic acids used in chalcone synthesis. A second strain carried out production of a flavanone, which was designed to be converted to a dihydroflavonol by a third strain. This allowed a fourth strain to achieved production of the anthocyanin callistephin [16]. A 5th strain was proposed that could be used for further downstream modifications to synthesize non-natural flavonoid compounds. Both the flavonoid and stilbene classes of polyphenols have been altered in microbial organisms to produce non-natural variants, with some possibly having medical or pharmaceutical importance [45,86].

When attempting to fine tune a microbial synthesis platform for flavonoid biosynthesis, notable obstacles have been overcome. One example highlights the rapidly improving potential for optimizing commercial biosynthesis of these compounds. This example involves the precursor malonyl CoA. Three malonyl CoA are combined with 4- Coumaroyl CoA to produce chalcone, the flavonoid backbone from which different classes of flavonoids are derived. Malonyl CoA avail- ability has been shown to be a limiting factor when trying to express flavonoids in microbial systems [31,87]. Recently, *C. glutamicum* has been used to express naringenin and resveratrol at production titers markedly improved due to engineering strategies that increased malonyl CoA availability [44]. A similar endeavor to modulate malonyl CoA metabolism used *E. coli* and CRISPR editing to produce naringenin at greater amounts due to a larger malonyl CoA pool [88].

Another approach used the oleaginous yeast *Y. lipolytica* to engineer hydroxylated flavonoid production [27]. This microbe provides a lipophilic environment that is beneficial for the expression of plant enzymes needed to produce flavonoids, such as chalcone synthase and cytochrome P450 reductase [81]. Furthermore, *Y. liplytica* has a naturally high flux of malonyl-CoA [27]. This microbe benefitted from engineered expression of genes responsible for chalcone synthase and cytochrome p450 reductase, as well as genes that upregulate malonyl CoA production. The engineered strain was able to produce titers of naringenin at concentrations of 250 mg/L and 134 mg/L respectively [27].

4.4 Alkaloids

Different structural classes of alkaloids exist that include benzylisoquinoline alkaloids, monoterpene indole alkaloids, bisbenzylisoquinoline alkaloids, and the tropane and nicotine alkaloids (Fig. 4.2) [89,90]. Numerous pathways exist for synthesizing these compounds but all

originate from precursor amino acids, such as tryptophan or aspartic acid. For example, the monoterpene indole pathways begin when L-tryptophan is converted to tryptamine, which is then used as a substrate to form strictosidine, the precursor for all monoterpene indole derivatives (Fig. 4.2) [91].

Alkaloids are used in plants for self-defense and can exhibit direct toxicity to herbivorous organisms [92]. Alkaloids may also inhibit or alter important metabolic pathways within the consuming animal. Plants use alkaloids as nitrogen sinks for suppling pools of chemicals important to the synthesis of wound-healing metabolites [93,94]. Like flavonoids, alkaloids have also been shown to influence plant-insect interactions and can deter insect feeding [92]. Elucidated pathways of monoterpene indole alkaloids (MIAs) have shown that when under attack, *Catharanthus roseus* deglucosylates the alkaloid strictosidine-4, converting it into a dialdehyde compound that can cross link proteins [95,96]. This mechanism of defense has also been found for glucosinolates found in mustard oil discussed further below [53].

Plant alkaloids have long been used by humans, with their usage recorded prominently throughout ancient history and into today [90,97 and references therein]. The thousands of plant alkaloids currently known to exist represent a vast pharmacopeia of useful compounds that include medicines, stimulants, and many psychoactive substances [98,99]. Consequently, there is a strong incentive to produce a wide range of alkaloid compounds in microbial systems, with many recent reviews highlighting its potential to transform drug manufacture and other fields of industry [33,100,101].

Chemotherapeutic alkaloids such as taxol are derived from plant sources, as are other alkaloids such as opiates for treating pain [102,103]. These compounds are fundamental to modern medicine yet extracting high value medicinal compounds from plant sources can be

complicated by agricultural or environmental factors [104]. Biosynthesizing plant alkaloids in microbial systems is not subject to these constraints. Consequently, optimizing microbial engineering towards the synthesis of useful alkaloids has garnered an abundance of research interest. Recent advances have been made regarding complete biosynthesis of alkaloids in microbial systems, semisynthesis of alkaloids, and engineering alkaloids *in planta* [105].

Complete biosynthesis of alkaloids in microbial systems offers the greatest potential for generating a safe and affordable basis for pharmaceutical manufacture. A proof of concept manufacturing process was recently described using *S. cerevisiae* to engineer benzylisoquinoline alkaloids used in opiate production, specifically hydrocodone [106]. This process achieved titers of less than 1 μg/L, which fell short of the 5 g/L projection needed to be a suitable alternative to poppy cultivation, yet an improved process was reported shortly thereafter that used engineered *E. coli* to synthesize the alkaloid thebaine, a precursor used in the semi-synthetic production of hydrocodone and codeine [107]. At titers of approximately 2 mg/L, this method showed a 300-fold increase in production capacity over previous methods. Furthermore, these precursors were used to biosynthesize hydrocodone in *S. cerevisiae*, demonstrating that complete biosynthesis of valuable alkaloid com- pounds in microbes is possible [107]. Rapidly advancing techniques in microbial engineering could reduce the need for plant-based extraction of alkaloids as well as lowering the cost of chemically based synthesis strategies. However, not all alkaloids can be completely synthesized in microbes. A recent review discussed in great detail the various chemical and biochemical routes needed for complete biosynthesis of many alkaloids [54].

Employing complex chemical syntheses of alkaloid based drugs often cannot meet the demand for large-scale production, resulting in limited supplies of many lifesaving therapeutics. Semisynthesis of alkaloids involves using microbe biosynthesis platforms to generate high levels

of precursor alkaloids used in drug manufacture. The desired precursor is biosynthesized in the microbe of choice, extracted, and subjected to further synthesizing steps. This technique has emerged as an attractive way to simplify multistep synthesis strategies and drive down the cost of drug manufacture. A recent approach engineered *E. coli* and *S. cerevisiae* to produce guaia-6,10(14)-diene via the mevalonate pathway [105]. This starting material is used in the synthesis of englerin A, a chemotherapeutic agonist of the transient receptor potential (TRP) channels implicated in some cancers [55,108,109]. For alkaloids that cannot be completely expressed in microbial systems, this semisynthetic approach shows great promise in expanding drug availability and reducing cost of manufacture, as highlighted in recent reviews [33,54,110].

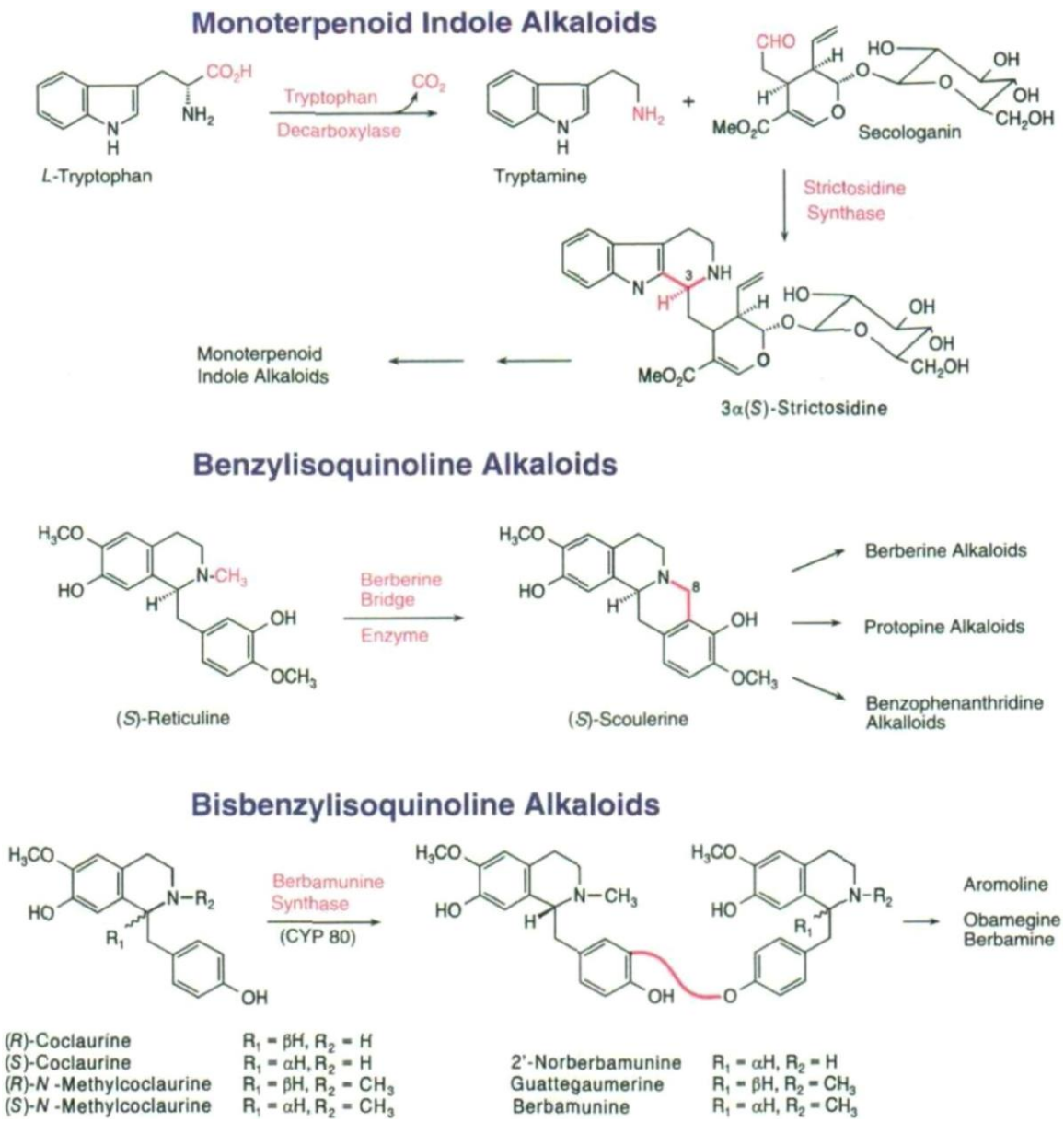

Fig. 4.2 Pathways for 3 subclasses of alkaloids (adapted from [89])

Despite exciting developments in full and semi biosynthesis of alkaloids, plant extraction remains a vital component of producing important alkaloid drugs [111]. Significant research has been focused towards engineering optimized and enhanced expression of alkaloids in plants [112–115]. The opium poppy, *Papavar somniferum*, is perhaps the most recognizable plant source due to its supply of alkaloids necessary for synthesizing pain relieving medication. Not

surprisingly, gene editing techniques have been employed to upregulate expression of metabolites that regulate alkaloid biosynthesis. Recently, a CRISPER/Cas9 system was designed to knock out the 4′OMT2 gene in *P. somniferum,* which has been shown to control biosynthesis of thebaine, noscapine, and morphine [100,112]. This technique generated a transgenic poppy plant that had markedly reduced biosynthesis of benzylisoquinoline alkaloids, indicating alkaloid biosynthesis can be regulated in transgenic plants using CRISPER/Cas9 gene drives [112]. Other plants, such as *C. roseus*, have been targeted for genetically engineering their production of the anti-cancerous vinca alkaloids vinblastine and vincristine [104,115,116]. A recent review highlighted a wide scope of engineering practices for upregulating vinca alkaloids in this plant [113]. Hairy root cultures of the Iranian flower *Hyoscyamus senecionis* and *H. muticus*, were also recently targeted for engineering of the anticholinergic and deliriant tropane alkaloids [114].

4.5 Betalains

Betalains are nitrogen containing compounds produced largely in plants in the Caryophyllales such as cacti and red beets. They are synthesized from tyrosine and fall into two main subclasses, the beta- cyanins that are red/purple and the betaxanthins that are yellow (Fig. 4.3).

Because of the similarity of colors of betacyanins and the anthocyanin flavonoid pigments, significant research has been conducted with respect to biosynthesis and addressing the interesting phenomenon that anthocyanins are not found in plants producing betalains even as other flavonoid compounds are found in these plants [35,117]. Different scenarios have been suggested for this ranging from loss of function to gene deletion. Richardson [117] and Polturak et al. [35] recently published results from a study of transcriptome and metabolite composition in *Mirabilis jalapa* that showed that anthocyanin synthase was present and highly expressed

however the gene contained a deletion in the active site [35,117] rendering the enzyme inactive. These data support the idea of loss of function as one possible explanation of this exclusivity between betalains and anthocyanins and how anthocyanin production could have been lost over time in this group of plants. It will be interesting to see if this pattern of the ANS gene being present and expressed but having a loss of function through mutation will observed in other betalain producing plants, or if there are other variations.

Betalains are known for the vibrant colors that they impart to plants where they are found, serving as pollinator cues as well as for their potent antioxidant properties [35,118]. Synthesis of these compounds in plants is increased in response to stress [118]. Several contributions to improvement of human health have been identified that have contributed to the increased interest in these compounds. These include positive effects on hypertension, cancer, and more and has recently been reviewed [119,120]. These properties, including the ability to use these compounds as natural food colorants and dyes, show significant potential for commercial applications [121]. Elucidation of the biosynthetic pathway has been sought for decades and was resolved through recognition that some steps involve spontaneous (i.e., not enzyme catalyzed) reactions [35,122]. As stated previously, betalains are synthesized from tyrosine leading to production of a key intermediate, betalamic acid. This is done by a cytochrome P450 converting tyrosine to L-DOPA followed by conversion of L-DOPA to betalamic acid through the action of DOPA 4,5-dioxygenase. Synthesis of betaxanthins occurs by spontaneous condensation of betalamic acid with amino acids such as glycine or proline or from condensation with amines such as tyramine or dopa- mine [35 and references therein]. Synthesis of betacyanins occurs through spontaneous condensation of betalamic acid with cyclo-DOPA followed by

derivatization or “decorating” reactions such as addition of sugars and/or acyl groups [40]. As a result, over 70 naturally occurring betalains have been identified from plants [35,122].

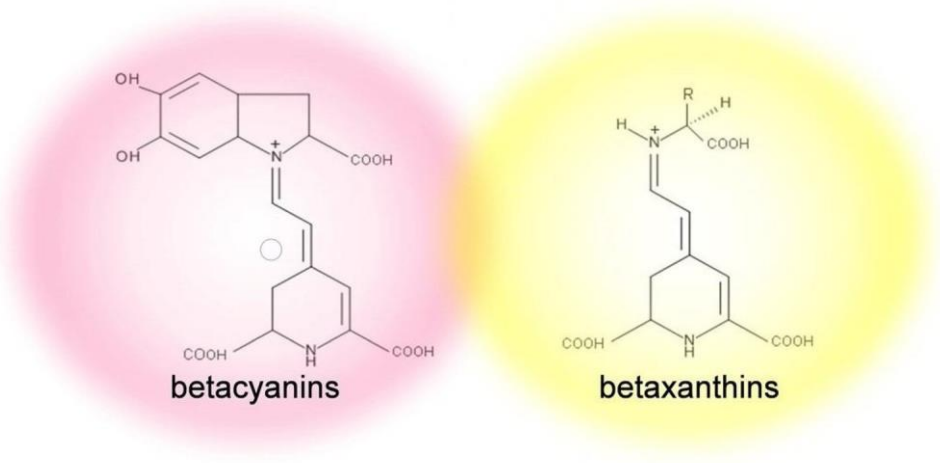

Fig. 4.3 The two main subclasses of betalains, betacyanins and betaxanthins [20]

This group of compounds has found renewed interest as a target of metabolic engineering at least in part due to the properties mentioned above as well as elucidation of betalain biosynthesis. Recent reviews have been published by Polturak and Aharoni [35,123]. These reviews included consideration of biosynthesis, metabolism, applications, and current state of metabolic engineering. Because of the interest in this field, significant publications have appeared since those reviews. Additional factors to consider for engineering in plants includes tyrosine and other amino acid partitioning [124,125]. Other key contributions to the metabolic engineering toolbox have included identification of key genes and transcription factors [118,123,124].

Tian et al. published results of their work to engineer betanin into rice endosperm as a health promoting food additive to increase health benefits of rice as well as provide a potential source of raw material for commercial supplement production [126]. Because of the relative

simplicity of the pathway, they introduced only 3 genes to be overexpressed in rice, meloS, BvDODAIS, and BcCYP76ADIS. This work included determination of antioxidant capacity, grain yield, and quality of starch granules. The latter two parameters were comparable to wild type, while the transgenic rice had 4−6x higher antioxidant activity. This as well as the successful engineering of betalain synthesis into microbes, serve as concrete examples of the success of metabolic engineering of this pathway and likely will lead to further successful applications [21,120].

4.6 Glucosinolates

Glucosinolates are plant secondary natural products derived from both aliphatic and aromatic amino acids that are notably found in the Brassicales and related plants. Their structure is composed of a sulfonated oxime, a thioglucose moiety, and a side chain originating from the aliphatic and aromatic amino acids (Fig. 4.4) [127–129]. Further structural diversity occurs through modification reactions. A rigorous vetting of structures and compound identification has recently been published by Blazevic et al. [129]. To date, there are over 150 naturally occurring glucosinolates identified from plants.

Like the cyanogenic glycosides, glucosinolates are compartmentalized from degrading enzymes in plant tissues. Following mechanical disruption of tissue, they are hydrolyzed by endogenous myrosinases, a specific group of glucosidases that act on glucosinolates [128,129]. The products of hydrolysis include isothiocyanates and thiocyanates that have diverse biological activity (Fig. 4.5). In plants, these compounds provide defense against insects and microbes, as the hydrolysis products can lead to production of phytoalexins [130,131]. They have been shown to serve as antifeedant defense compounds for both insects and animal foragers, as well as feeding attractants and oviposition signals in insects preferring crucifers [127,128,130].

In addition to the mustard and other taste properties of the Brassica, glucosinolates also impact animal health and nutrition. Examples of health benefit include cancer prevention and inhibition of proliferation, stimulation of the immune system, and reduction of heart disease [132,133]. Nutritional and health considerations include bioavailability of glucosinolates and/or their breakdown products isothiocyanates [134].

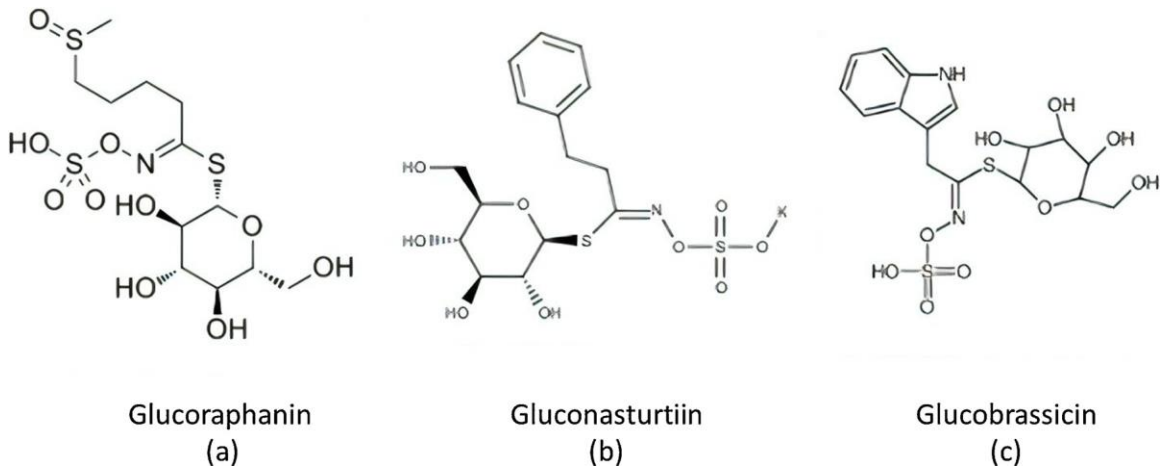

Fig. 4.4 Examples of glucosinolates synthesized from a) methionine, b) phenylalanine, and c) tryptophan.

Because of the benefits to plants and animals, there is significant interest in the metabolic engineering of glucosinolates into plants, enhancement of production in plants, and commercial production of these compounds for use as nutritional supplements. As stated previously, for successful engineering it is imperative that the biosynthetic pathway, metabolism, regulation, and physiological impact of altering metabolism in the target organism is understood.

Blazevic et al. recently published a comprehensive review of glucosinolate biosynthesis [129]. It includes consideration of synthesis from aliphatic and aromatic amino acids as well as

key studies to identify genes and elucidate their functions (see also [135]); seven biosynthetic steps have been fully elucidated. Arabidopsis synthesizes glucosinolates thereby providing an amenable model system and powerful tool for identification of key genes and potential regulatory factors [136,137]. Transformation of genes into non-glucosinolate producing plants or systems was an important approach to further confirm function of the cytochrome P450 enzymes as well as the hydrolytic enzymes [22,136,137]. Knowledge of key genes in Arabidopsis has been used to identify potential genes in other species through comparative genetics [24].

Glucosinolate metabolism and regulation in plants has been studied in several systems including consideration of levels in different tissues and different stages [22–24,135]. One important aspect is the use of tryptophan for synthesis of both indole glucosinolates and indole acetic acid (IAA) and the metabolic balance of tryptophan use between these pathways [135,138]. MYB transcription factors have been identified that control synthesis [e.g. [130,139 and references therein]. Circadian regulation has also been established [130].

Recent reviews on development of tools and strategies for metabolic engineering and on use of *E. coli* as a host have shown the significance of these contributions and the advancements in technological approaches [32,140]. Glucosinolate metabolic engineering has encompassed transformation of genes into plants or systems that do not normally synthesize glucosinolates, as well as efforts to improve production in crop plants to improve nutritional value [22–24,136]. Building upon transformation of *E. coli* to produce glucoraphanin, Petersen et al. recently were able to engineer *E. coli* to produce benzyl glucosinolate [29,141]. They used a series of strategies evaluating host strain and culture conditions as well as monitoring protein expression levels to optimize levels of production and to further increase synthesis five-fold [141]. In a perhaps somewhat related effort, Liou et al. have recently identified the operon used by a human gut

symbiont responsible for transformation of glucosinolates into health promoting isothiocyanates by engineering this operon into a non-metabolizing strain and showing gain of function [142].

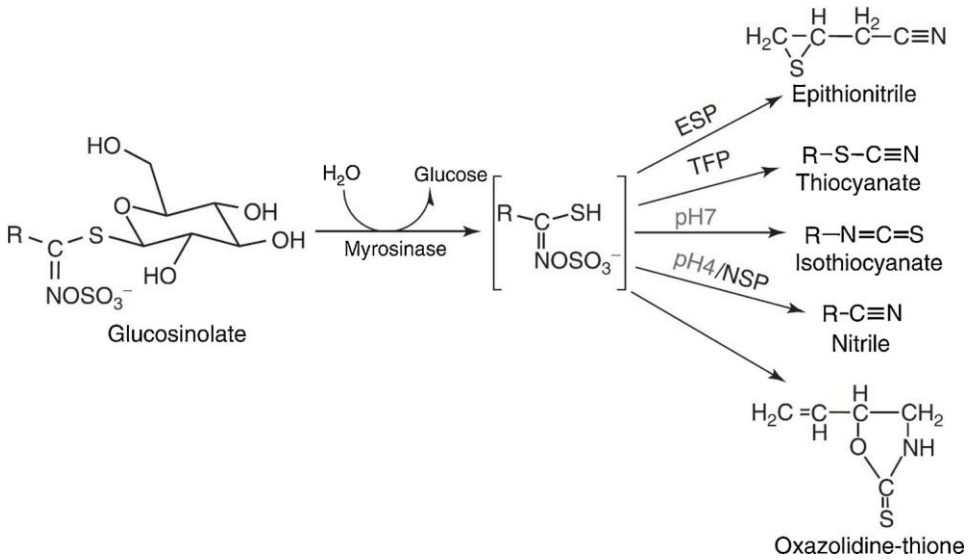

Fig. 4.5 Possible Products of Glucosinolate Hydrolysis [143]. ESP = Epithiospecifier protein; TFP = thiocyanate forming protein; NSP = nitrile specifier protein.

4.7 Summary

This mini-review summarized recent advances in synthetic biology and metabolic engineering for four classes of plant natural products; flavonoid, alkaloids, betalains, and glucosinolates. The importance of plant natural products to both plants and animals coupled with advances in technological approaches has resulted in being able to engineer plants for value-added crops, increased resistance/defense, enhanced coloration, and more. As we learn more about *in planta* regulation and control of metabolism and metabolic networks, we will continue to see increased success and efficiency in this area.

An overview of recent technological advances in synthetic biology was also given. Emphasis was placed on microbial platforms as a highly useful tool for engineering these natural

products with additional consideration given to microbial consortia. The ability to culture robust assortments of microbes with varying functionality has dramatically expanded to allow a wider scope of products to be engineered in increasingly efficient ways. These technological breakthroughs are continuing to develop at a rapid pace and coincide with new discoveries regarding the use of plant natural products. This is especially so for those compounds with benefits to human health. Improved approaches for synthetic biology and scalable production for pharmaceutically effective compounds as well as nutraceuticals have been significant. Further optimization and technological advances as well as continued enlightenment of the intricacies of the metabolic networks involving plant natural products will continue to advance the field.

4.8 Authorship Contribution Statement

Both authors contributed equally to the preparation and writing of this manuscript.

4.9 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Wink, M. Plant Secondary Metabolites Modulate Insect Behavior-Steps Toward Addiction? *Front Physiol* **2018**, *9*, doi:10.3389/fphys.2018.00364.

2. Wasson, A.P.; Pellerone, F.I.; Mathesius, U. Silencing the Flavonoid Pathway in *Medicago Truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. *Plant Cell* **2006**, *18*, 1617–1629, doi:10.1105/tpc.105.038232.

3. Verpoorte, R.; Memelink, J. Engineering Secondary Metabolite Production in Plants. *Curr Opin Biotechnol* **2002**, *13*, 181–187, doi:10.1016/S0958-1669(02)00308-7.

4. Logemann, E.; Wu, S.-C.; Schröder, J.; Schmelzer, E.; Somssich, I.E.; Hahlbrock, K. Gene Activation by UV Light, Fungal Elicitor or Fungal Infection in *Petroselinum Crispum* Is Correlated with Repression of Cell Cycle-Related Genes. *The Plant Journal* **1995**, *8*, 865–876, doi:10.1046/j.1365-313X.1995.8060865.x.

5. Stevenson, P.C.; Nicolson, S.W.; Wright, G.A. Plant Secondary Metabolites in Nectar: Impacts on Pollinators and Ecological Functions. *Funct Ecol* **2017**, *31*, 65–75, doi:10.1111/1365-2435.12761.

6. Yang, L.; Wen, K.-S.; Ruan, X.; Zhao, Y.-X.; Wei, F.; Wang, Q. Response of Plant Secondary Metabolites to Environmental Factors. *Molecules* **2018**, *23*, 762, doi:10.3390/molecules23040762.

7. Wallace, R.J. Antimicrobial Properties of Plant Secondary Metabolites. *Proceedings of the Nutrition Society* **2004**, *63*, 621–629, doi:10.1079/PNS2004393.

8. Seca, A.; Pinto, D. Plant Secondary Metabolites as Anticancer Agents: Successes in Clinical Trials and Therapeutic Application. *Int J Mol Sci* **2018**, *19*, 263, doi:10.3390/ijms19010263.

9. Siva, R.; Mayes, S.; Behera, S.K.; Rajasekaran, C. Anthraquinones Dye Production Using Root Cultures of *Oldenlandia Umbellata* L. *Ind Crops Prod* **2012**, *37*, 415–419, doi:10.1016/j.indcrop.2011.12.027.

10. Kennedy, D.O.; Wightman, E.L. Herbal Extracts and Phytochemicals: Plant Secondary Metabolites and the Enhancement of Human Brain Function. *Advances in Nutrition* **2011**, *2*, 32–50, doi:10.3945/an.110.000117.

11. *Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet*; Crozier, A., Clifford, M.N., Ashihara, H., Eds.; Blackwell Publishing, 2006; ISBN 9781405125093.

12. Sholikhah, E.N. Indonesian Medicinal Plants as Sources of Secondary Metabolites for Pharmaceutical Industry. *Journal of thee Medical Sciences (Berkala Ilmu Kedokteran)* **2016**, *48*, 226–239, doi:10.19106/JMedSci004804201606.

13. Suástegui, M.; Shao, Z. Yeast Factories for the Production of Aromatic Compounds: From Building Blocks to Plant Secondary Metabolites. *J Ind Microbiol Biotechnol* **2016**, *43*, 1611–1624, doi:10.1007/s10295-016-1824-9.

14. Isah, T.; Umar, S.; Mujib, A.; Sharma, M.P.; Rajasekharan, P.E.; Zafar, N.; Frukh, A. Secondary Metabolism of Pharmaceuticals in the Plant in Vitro Cultures: Strategies, Approaches, and Limitations to Achieving Higher Yield. *Plant Cell, Tissue and Organ Culture (PCTOC)* **2018**, *132*, 239–265, doi:10.1007/s11240-017-1332-2.

15. Rafińska, K.; Pomastowski, P.; Wrona, O.; Górecki, R.; Buszewski, B. *Medicago Sativa* as a Source of Secondary Metabolites for Agriculture and Pharmaceutical Industry. *Phytochem Lett* **2017**, *20*, 520–539, doi:10.1016/j.phytol.2016.12.006.

16. Jones, J.A.; Vernacchio, V.R.; Collins, S.M.; Shirke, A.N.; Xiu, Y.; Englaender, J.A.; Cress, B.F.; McCutcheon, C.C.; Linhardt, R.J.; Gross, R.A.; et al. Complete Biosynthesis of Anthocyanins Using *E. Coli* Polycultures. *mBio* **2017**, *8*, doi:10.1128/mBio.00621-17.

17. Pyne, M.E.; Narcross, L.; Martin, V.J.J. Engineering Plant Secondary Metabolism in Microbial Systems. *Plant Physiol* **2019**, *179*, 844–861, doi:10.1104/pp.18.01291.

18. Ajikumar, P.K.; Xiao, W.-H.; Tyo, K.E.J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T.H.; Pfeifer, B.; Stephanopoulos, G. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia Coli*. *Science (1979)* **2010**, *330*, 70–74, doi:10.1126/science.1191652.

19. Ding, M.; Yan, H.; Li, L.; Zhai, F.; Shang, L.; Yin, Z.; Yuan, Y. Biosynthesis of Taxadiene in *Saccharomyces Cerevisiae*: Selection of Geranylgeranyl Diphosphate Synthase Directed by a Computer-Aided Docking Strategy. *PLoS One* **2014**, *9*, e109348, doi:10.1371/journal.pone.0109348.

20. Polturak, G.; Grossman, N.; Vela-Corcia, D.; Dong, Y.; Nudel, A.; Pliner, M.; Levy, M.; Rogachev, I.; Aharoni, A. Engineered Gray Mold Resistance, Antioxidant Capacity, and Pigmentation in Betalain-Producing Crops and Ornamentals. *Proc Natl Acad Sci U S A* **2017**, *114*, 9062–9067, doi:10.1073/pnas.1707176114.

21. Grewal, P.S.; Modavi, C.; Russ, Z.N.; Harris, N.C.; Dueber, J.E. Bioproduction of a Betalain Color Palette in *Saccharomyces Cerevisiae*. *Metab Eng* **2018**, *45*, 180–188, doi:10.1016/j.ymben.2017.12.008.

22. Pfalz, M.; Mikkelsen, M.D.; Bednarek, P.; Olsen, C.E.; Halkier, B.A.; Kroymann, J. Metabolic Engineering in *Nicotiana Benthamiana* Reveals Key Enzyme Functions in *Arabidopsis* Indole Glucosinolate Modification. *Plant Cell* **2011**, *23*, 716–729, doi:10.1105/tpc.110.081711.

23. Qian, H.; Sun, B.; Miao, H.; Cai, C.; Xu, C.; Wang, Q. Variation of Glucosinolates and Quinone Reductase Activity among Different Varieties of Chinese Kale and Improvement of Glucoraphanin by Metabolic Engineering. *Food Chem* **2015**, *168*, 321–326, doi:10.1016/j.foodchem.2014.07.073.

24. Wang, J.; Qiu, Y.; Wang, X.; Yue, Z.; Yang, X.; Chen, X.; Zhang, X.; Shen, D.; Wang, H.; Song, J.; et al. Insights into the Species-Specific Metabolic Engineering of Glucosinolates in Radish (*Raphanus Sativus* L.) Based on Comparative Genomic Analysis. *Sci Rep* **2017**, *7*, 16040, doi:10.1038/s41598-017-16306-4.

25. Zhou, K.; Qiao, K.; Edgar, S.; Stephanopoulos, G. Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products. *Nat Biotechnol* **2015**, *33*, 377–383, doi:10.1038/nbt.3095.

26. Trantas, E.A.; Koffas, M.A.G.; Xu, P.; Ververidis, F. When Plants Produce Not Enough or at All: Metabolic Engineering of Flavonoids in Microbial Hosts. *Front Plant Sci* **2015**, *6*, doi:10.3389/fpls.2015.00007.

27. Lv, Y.; Marsafari, M.; Koffas, M.; Zhou, J.; Xu, P. Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis. *ACS Synth Biol* **2019**, *8*, 2514–2523, doi:10.1021/acssynbio.9b00193.

28. Yang, Y.; Lin, Y.; Li, L.; Linhardt, R.J.; Yan, Y. Regulating Malonyl-CoA Metabolism via Synthetic Antisense RNAs for Enhanced Biosynthesis of Natural Products. *Metab Eng* **2015**, *29*, 217–226, doi:10.1016/j.ymben.2015.03.018.

29. Yang, H.; Liu, F.; Li, Y.; Yu, B. Reconstructing Biosynthetic Pathway of the Plant-Derived Cancer Chemopreventive-Precursor Glucoraphanin in *Escherichia Coli*. *ACS Synth Biol* **2018**, *7*, 121–131, doi:10.1021/acssynbio.7b00256.

30. Lim, C.G.; Fowler, Z.L.; Hueller, T.; Schaffer, S.; Koffas, M.A.G. High-Yield Resveratrol Production in Engineered *Escherichia Coli*. *Appl Environ Microbiol* **2011**, *77*, 3451–3460, doi:10.1128/AEM.02186-10.

31. Li, M.; Schneider, K.; Kristensen, M.; Borodina, I.; Nielsen, J. Engineering Yeast for High-Level Production of Stilbenoid Antioxidants. *Sci Rep* **2016**, *6*, 36827, doi:10.1038/srep36827.

32. Chae, T.U.; Choi, S.Y.; Kim, J.W.; Ko, Y.-S.; Lee, S.Y. Recent Advances in Systems Metabolic Engineering Tools and Strategies. *Curr Opin Biotechnol* **2017**, *47*, 67–82, doi:10.1016/j.copbio.2017.06.007.

33. Ehrenworth, A.M.; Peralta-Yahya, P. Accelerating the Semisynthesis of Alkaloid-Based Drugs through Metabolic Engineering. *Nat Chem Biol* **2017**, *13*, 249–258, doi:10.1038/nchembio.2308.

34. Vavitsas, K.; Fabris, M.; Vickers, C. Terpenoid Metabolic Engineering in Photosynthetic Microorganisms. *Genes (Basel)* **2018**, *9*, 520, doi:10.3390/genes9110520.

35. Polturak, G.; Aharoni, A. "La Vie En Rose": Biosynthesis, Sources, and Applications of Betalain Pigments. *Mol Plant* 2018, *11*, 7–22.

36. De Masi, L.; Bontempo, P.; Rigano, D.; Stiuso, P.; Carafa, V.; Nebbioso, A.; Piacente, S.; Montoro, P.; Aversano, R.; D'Amelia, V.; et al. Comparative Phytochemical Characterization, Genetic Profile, and Antiproliferative Activity of Polyphenol-Rich Extracts from Pigmented Tubers of Different *Solanum Tuberosum* Varieties. *Molecules* **2020**, *25*, doi:10.3390/molecules25010233.

37. Tippmann, S.; Scalcinati, G.; Siewers, V.; Nielsen, J. Production of Farnesene and Santalene by *Saccharomyces Cerevisiae* Using Fed-batch Cultivations with *RQ* -controlled Feed. *Biotechnol Bioeng* **2016**, *113*, 72–81, doi:10.1002/bit.25683.

38. Lehka, B.J.; Eichenberger, M.; Bjørn-Yoshimoto, W.E.; Vanegas, K.G.; Buijs, N.; Jensen, N.B.; Dyekjær, J.D.; Jenssen, H.; Simon, E.; Naesby, M. Improving Heterologous Production of Phenylpropanoids in *Saccharomyces Cerevisiae* by Tackling an Unwanted Side Reaction of Tsc13, an Endogenous Double Bond Reductase. *FEMS Yeast Res* **2017**, fox004, doi:10.1093/femsyr/fox004.

39. Kwak, S.; Jin, Y.-S. Production of Fuels and Chemicals from Xylose by Engineered *Saccharomyces Cerevisiae*: A Review and Perspective. *Microb Cell Fact* **2017**, *16*, 82, doi:10.1186/s12934-017-0694-9.

40. Mark, R.; Lyu, X.; Ng, K.R.; Chen, W.N. Gene Source Screening as a Tool for Naringenin Production in Engineered *Saccharomyces Cerevisiae*. *ACS Omega* **2019**, *4*, 12872–12879, doi:10.1021/acsomega.9b00364.

41. Ruprecht, C.; Bönisch, F.; Ilmberger, N.; Heyer, T. V.; Haupt, E.T.K.; Streit, W.R.; Rabausch, U. High Level Production of Flavonoid Rhamnosides by Metagenome-Derived Glycosyltransferase C in *Escherichia Coli* Utilizing Dextrins of Starch as a Single Carbon Source. *Metab Eng* **2019**, *55*, 212–219, doi:10.1016/j.ymben.2019.07.002.

42. Shrestha, A.; Pandey, R.P.; Dhakal, D.; Parajuli, P.; Sohng, J.K. Biosynthesis of Flavone C-Glucosides in Engineered *Escherichia Coli*. *Appl Microbiol Biotechnol* **2018**, *102*, 1251–1267, doi:10.1007/s00253-017-8694-6.

43. Jeandet, P.; Sobarzo-Sánchez, E.; Clément, C.; Nabavi, S.F.; Habtemariam, S.; Nabavi, S.M.; Cordelier, S. Engineering Stilbene Metabolic Pathways in Microbial Cells. *Biotechnol Adv* **2018**, *36*, 2264–2283, doi:10.1016/j.biotechadv.2018.11.002.

44. Milke, L.; Ferreira, P.; Kallscheuer, N.; Braga, A.; Vogt, M.; Kappelmann, J.; Oliveira, J.; Silva, A.R.; Rocha, I.; Bott, M.; et al. Modulation of the Central Carbon Metabolism of *Corynebacterium Glutamicum* Improves Malonyl-CoA Availability and Increases Plant Polyphenol Synthesis. *Biotechnol Bioeng* **2019**, *116*, 1380–1391, doi:10.1002/bit.26939.

45. Pandey, R.P.; Parajuli, P.; Koffas, M.A.G.; Sohng, J.K. Microbial Production of Natural and Non-Natural Flavonoids: Pathway Engineering, Directed Evolution and Systems/Synthetic Biology. *Biotechnol Adv* 2016, *34*, 634–662.

46. Turanlı-Yıldız, B.; Hacısalihoğlu, B.; Çakar, Z.P. Advances in Metabolic Engineering of *Saccharomyces Cerevisiae* for the Production of Industrially and Clinically Important Chemicals. In *Old Yeasts - New Questions*; InTech, 2017.

47. Blount, Z.D. The Unexhausted Potential of *E. Coli*. *Elife* **2015**, *4*, doi:10.7554/eLife.05826.

48. Bhatia, S.K.; Bhatia, R.K.; Choi, Y.-K.; Kan, E.; Kim, Y.-G.; Yang, Y.-H. Biotechnological Potential of Microbial Consortia and Future Perspectives. *Crit Rev Biotechnol* **2018**, *38*, 1209–1229, doi:10.1080/07388551.2018.1471445.

49. Jones, J.A.; Vernacchio, V.R.; Sinkoe, A.L.; Collins, S.M.; Ibrahim, M.H.A.; Lachance, D.M.; Hahn, J.; Koffas, M.A.G. Experimental and Computational Optimization of an *Escherichia Coli* Co-Culture for the Efficient Production of Flavonoids. *Metab Eng* **2016**, *35*, 55–63, doi:10.1016/j.ymben.2016.01.006.

50. Roell, G.W.; Zha, J.; Carr, R.R.; Koffas, M.A.; Fong, S.S.; Tang, Y.J. Engineering Microbial Consortia by Division of Labor. *Microb Cell Fact* **2019**, *18*, 35, doi:10.1186/s12934-019-1083-3.

51. Honjo, H.; Iwasaki, K.; Soma, Y.; Tsuruno, K.; Hamada, H.; Hanai, T. Synthetic Microbial Consortium with Specific Roles Designated by Genetic Circuits for Cooperative Chemical Production. *Metab Eng* **2019**, *55*, 268–275, doi:10.1016/j.ymben.2019.08.007.

52. Willson, M.L.; Burke, L.; Ferguson, T.; Ghersi, D.; Nowak, A.K.; Wilcken, N. Taxanes for Adjuvant Treatment of Early Breast Cancer. *Cochrane Database of Systematic Reviews* **2019**, doi:10.1002/14651858.CD004421.pub3.

53. Ratzka, A.; Vogel, H.; Kliebenstein, D.J.; Mitchell-Olds, T.; Kroymann, J. Disarming the Mustard Oil Bomb. *Proc Natl Acad Sci U S A* **2002**, *99*, 11223–11228, doi:10.1073/pnas.172112899.

54. Tanifuji, R.; Minami, A.; Oguri, H.; Oikawa, H. Total Synthesis of Alkaloids Using Both Chemical and Biochemical Methods. *Nat Prod Rep* **2020**, doi:10.1039/c9np00073a.

55. Santoni, G.; Farfariello, V. TRP Channels and Cancer: New Targets for Diagnosis and Chemotherapy. *Endocr Metab Immune Disord Drug Targets* **2012**, *11*, 54–67, doi:10.2174/187153011794982068.

56. Wang, P.; Zhao, F.-J.; Kopittke, P.M. Engineering Crops without Genome Integration Using Nanotechnology. *Trends Plant Sci* **2019**, *24*, 574–577, doi:10.1016/j.tplants.2019.05.004.

57. Ortiz, R.; Geleta, M.; Gustafsson, C.; Lager, I.; Hofvander, P.; Löfstedt, C.; Cahoon, E.B.; Minina, E.; Bozhkov, P.; Stymne, S. Oil Crops for the Future. *Curr Opin Plant Biol* **2020**, *56*, 181–189, doi:10.1016/j.pbi.2019.12.003.

58. Parray, J.A.; Yaseen Mir, M.; Shameem, N. Plant Genetic Engineering and GM Crops: Merits and Demerits. In *Sustainable Agriculture: Biotechniques in Plant Biology*; Springer, Singapore, 2019; pp. 155–229.

59. Hoshino, A.; Mizuno, T.; Shimizu, K.; Mori, S.; Fukada-Tanaka, S.; Furukawa, K.; Ishiguro, K.; Tanaka, Y.; Iida, S. Generation of Yellow Flowers of the Japanese Morning Glory by Engineering Its Flavonoid Biosynthetic Pathway toward Aurones. *Plant Cell Physiol* **2019**, *60*, 1871–1879, doi:10.1093/pcp/pcz101.

60. Ono, E.; Fukuchi-Mizutani, M.; Nakamura, N.; Fukui, Y.; Yonekura-Sakakibara, K.; Yamaguchi, M.; Nakayama, T.; Tanaka, T.; Kusumi, T.; Tanaka, Y. Yellow Flowers Generated by Expression of the Aurone Biosynthetic Pathway. *Proceedings of the National Academy of Sciences* **2006**, *103*, 11075–11080, doi:10.1073/pnas.0604246103.

61. Pixley, K. V.; Falck-Zepeda, J.B.; Giller, K.E.; Glenna, L.L.; Gould, F.; Mallory-Smith, C.A.; Stelly, D.M.; Stewart, C.N. Genome Editing, Gene Drives, and Synthetic Biology: Will They Contribute to Disease-Resistant Crops, and Who Will Benefit? *Annu Rev Phytopathol* **2019**, *57*, 165–188, doi:10.1146/annurev-phyto-080417-045954.

62. Borrelli, G.; Trono, D. Molecular Approaches to Genetically Improve the Accumulation of Health-Promoting Secondary Metabolites in Staple Crops—A Case Study: The Lipoxygenase-B1 Genes and Regulation of the Carotenoid Content in Pasta Products. *Int J Mol Sci* **2016**, *17*, 1177, doi:10.3390/ijms17071177.

63. Tatsis, E.C.; O'Connor, S.E. New Developments in Engineering Plant Metabolic Pathways. *Curr Opin Biotechnol* **2016**, *42*, 126–132, doi:10.1016/j.copbio.2016.04.012.

64. Wilson, S.A.; Roberts, S.C. Metabolic Engineering Approaches for Production of Biochemicals in Food and Medicinal Plants. *Curr Opin Biotechnol* **2014**, *26*, 174–182, doi:10.1016/j.copbio.2014.01.006.

65. Kootstra, A. Protection from UV-B-Induced DNA Damage by Flavonoids. *Plant Mol Biol* **1994**, *26*, 771–774, doi:10.1007/BF00013762.

66. Sosa, T.; Chaves, N.; Alias, J.C.; Escudero, J.C.; Henao, F.; Gutiérrez-Merino, C. Inhibition of Mouth Skeletal Muscle Relaxation by Flavonoids of *Cistus Ladanifer* L.: A Plant

Defense Mechanism Against Herbivores. *J Chem Ecol* **2004**, *30*, 1087–1101, doi:10.1023/B:JOEC.0000030265.45127.08.

67. Hartwig, U.A.; Joseph, C.M.; Phillips, D.A. Flavonoids Released Naturally from Alfalfa Seeds Enhance Growth Rate of *Rhizobium Meliloti*. *Plant Physiol* **1991**, *95*, 797–803, doi:10.1104/pp.95.3.797.

68. Nix, A.; Paull, C.; Colgrave, M. Flavonoid Profile of the Cotton Plant, *Gossypium Hirsutum*: A Review. *Plants* **2017**, *6*, doi:10.3390/plants6040043.

69. Weseler, A.R.; Bast, A. Masquelier's Grape Seed Extract: From Basic Flavonoid Research to a Well-Characterized Food Supplement with Health Benefits. *Nutr J* 2017, *16*, 1–19.

70. Owens, D.K.; McIntosh, C.A. Biosynthesis and Function of Citrus Glycosylated Flavonoids. In *Recent Advances in Phytochemistry: The Biological Activity of Phytochemicals*; Gang, D., Ed.; Springer New York, 2011; Vol. 41, pp. 67–96 ISBN 978-1-4419-7299-6.

71. Abotaleb, M.; Samuel, S.M.; Varghese, E.; Varghese, S.; Kubatka, P.; Liskova, A.; Büsselberg, D. Flavonoids in Cancer and Apoptosis. *Cancers (Basel)* 2019, *11*.

72. Batra, P.; Sharma, A.K. Anti-Cancer Potential of Flavonoids: Recent Trends and Future Perspectives. *3 Biotech* **2013**, *3*, 439–459, doi:10.1007/s13205-013-0117-5.

73. Tapas, A.R.; Sakarkar, D.M.; Kakde, R.B. Flavonoids as Nutraceuticals: A Review. *Tropical Journal of Pharmaceutical Research* **2008**, *7*, 1089–1099, doi:10.4314/tjpr.v7i3.14693.

74. Ravishankar, D.; Rajora, A.K.; Greco, F.; Osborn, Helen.M.I. Flavonoids as Prospective Compounds for Anti-Cancer Therapy. *Int J Biochem Cell Biol* **2013**, *45*, 2821–2831, doi:10.1016/j.biocel.2013.10.004.

75. Panche, A.N.; Diwan, A.D.; Chandra, S.R. Flavonoids: An Overview. *J Nutr Sci* **2016**, *5*, e47, doi:10.1017/jns.2016.41.

76. Sezer, E.D.; Oktay, L.M.; Karadadaş, E.; Memmedov, H.; Gunel, N.S.; Sözmen, E. Assessing Anticancer Potential of Blueberry Flavonoids, Quercetin, Kaempferol, and Gentisic Acid, Through Oxidative Stress and Apoptosis Parameters on HCT-116 Cells. *J Med Food* **2019**, *22*, 1118–1126, doi:10.1089/jmf.2019.0098.

77. Karak, P. Biological Activities of Flavonoids: An Overview. *Int J Pharm Sci Res* **2019**, *10*, 1567–1574, doi:10.13040/IJPSR.0975-8232.10(4).1567-74.

78. Sharma, A.; Tuli, H.S.; Sharma, A.K. Chemistry and Synthetic Overview of Flavonoids. *Current Aspects of Flavonoids: Their Role in Cancer Treatment* 2019, 23–38.

79. Kumar, S.; Pandey, A.K. Chemistry and Biological Activities of Flavonoids: An Overview. *The Scientific World Journal* **2013**, *2013*, 13, doi:https://doi.org/10.1155/2013/162750.

80. Menezes, J.C.J.M.D.S.; Orlikova, B.; Morceau, F.; Diederich, M. Natural and Synthetic Flavonoids: Structure–Activity Relationship and Chemotherapeutic Potential for the Treatment of Leukemia. *Crit Rev Food Sci Nutr* **2016**, *56*, S4–S28, doi:10.1080/10408398.2015.1074532.

81. Dao, T.T.H.; Linthorst, H.J.M.; Verpoorte, R. Chalcone Synthase and Its Functions in Plant Resistance. *Phytochemistry Reviews* **2011**, *10*, 397–412, doi:10.1007/s11101-011-9211-7.

82. Sudheeran, P.K.; Ovadia, R.; Galsarker, O.; Maoz, I.; Sela, N.; Maurer, D.; Feygenberg, O.; Shamir, M.O.; Alkan, N. Glycosylated Flavonoids: Fruit's Concealed Antifungal Arsenal. *New Phytologist* **2020**, *225*, 1788–1798, doi:10.1111/nph.16251.

83. Vasudevan, U.M.; Lee, E.Y. Flavonoids, Terpenoids, and Polyketide Antibiotics: Role of Glycosylation and Biocatalytic Tactics in Engineering Glycosylation. *Biotechnol Adv* 2020, *41*, 107550.

84. Kim, B.G.; Yang, S.M.; Kim, S.Y.; Cha, M.N.; Ahn, J.-H. Biosynthesis and Production of Glycosylated Flavonoids in *Escherichia Coli*: Current State and Perspectives. *Appl Microbiol Biotechnol* **2015**, *99*, 2979–2988, doi:10.1007/s00253-015-6504-6.

85. Wang, H.; Yang, Y.; Lin, L.; Zhou, W.; Liu, M.; Cheng, K.; Wang, W. Engineering *Saccharomyces Cerevisiae* with the Deletion of Endogenous Glucosidases for the Production of Flavonoid Glucosides. *Microb Cell Fact* **2016**, *15*, 134, doi:10.1186/s12934-016-0535-2.

86. Katsuyama, Y.; Funa, N.; Miyahisa, I.; Horinouchi, S. Synthesis of Unnatural Flavonoids and Stilbenes by Exploiting the Plant Biosynthetic Pathway in *Escherichia Coli*. *Chem Biol* **2007**, *14*, 613–621, doi:10.1016/j.chembiol.2007.05.004.

87. Kim, B.G.; Lee, H.; Ahn, J.H. Biosynthesis of Pinocembrin from Glucose Using Engineered *Escherichia Coli*. *J Microbiol Biotechnol* **2014**, *24*, 1536–1541, doi:10.4014/jmb.1406.06011.

88. Wu, J.; Du, G.; Chen, J.; Zhou, J. Enhancing Flavonoid Production by Systematically Tuning the Central Metabolic Pathways Based on a CRISPR Interference System in *Escherichia Coli*. *Sci Rep* **2015**, *5*, doi:10.1038/srep13477.

89. Kutchan, T.M. Alkaloid Biosynthesis - The Basis of Metabolic Engineering of Medicinal Plants. *Plant Cell* **1995**, *7*, 1059–1070, doi:10.1105/tpc.7.7.1059.

90. Facchini, P.J. Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. *Annu Rev Plant Physiol Plant Mol Biol* **2001**, *52*, 29–66, doi:10.1146/annurev.arplant.52.1.29.

91. Li, Y.; Li, S.; Thodey, K.; Trenchard, I.; Cravens, A.; Smolke, C.D. Complete Biosynthesis of Noscapine and Halogenated Alkaloids in Yeast. *Proceedings of the National Academy of Sciences* **2018**, *115*, doi:10.1073/pnas.1721469115.

92. Matsuura, H.N.; Fett-Neto, A.G. Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action. *Plant Toxins* 2015, 1–15.

93. Waller, G.R.; Nowacki, E.K. The Role of Alkaloids in Plants. In *Alkaloid Biology and Metabolism in Plants*; Springer : Boston, MA, 1978; pp. 143–181.

94. Xi, A.; Li, C.; Zhang, F. Tobacco Plants Can Use Nitrogen Taken up before Mechanical Wounding to Synthesize Nicotine Afterwards. *Plant Signal Behav* **2008**, *3*, 87–90, doi:10.4161/psb.3.2.5121.

95. Guirimand, G.; Courdavault, V.; Lanoue, A.; Mahroug, S.; Guihur, A.; Blanc, N.; Giglioli-Guivarc'h, N.; St-Pierre, B.; Burlat, V. Strictosidine Activation in Apocynaceae: Towards a "Nuclear Time Bomb"? *BMC Plant Biol* **2010**, *10*, 1–20, doi:10.1186/1471-2229-10-182.

96. Guirimand, G.; Guihur, A.; Ginis, O.; Poutrain, P.; Héricourt, F.; Oudin, A.; Lanoue, A.; St-Pierre, B.; Burlat, V.; Courdavault, V. The Subcellular Organization of Strictosidine Biosynthesis in *Catharanthus Roseus* Epidermis Highlights Several Trans-Tonoplast Translocations of Intermediate Metabolites. *FEBS Journal* **2011**, *278*, 749–763, doi:10.1111/j.1742-4658.2010.07994.x.

97. Hagel, J.M.; Facchini, P.J. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. *Plant Cell Physiol* **2013**, *54*, 647–672, doi:10.1093/pcp/pct020.

98. Waller, G.R.; Nowacki, E.K. *Alkaloid Biology and Metabolism in Plants*; Springer New York, NY, 1978; ISBN 978-1-4684-0774-7.

99. Ballout, F.; Habli, Z.; Monzer, A.; Rahal, O.N.; Fatfat, M.; Gali-Muhtasib, H. Anticancer Alkaloids: Molecular Mechanisms and Clinical Manifestations. *Bioactive Natural Products for the Management of Cancer: from Bench to Bedside* 2019, 1–35.

100. Singh, A.; Menéndez-Perdomo, I.M.; Facchini, P.J. Benzylisoquinoline Alkaloid Biosynthesis in Opium Poppy: An Update. *Phytochemistry Reviews* 2019, *18*, 1457–1482.

101. Diamond, A.; Desgagné-Penix, I. Metabolic Engineering for the Production of Plant Isoquinoline Alkaloids. *Plant Biotechnol J* **2016**, *14*, 1319–1328, doi:10.1111/pbi.12494.

102. Beaudoin, G.A.W.; Facchini, P.J. Benzylisoquinoline Alkaloid Biosynthesis in Opium Poppy. *Planta* 2014, *240*, 19–32.

103. Witherup, K.M.; Look, S.A.; Stasko, M.W.; Ghiorzi, T.J.; Muschik, G.M.; Cragg, G.M. *Taxus* Spp. Needles Contain Amounts of Taxol Comparable to the Bark of *Taxus Brevifolia*: Analysis and Isolation. *J Nat Prod* **1990**, *53*, 1249–1255, doi:10.1021/np50071a017.

104. Idrees, M.; Hassan, I.U.; Naeem, M.; Ali, A.; Aftab, T.; Masroor, M.A.K. The Accumulation and Degradation of Alkaloids in *Catharanthus Roseus* Supported by Various External Agents under Different Environmental Conditions. In *Catharanthus Roseus: Current*

Research and Future Prospects; Springer International Publishing: Cham, 2017; pp. 321–329 ISBN 9783319516202.

105. Siemon, T.; Wang, Z.; Bian, G.; Seitz, T.; Ye, Z.; Lu, Y.; Cheng, S.; Ding, Y.; Huang, Y.; Deng, Z.; et al. Semisynthesis of Plant-Derived Englerin A Enabled by Microbe Engineering of Guaia-6,10(14)-Diene as Building Block. *J Am Chem Soc* **2020**, *142*, 2760–2765, doi:10.1021/jacs.9b12940.

106. Galanie, S.; Thodey, K.; Trenchard, I.J.; Interrante, M.F.; Smolke, C.D. Complete Biosynthesis of Opioids in Yeast. *Science (1979)* **2015**, *349*, 1095–1100, doi:10.1126/science.aac9373.

107. Nakagawa, A.; Matsumura, E.; Koyanagi, T.; Katayama, T.; Kawano, N.; Yoshimatsu, K.; Yamamoto, K.; Kumagai, H.; Sato, F.; Minami, H. Total Biosynthesis of Opiates by Stepwise Fermentation Using Engineered *Escherichia Coli*. *Nat Commun* **2016**, *7*, doi:10.1038/ncomms10390.

108. Prevarskaya, N.; Zhang, L.; Barritt, G. TRP Channels in Cancer. *Biochim Biophys Acta Mol Basis Dis* 2007, *1772*, 937–946.

109. Ratnayake, R.; Covell, D.; Ransom, T.T.; Gustafson, K.R.; Beutler, J.A. Englerin a, a Selective Inhibitor of Renal Cancer Cell Growth, from *Phyllanthus Engleri*. *Org Lett* **2009**, *11*, 57–60, doi:10.1021/ol802339w.

110. Rosa, R. da; Schenkel, E.P.; Bernardes, L.S.C. Semisynthetic and Newly Designed Derivatives Based on Natural Chemical Scaffolds: Moving beyond Natural Products to Fight *Trypanosoma Cruzi*. *Phytochemistry Reviews* 2020, *19*, 105–122.

111. Alamgir, A.N.M. Drugs: Their Natural, Synthetic, and Biosynthetic Sources. *Progress in Drug Research* 2017, *73*, 105–123.

112. Alagoz, Y.; Gurkok, T.; Zhang, B.; Unver, T. Manipulating the Biosynthesis of Bioactive Compound Alkaloids for Next-Generation Metabolic Engineering in Opium Poppy Using CRISPR-Cas 9 Genome Editing Technology. *Sci Rep* **2016**, *6*, 30910, doi:10.1038/srep30910.

113. Tang, K.; Pan, Q. Strategies for Enhancing Alkaloids Yield in *Catharanthus Roseus* via Metabolic Engineering Approaches. In *Catharanthus Roseus: Current Research and Future Prospects*; Naeem, M., Aftab, T., Khan, M., Eds.; Springer, Cham, 2017; pp. 1–16 ISBN 9783319516202.

114. Dehghan, E.; Reed, D.W.; Covello, P.S.; Hasanpour, Z.; Palazon, J.; Oksman-Caldentey, K.M.; Ahmadi, F.S. Genetically Engineered Hairy Root Cultures of *Hyoscyamus Senecionis* and *H. Muticus*: Ploidy as a Promising Parameter in the Metabolic Engineering of Tropane Alkaloids. *Plant Cell Rep* **2017**, *36*, 1615–1626, doi:10.1007/s00299-017-2178-0.

115. Sharma, A.; Verma, P.; Mathur, A.; Mathur, A.K. Genetic Engineering Approach Using Early Vinca Alkaloid Biosynthesis Genes Led to Increased Tryptamine and Terpenoid Indole Alkaloids Biosynthesis in Differentiating Cultures of *Catharanthus Roseus*. *Protoplasma* **2018**, *255*, 425–435, doi:10.1007/s00709-017-1151-7.

116. Schweizer, F.; Colinas, M.; Pollier, J.; Moerkercke, A. Van; Bossche, R. Vanden; de Clercq, R.; Goossens, A. An Engineered Combinatorial Module of Transcription Factors Boosts Production of Monoterpenoid Indole Alkaloids in *Catharanthus Roseus*. *Metab Eng* **2018**, *48*, 150–162, doi:10.1016/j.ymben.2018.05.016.

117. Richardson, M. Flavonols and C-Glycosylflavonoids of the Caryophyllales. *Biochem Syst Ecol* **1978**, *6*, 283–286, doi:10.1016/0305-1978(78)90046-7.

118. Li, G.; Meng, X.; Zhu, M.; Li, Z. Research Progress of Betalain in Response to Adverse Stresses and Evolutionary Relationship Compared with Anthocyanin. *Molecules* **2019**, *24*, 3078, doi:10.3390/molecules24173078.

119. Rahimi, P.; Abedimanesh, S.; Mesbah-Namin, S.A.; Ostadrahimi, A. Betalains, the Nature-Inspired Pigments, in Health and Diseases. *Crit Rev Food Sci Nutr* 2019, *59*, 2949–2978.

120. Contreras-Llano, L.E.; Guerrero-Rubio, M.A.; Lozada-Ramírez, J.D.; García-Carmona, F.; Gandía-Herrero, F. First Betalain-Producing Bacteria Break the Exclusive Presence of the Pigments in the Plant Kingdom. *mBio* **2019**, *10*, doi:10.1128/mBio.00345-19.

121. Esquivel, P. *Handbook on Natural Pigments in Food and Beverages. Chapter 4: Betalains*; Schweiggert, R.C.R., Ed.; Elsevier, 2016; ISBN 9780081003718.

122. Miguel, M.G. Betalains in Some Species of the Amaranthaceae Family: A Review. *Antioxidants* 2018, *7*.

123. Polturak, G.; Aharoni, A. Advances and Future Directions in Betalain Metabolic Engineering. *New Phytologist* **2019**, *224*, 1472–1478, doi:10.1111/nph.15973.

124. Timoneda, A.; Feng, T.; Sheehan, H.; Walker-Hale, N.; Pucker, B.; Lopez-Nieves, S.; Guo, R.; Brockington, S. The Evolution of Betalain Biosynthesis in Caryophyllales. *New Phytologist* **2019**, *224*, 71–85, doi:10.1111/nph.15980.

125. Schenck, C.A.; Maeda, H.A. Tyrosine Biosynthesis, Metabolism, and Catabolism in Plants. *Phytochemistry* 2018, *149*, 82–102.

126. Tian, Y.; Fu, X.; Yang, Z.; Wang, B.; Gao, J.; Wang, M.; Xu, J.; Han, H.; Li, Z.; Yao, Q.; et al. Metabolic Engineering of Rice Endosperm for Betanin Biosynthesis. *New Phytologist* **2020**, *225*, 1915–1922, doi:10.1111/nph.16323.

127. Mikkelsen, M.D.; Petersen, B.L.; Olsen, C.E.; Halkier, B.A. Biosynthesis and Metabolic Engineering of Glucosinolates. *Amino Acids* 2002, *22*, 279–295.

128. Halkier, B.A.; Gershenzon, J. Biology and Biochemistry of Glucosinolates. *Annu Rev Plant Biol* **2006**, *57*, 303–333, doi:10.1146/annurev.arplant.57.032905.105228.

129. Blažević, I.; Montaut, S.; Burčul, F.; Olsen, C.E.; Burow, M.; Rollin, P.; Agerbirk, N. Glucosinolate Structural Diversity, Identification, Chemical Synthesis and Metabolism in Plants. *Phytochemistry* 2020, *169*, 112100.

130. Lei, J.; Jayaprakasha, G.K.; Singh, J.; Uckoo, R.; Borrego, E.J.; Finlayson, S.; Kolomiets, M.; Patil, B.S.; Braam, J.; Zhu-Salzman, K. CIRCADIAN CLOCK-ASSOCIATED1 Controls Resistance to Aphids by Altering Indole Glucosinolate Production. *Plant Physiol* **2019**, *181*, 1344–1359, doi:10.1104/pp.19.00676.

131. Klein, A.P.; Sattely, E.S. Biosynthesis of Cabbage Phytoalexins from Indole Glucosinolate. *Proc Natl Acad Sci U S A* **2017**, *114*, 1910–1915, doi:10.1073/pnas.1615625114.

132. Traka, M.; Mithen, R. Glucosinolates, Isothiocyanates and Human Health. *Phytochemistry Reviews* 2009, *8*, 269–282.

133. Sanlier, N.; Salsan, M.G. The Benefits of Brassica Vegetables on Human Health. *J Human Health Res* **2018**, *1*, 104.

134. Barba, F.J.; Nikmaram, N.; Roohinejad, S.; Khelfa, A.; Zhu, Z.; Koubaa, M. Bioavailability of Glucosinolates and Their Breakdown Products: Impact of Processing. *Front Nutr* 2016, *3*.

135. Grubb, C.D.; Abel, S. Glucosinolate Metabolism and Its Control. *Trends Plant Sci* 2006, *11*, 89–100.

136. Zang, Y.X.; Kim, D.H.; Park, B.S.; Hong, S.B. Metabolic Engineering of Indole Glucosinolates in Chinese Cabbage Hairy Roots Expressing *Arabidopsis* CYP79B2, CYP79B3, and CYP83B1. *Biotechnology and Bioprocess Engineering* **2009**, *14*, 467–473, doi:10.1007/s12257-008-0294-y.

137. Wang, C.; Dissing, M.M.; Agerbirk, N.; Crocoll, C.; Halkier, B.A. Characterization of *Arabidopsis* CYP79C1 and CYP79C2 by Glucosinolate Pathway Engineering in *Nicotiana Benthamiana* Shows Substrate Specificity Toward a Range of Aliphatic and Aromatic Amino Acids. *Front Plant Sci* **2020**, *11*, 57, doi:10.3389/fpls.2020.00057.

138. Malka, S.K.; Cheng, Y. Possible Interactions between the Biosynthetic Pathways of Indole Glucosinolate and Auxin. *Front Plant Sci* 2017, *8*.

139. Celenza, J.L.; Quiel, J.A.; Smolen, G.A.; Merrikh, H.; Silvestro, A.R.; Normanly, J.; Bender, J. The *Arabidopsis* ATR1 Myb Transcription Factor Controls Indolic Glucosinolate Homeostasis. *Plant Physiol* **2005**, *137*, 253–262, doi:10.1104/pp.104.054395.

140. Pontrelli, S.; Chiu, T.Y.; Lan, E.I.; Chen, F.Y.H.; Chang, P.; Liao, J.C. *Escherichia Coli* as a Host for Metabolic Engineering. *Metab Eng* **2018**, *50*, 16–46, doi:10.1016/j.ymben.2018.04.008.

141. Petersen, A.; Crocoll, C.; Halkier, B.A. *De Novo* Production of Benzyl Glucosinolate in *Escherichia Coli*. *Metab Eng* **2019**, *54*, 24–34, doi:10.1016/j.ymben.2019.02.004.

142. Liou, C.S.; Sirk, S.J.; Diaz, C.A.C.; Klein, A.P.; Fischer, C.R.; Higginbottom, S.K.; Erez, A.; Donia, M.S.; Sonnenburg, J.L.; Sattely, E.S. A Metabolic Pathway for Activation of Dietary Glucosinolates by a Human Gut Symbiont. *Cell* **2020**, *180*, 717-728.e19, doi:10.1016/j.cell.2020.01.023.

143. Bell, L. The Biosynthesis of Glucosinolates: Insights, Inconsistencies, and Unknowns. In *Annual Plant Reviews online*; Wiley, 2019; pp. 969–1000.

CHAPTER 5. SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

This research undertook an in-depth examination of the effects of a recombinant c-terminal c-myc/6x histidine tag on the activity and specificity of a flavonol-specific 3-hydroxy glucosyltransferase (Cp3GT) isolated from the citrus plant, *Citrus paradisi*. Additionally, the study involved the purification of this enzyme to apparent homogeneity, laying the groundwork for potential crystallographic and other functional analyses. To provide a comprehensive modeling analysis prior to experimental structure determination using crystallography, the AI-driven *in silico* structural modeling tool D-I-TASSER paired with its integrated COFACTOR algorithm to elucidate new insights in Cp3GT flavonol specificity was used. Furthermore, a review of recent advances in the synthetic biology and bioengineering of four classes of plant secondary metabolites was conducted, highlighting recent progress in producing beneficial compounds via microbial expression platforms and plant metabolic engineering. These classes comprise flavonoids, alkaloids, betalains, and glucosinolates.

The exploration of recombinant tag influence on Cp3GT underscored a critical balance between maintaining enzymatic integrity and facilitating purification, with a particular emphasis on potential implications for further analysis. Cp3GT exhibited unaltered activity, substrate affinity, kinetic values, and pH optima with a C-terminal c-myc/6x Histidine tag indicating no change in overall folding of structure and activity. *In silico* modeling predicted that these tags would localize to the periphery of Cp3GT, further supporting the assertion they do not interfere with Cp3GT's core structural and functional attributes.

While the initial objective was to discern the impact of tags for structural analysis via x-ray crystallography, the findings from this research substantiate the claim that cleaving tags prior to subjecting Cp3GT to crystallographic studies may not be imperative. Nevertheless, it is crucial

to acknowledge that the specific conditions facilitating protein crystallization often necessitate empirical determination, given the unique attributes of each protein. Consequently, forthcoming research should prioritize assessing the tangible impact of these tags during the crystallization condition screening phase to ensure the accurate and reliable structural elucidation of Cp3GT.

The multifaceted journey to purify Cp3GT to apparent homogeneity and to delve into its structure and interactions through *in silico* modeling presented a series of challenges and discoveries. The challenges encountered during this process, particularly the persistent degradation of Cp3GT and non-specific binding during purification, necessitated the exploration of various chromatographic techniques. Cobalt-affinity chromatography was strategically selected due to Cp3GT's inactivity in the presence of nickel and cobalt's reduced susceptibility to non-specific binding. A carefully refined low astringency wash successfully removed a majority of contaminants and prepared Cp3GT for additional chromatographic techniques.

Subsequent use of MonoQ anion exchange chromatography was effective in isolating Cp3GT from numerous contaminants, however an unidentified 100kDa protein was still present. Subsequent use of size exclusion chromatography (SEC) successfully purified Cp3GT to apparent homogeneity. Future research may delve into further optimizing chromatographic techniques, exploring alternative or supplementary purification methods, and/or possibly integrating inline concentration systems to uphold protein concentrations post-purification. This could refine the purification process and facilitate subsequent high-resolution structural analyses.

In-depth analysis of Cp3GT and its truncated variants, Cp3GTΔ80 and Cp3GTΔ10, elucidated the enzyme's flavonol 3-O specificity and potential applications, utilizing the advanced structural prediction tools D-I TASSER and COFACTOR. While comparative and docking analyses, especially with VvGT1, have enriched understanding of Cp3GT's stringent

specificity, the truncated variants revealed structural deviations and unique binding dynamics that could impact enzymatic activity. Future research could benefit from exploring alternative *in silico* methodologies, such as molecular dynamics (MD) simulations via GROMACS, to gain insights into real-time protein-ligand interactions, stability of complexes, binding energetics, and potential conformational shifts upon ligand binding, thereby providing a dynamic and temporal perspective on Cp3GT interactions and functionality.

The successful development of a Cp3GT purification protocol has paved the way for obtaining its high-resolution crystal structure in the future. This is crucial for understanding its unique glucosylation of only flavonols at the 3-OH position. Further research exploring the structural and functional divergences in Cp3GT truncated variants, Cp3GTΔ10 and Cp3GTΔ80, could unveil novel functional insights. Leveraging advancements in plant glycosyltransferase structural elucidation, particularly the glycosyltransferase VvGT1 that shares 56.7% identity with Cp3GT, could provide a foundational guide for Cp3GT crystallization efforts, despite the empirical nature of protein crystallization (Appendix C).

X-ray crystallography demands a tailored approach due to the distinct nature of each protein. Information on crystallization conditions for seven plant glycosyltransferases is compiled in Table S1 (Appendix C), providing a valuable reference point for initiating Cp3GT crystallization experiments. Notably, the glycosyltransferase VvGT1, sharing 56.7% identity with Cp3GT, emerges as a particularly relevant model for guiding preliminary crystallization trials for Cp3GT. While it is necessary to empirically determine the optimal conditions for each protein, leveraging the crystallization conditions of structurally and functionally similar proteins, such as VvGT1, provides a strategic starting point. Thus, a screening matrix for Cp3GT could be designed by incorporating and modifying the conditions elucidated for VvGT1. This approach

not only capitalizes on existing data but also pragmatically navigates the empirical landscape of protein crystallization, enhancing the likelihood of obtaining high-quality Cp3GT crystals conducive for high-resolution structural studies.

Building upon the strategic foundation laid by VvGT1, specific parameters from the crystallization conditions of other plant GTs (Appendix C) can be distilled to inform initial trials for Cp3GT. Notably, the recurrent utilization of the hanging drop method and protein concentrations between 5mg/mL and 20mg/mL in the documented cases provide a pragmatic starting point. Additionally, a pH range between 6.5 and 8.0 has been prevalently employed, offering a focused window for initial pH condition explorations for Cp3GT. Although Tris buffers are commonly utilized, their use warrants caution, as evidenced by the case of UGT72B1, where Bis-tris propane was identified in the active site, implying potential interference with substrate docking and glucosylation (Brazier-Hicks et al. 2007). This is further substantiated by consistent observations of significant reductions in Cp3GT activity in the presence of Tris buffers, underscoring the need to use alternative buffers during crystallography experiments (Owens and McIntosh 2009; Daniel et al. 2011; Devaiah et al. 2016).

The insights derived from Cp3GT's purification and *in silico* modeling, when interfaced with these informed crystallization explorations, not only deepens the understanding of its substrate specificity but also propels forward the enzyme engineering endeavors. This integrative approach harbors the potential to unlock novel applications across synthetic biology, medicine, and agriculture, thereby sculpting new trajectories in research and application, underpinned by a nuanced understanding of Cp3GT's structural and functional landscape.

A comprehensive review of the metabolic engineering and synthetic biology of plant natural products was undertaken, focusing on flavonoids, alkaloids, betalains, and glucosinolates.

These plant-derived compounds have a profound impact on human health and find applications across pharmaceuticals and various other industries. The study delved into advanced techniques and methodologies, highlighting the potential of microbial consortia in biotechnological applications. Emphasis was placed on the benefits of these plant natural products, particularly their significance in human health and the intricate metabolic networks associated with their synthesis. With rapid advancements in the field, there's a growing importance attached to improved synthetic biology approaches and scalable production techniques. This work underscored the promising future of plant natural products in the pharmaceutical sector and their potential to bring about therapeutic revolutions.

An in-depth understanding of the structure and function of enzymes, as well as their kinetic requirements, emerges as a cornerstone in the successful application of synthetic biology and bioengineering to produce desired compounds. Enzymes, being the catalysts of metabolic pathways, dictate the efficiency and specificity of biosynthetic processes. Therefore, elucidating the mechanistic and structural nuances of enzymes involved in the synthesis of plant natural products is pivotal. This knowledge not only facilitates the optimization of existing biosynthetic pathways but also paves the way for the innovative design of novel, efficient metabolic routes, thereby enhancing the scalability and sustainability of these production processes in a biotechnological framework.

REFERENCES

Abad-García B, Garmón-Lobato S, Sánchez-Ilárduya MB, Berrueta LA, Gallo B, Vicente F, Alonso-Salces RM. 2014. Polyphenolic contents in Citrus fruit juices: Authenticity assessment. European Food Research and Technology. 238(5):803–818. doi:10.1007/S00217-014-2160-9/METRICS. [accessed 2023 Aug 22]. https://link.springer.com/article/10.1007/s00217-014-2160-9.

Abbas F, Zhou Y, O'Neill Rothenberg D, Alam I, Ke Y, Wang HC. 2023. Aroma Components in Horticultural Crops: Chemical Diversity and Usage of Metabolic Engineering for Industrial Applications. Plants 2023, Vol 12, Page 1748. 12(9):1748. doi:10.3390/PLANTS12091748.

Abotaleb M, Samuel SM, Varghese E, Varghese S, Kubatka P, Liskova A, Büsselberg D. 2019. Flavonoids in cancer and apoptosis. Cancers (Basel). 11(1). doi:10.3390/cancers11010028.

Achnine L, Huhman D V, Farag MA, Sumner LW, Blount JW, Dixon RA. 2005. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*. Plant J. 41(6):875–87. doi:10.1111/j.1365-313X.2005.02344.x.

Addi M, Elbouzidi A, Abid M, Tungmunnithum D, Elamrani A, Hano C. 2022. An Overview of Bioactive Flavonoids from Citrus Fruits. Applied Sciences (Switzerland). 12(1). doi:10.3390/app12010029.

Adepoju O. 2014. Using Site-Directed Mutagenesis to Determine Impact of Amino Acid Substitution on Substrate and Regiospecificity of Grapefruit Flavonol Specific 3-O-

Glucosyltransferase. [Johnson City]: East Tennessee State University. [accessed 2023 Aug 9]. https://dc.etsu.edu/etd/2404.

Adewole SO, Caxton-Martins EA, Ojewole JAO. 2007. Protective effect of quercetin on the morphology of pancreatic β-cells of streptozotocin-treated diabetic rats. African Journal of Traditional, Complementary, and Alternative Medicines. 4(1):64–74. doi:10.4314/AJTCAM.V4I1.31196.

Agati G, Azzarello E, Pollastri S, Tattini M. 2012. Flavonoids as antioxidants in plants: Location and functional significance. Plant Science. 196:67–76. doi:10.1016/J.PLANTSCI.2012.07.014.

Agati G, Tattini M. 2010. Multiple functional roles of flavonoids in photoprotection. New Phytologist. 186(4):786–793. doi:10.1111/J.1469-8137.2010.03269.X. [accessed 2023 Aug 29]. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1469-8137.2010.03269.x.

Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. Science (1979). 330(6000):70–74. doi:10.1126/science.1191652.

Alagoz Y, Gurkok T, Zhang B, Unver T. 2016. Manipulating the Biosynthesis of Bioactive Compound Alkaloids for Next-Generation Metabolic Engineering in Opium Poppy Using CRISPR-Cas 9 Genome Editing Technology. Sci Rep. 6:30910. doi:10.1038/srep30910.

Alam F, Mohammadin K, Shafique Z, Amjad ST, bin Asad MHH. 2022. Citrus flavonoids as potential therapeutic agents: A review. Phytotherapy Research. 36(4):1417–1441. doi:10.1002/PTR.7261.

Alam MA, Subhan N, Rahman MM, Uddin SJ, Reza HM, Sarker SD. 2014. Effect of citrus flavonoids, naringin and naringenin, on metabolic syndrome and their mechanisms of action. Adv Nutr. 5(4):404–417. doi:10.3945/AN.113.005603.

Alamgir ANM. 2017. Drugs: Their natural, synthetic, and biosynthetic sources. Progress in Drug Research. 73:105–123. doi:10.1007/978-3-319-63862-1_4.

Alseekh S, Perez de Souza L, Benina M, Fernie AR. 2020. The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry. 174:112347. doi:10.1016/J.PHYTOCHEM.2020.112347.

Amin K, Tranchimand S, Benvegnu T, Abdel-Razzak Z, Chamieh H. 2021. Glycoside Hydrolases and Glycosyltransferases from Hyperthermophilic Archaea: Insights on Their Characteristics and Applications in Biotechnology. Biomolecules. 11(11). doi:10.3390/BIOM11111557.

Anbualakan K, Urus NQT, Makpol S, Jamil A, Ramli ESM, Pauzi SHM, Muhammad N. 2022. A Scoping Review on the Effects of Carotenoids and Flavonoids on Skin Damage Due to Ultraviolet Radiation. Nutrients 2023, Vol 15, Page 92. 15(1):92. doi:10.3390/NU15010092. https://www.mdpi.com/2072-6643/15/1/92/htm.

Aniszewski T. 2015. Alkaloids: Chemistry, Biology, Ecology, and Applications: Second Edition. Alkaloids: Chemistry, Biology, Ecology, and Applications: Second Edition.:1–475. doi:10.1016/C2011-0-04166-2.

Araghi SH, Amalraj J, Googheri MS, Pyarasani RD, Googheri MSS. 2023. An in-silico study to gain a comprehensive understanding of the effects of glucosylation on quercetin properties. Comput Theor Chem. 1220:113981. doi:10.1016/J.COMPTC.2022.113981.

Arcas MC, Botía JM, Ortuño AM, Del Río JA. 2000. UV irradiation alters the levels of flavonoids involved in the defense mechanism of *Citrus aurantium* fruits against *Penicillium digitatum*. Eur J Plant Pathol. 106(7):617–622. doi:10.1023/A:1008704102446/METRICS.

Arnau J, Lauritzen C, Petersen GE, Pedersen J. 2006. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. Protein Expr Purif. 48(1):1–13. doi:10.1016/J.PEP.2005.12.002.

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues A V, van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin N V, Adams PD, Read RJ, Baker D. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. Science. 373(6557):871–876. doi:10.1126/science.abj8754.

Bai L, Li H. 2021. Protein N-glycosylation and O-mannosylation are catalyzed by two evolutionarily related GT-C glycosyltransferases. Curr Opin Struct Biol. 68:66–73. doi:10.1016/J.SBI.2020.12.009.

Ballout F, Habli Z, Monzer A, Rahal ON, Fatfat M, Gali-Muhtasib H. 2019. Anticancer Alkaloids: Molecular Mechanisms and Clinical Manifestations. Bioactive Natural Products for the Management of Cancer: from Bench to Bedside.:1–35. doi:10.1007/978-981-13-7607-8_1.

Bangar SP, Chaudhary V, Sharma N, Bansal V, Ozogul F, Lorenzo JM. 2022. Kaempferol: A flavonoid with wider biological activities and its applications. Crit Rev Food Sci Nutr.:1–25. doi:10.1080/10408398.2022.2067121.

Barba FJ, Nikmaram N, Roohinejad S, Khelfa A, Zhu Z, Koubaa M. 2016. Bioavailability of Glucosinolates and Their Breakdown Products: Impact of Processing. Front Nutr. 3. doi:10.3389/fnut.2016.00024.

Bashllari R, Molonia MS, Muscarà C, Speciale A, Wilde PJ, Saija A, Cimino F. 2023. Cyanidin-3-O-glucoside protects intestinal epithelial cells from palmitate-induced lipotoxicity. Arch Physiol Biochem. 129(2):379–386. doi:10.1080/13813455.2020.1828480. [accessed 2023 Aug 18]. https://www.tandfonline.com/doi/abs/10.1080/13813455.2020.1828480.

Batra P, Sharma AK. 2013. Anti-cancer potential of flavonoids: recent trends and future perspectives. 3 Biotech. 3(6):439–459. doi:10.1007/s13205-013-0117-5. http://www.ncbi.nlm.nih.gov/pubmed/28324424.

Beaudoin GAW, Facchini PJ. 2014. Benzylisoquinoline alkaloid biosynthesis in opium poppy. Planta. 240(1):19–32. doi:10.1007/s00425-014-2056-8.

Bednar RA, Hadcock JR. 1988. Purification and characterization of chalcone isomerase from soybeans. J Biol Chem. 263(20):9582–9588. doi:10.1016/s0021-9258(19)81556-9.

Bell L. 2019. The Biosynthesis of Glucosinolates: Insights, Inconsistencies, and Unknowns. In: Annual Plant Reviews online. Wiley. p. 969–1000.

Bell MR, Engleka MJ, Malik A, Strickler JE. 2013. To fuse or not to fuse: What is your purpose? Protein Sci. 22(11):1466. doi:10.1002/PRO.2356.

Berhow M, Tisserat B, Kanes K, Vandercook C. 1998. Survey of Phenolic Compounds Produced in Citrus. US Department of Agriculture, Agricultural Research Service, Technical Bulletin No 1856.:158. [accessed 2023 Oct 7]. https://www.ars.usda.gov/ARSUserFiles/34764/MABSurveyCitrus.pdf.

Bhatia SK, Bhatia RK, Choi Y-K, Kan E, Kim Y-G, Yang Y-H. 2018. Biotechnological potential of microbial consortia and future perspectives. Crit Rev Biotechnol. 38(8):1209–1229. doi:10.1080/07388551.2018.1471445.

Bi H, Qu G, Wang S, Zhuang Y, Sun Z, Liu T, Ma Y. 2022. Biosynthesis of a rosavin natural product in *Escherichia coli* by glycosyltransferase rational design and artificial pathway construction. Metab Eng. 69:15–25. doi:10.1016/J.YMBEN.2021.10.010.

Birchfield A. 2019. Preparation of a Flavonol Specific Glucosyltransferase found in Grapefruit and Site-Directed Mutants for Protein Crystallization. [Johnson City]: East Tennessee State University. [accessed 2023 Sep 15]. https://dc.etsu.edu/etd/3523.

Birchfield AS, McIntosh CA. 2020. The Effect of Recombinant Tags on *Citrus paradisi* Flavonol-Specific 3-O Glucosyltransferase Activity. Plants. 9(3):402. doi:10.3390/plants9030402. https://www.mdpi.com/2223-7747/9/3/402.

Blažević I, Montaut S, Burčul F, Olsen CE, Burow M, Rollin P, Agerbirk N. 2020. Glucosinolate structural diversity, identification, chemical synthesis and metabolism in plants. Phytochemistry. 169:112100. doi:10.1016/j.phytochem.2019.112100.

Blount ZD. 2015. The unexhausted potential of *E. coli*. Elife. 4. doi:10.7554/eLife.05826.

Bohl T, Bai L, Li H. 2021. Recent Progress in Structural Studies on the GT-C Superfamily of Protein Glycosyltransferases. Subcell Biochem. 96:259–271. doi:10.1007/978-3-030-58971-4_6/COVER.

Boland MJ, Wong E. 1975. Purification and Kinetic Properties of Chalcone-Flavanone Isomerase from soybeans. Eur J Biochem. 50(2):383–389. doi:10.1111/j.1432-1033.1975.tb09814.x.

Borisova SA, Liu HW. 2010. Characterization of glycosyltransferase DesVII and its auxiliary partner protein DesVIII in the methymycin/picromycin biosynthetic pathway. Biochemistry. 49(37):8071–8084. doi:10.1021/BI1007657.

Borrelli G, Trono D. 2016. Molecular Approaches to Genetically Improve the Accumulation of Health-Promoting Secondary Metabolites in Staple Crops—A Case Study: The Lipoxygenase-B1 Genes and Regulation of the Carotenoid Content in Pasta Products. Int J Mol Sci. 17(7):1177. doi:10.3390/ijms17071177.

Bourgaud F, Gravot A, Milesi S, Gontier E. 2001. Production of plant secondary metabolites: a historical perspective. Plant Science. 161(5):839–851. doi:10.1016/S0168-9452(01)00490-3.

Brader G, Mikkelsen MD, Halkier BA, Tapio Palva E. 2006. Altering glucosinolate profiles modulates disease resistance in plants. The Plant Journal. 46(5):758–767. doi:10.1111/J.1365-313X.2006.02743.X.

Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ, Edwards R. 2007. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proc Natl Acad Sci U S A. 104(51):20238–20243. doi:10.1073/pnas.0706421104.

Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. 2006. Structures and mechanisms of glycosyltransferases. Glycobiology. 16(2):29R-37R. doi:10.1093/GLYCOB/CWJ016. [accessed 2023 Jul 7]. https://dx.doi.org/10.1093/glycob/cwj016.

Britsch L, Grisebach H. 1986. Purification and characterization of (2S)-flavanone 3-hydroxylase from *Petunia hybrida*. Eur J Biochem. 156(3):569–577. doi:10.1111/J.1432-1033.1986.TB09616.X. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/3699024/.

Britsch L, Heller W, Grisebach H. 1981. Conversion of flavanone to flavone, dihydroflavonol and flavonol with an enzyme system from cell cultures of parsley. Zeitschrift fur Naturforschung - Section C Journal of Biosciences. 36(9–10):742–750. [accessed 2023 Oct 7]. https://doi.org/10.1515/znc-1981-9-1009.

Carpenter EP, Beis K, Cameron AD, Iwata S. 2008. Overcoming the challenges of membrane protein crystallography. Curr Opin Struct Biol. 18(5):581–6. doi:10.1016/j.sbi.2008.07.001.

Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. 2007. His-tag impact on structure. Acta Crystallogr D Biol Crystallogr. 63(Pt 3):295–301. doi:10.1107/S0907444906052024. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/17327666/.

Celenza JL, Quiel JA, Smolen GA, Merrikh H, Silvestro AR, Normanly J, Bender J. 2005. The *Arabidopsis* ATR1 Myb transcription factor controls indolic glucosinolate homeostasis. Plant Physiol. 137(1):253–262. doi:10.1104/pp.104.054395.

Chae TU, Choi SY, Kim JW, Ko Y-S, Lee SY. 2017. Recent advances in systems metabolic engineering tools and strategies. Curr Opin Biotechnol. 47:67–82. doi:10.1016/j.copbio.2017.06.007.

Chang CJ, Tzeng TF, Liou SS, Chang YS, Liu IM. 2012. Myricetin Increases Hepatic Peroxisome Proliferator-Activated Receptor α Protein Expression and Decreases Plasma Lipids and Adiposity in Rats. Evidence-based Complementary and Alternative Medicine. 2012. doi:10.1155/2012/787152.

Chang TS, Wang TY, Chiang CM, Lin YJ, Chen HL, Wu YW, Ting HJ, Wu JY. 2021. Biotransformation of celastrol to a novel, well-soluble, low-toxic and anti-oxidative celastrol-29-O-β-glucoside by *Bacillus* glycosyltransferases. J Biosci Bioeng. 131(2):176–182. doi:10.1016/J.JBIOSC.2020.09.017.

Chen C, Wu T, Guo Z, Cheng J. 2021. Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction. Proteins. 89(6):697–707. doi:10.1002/prot.26052.

Chen F, Tholl D, Bohlmann J, Pichersky E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 66(1):212–229. doi:10.1111/J.1365-313X.2011.04520.X. [accessed 2023 Apr 7]. https://pubmed.ncbi.nlm.nih.gov/21443633/.

Chen H, Liu Y, Ren X, Wang J, Zhu L, Lu Y, Chen X. 2022. Engineering of Cyclodextrin Glycosyltransferase through a Size/Polarity Guided Triple-Code Strategy with Enhanced a-Glycosyl Hesperidin Synthesis Ability. Appl Environ Microbiol. 88(17).

doi:10.1128/AEM.01027-22/SUPPL_FILE/AEM.01027-22-S0001.PDF. [accessed 2023 Aug 6]. https://journals.asm.org/doi/10.1128/aem.01027-22.

Chen K, Hu Z-M, Song W, Wang Z-L, He J-B, Shi X-M, Cui Q-H, Qiao X, Ye M. 2019. Diversity of O-Glycosyltransferases Contributes to the Biosynthesis of Flavonoid and Triterpenoid Glycosides in *Glycyrrhiza uralensis*. ACS Synth Biol. 8(8):1858–1866. doi:10.1021/acssynbio.9b00171.

Chen K, Kurgan L. 2007. PFRES: protein fold classification by using evolutionary information and predicted secondary structure. Bioinformatics. 23(21):2843–50. doi:10.1093/bioinformatics/btm475.

Chen M, Song F, Qin Y, Han S, Rao Y, Liang S, Lin Y. 2022. Improving Thermostability and Catalytic Activity of Glycosyltransferase from *Panax ginseng* by Semi-Rational Design for Rebaudioside D Synthesis. Front Bioeng Biotechnol. 10:884898. doi:10.3389/FBIOE.2022.884898/BIBTEX.

Chen W, Yao J, Meng J, Han W, Tao Y, Chen Y, Guo Y, Shi G, He Y, Jin J-M, Tang S-Y. 2019. Promiscuous enzymatic activity-aided multiple-pathway network design for metabolic flux rearrangement in hydroxytyrosol biosynthesis. Nat Commun. 10(1):960. doi:10.1038/s41467-019-08781-2.

Chen YY, Xu ZS, Xiong AS. 2016. Identification and Characterization of DcUSAGT1, a UDP-Glucose: Sinapic Acid Glucosyltransferase from Purple Carrot Taproots. PLoS One. 11(5). doi:10.1371/JOURNAL.PONE.0154938. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/27171142/.

Chen Z, Li Y, Yuan Q. 2015. Study the effect of His-tag on chondroitinase ABC I based on characterization of enzyme. Int J Biol Macromol. 78:96–101. doi:10.1016/J.IJBIOMAC.2015.03.068. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/25861742/.

Chiocchio I, Mandrone M, Tomasi P, Marincich L, Poli F. 2021. Plant Secondary Metabolites: An Opportunity for Circular Economy. Molecules 2021, Vol 26, Page 495. 26(2):495. doi:10.3390/MOLECULES26020495. [accessed 2023 Aug 3]. https://www.mdpi.com/1420-3049/26/2/495/htm.

Choi HN, Kang MJ, Lee SJ, Kim JI. 2014. Ameliorative effect of myricetin on insulin resistance in mice fed a high-fat, high-sucrose diet. Nutr Res Pract. 8(5):544–549. doi:10.4162/NRP.2014.8.5.544. [accessed 2023 Aug 3]. https://synapse.koreamed.org/articles/1051413.

Chou PY, Fasman GD. 1974. Prediction of protein conformation. Biochemistry. 13(2):222–45. doi:10.1021/bi00699a002.

Ciesla WP, Bobak DA. 1998. *Clostridium difficile* toxins A and B are cation-dependent UDP-glucose hydrolases with differing catalytic activities. J Biol Chem. 273(26):16021–16026. doi:10.1074/JBC.273.26.16021. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/9632652/.

Clemensen AK, Provenza FD, Hendrickson JR, Grusak MA. 2020. Ecological Implications of Plant Secondary Metabolites - Phytochemical Diversity Can Enhance Agricultural Sustainability. Front Sustain Food Syst. 4:547826. doi:10.3389/FSUFS.2020.547826/BIBTEX.

Contreras-Llano LE, Guerrero-Rubio MA, Lozada-Ramírez JD, García-Carmona F, Gandía-Herrero F. 2019. First Betalain-Producing Bacteria Break the Exclusive Presence of the Pigments in the Plant Kingdom. Søgaard-Andersen L, editor. mBio. 10(2). doi:10.1128/mBio.00345-19. http://mbio.asm.org/lookup/doi/10.1128/mBio.00345-19.

Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. 2008. Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci. 17(12):2070–2079. doi:10.1110/PS.037564.108. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/18780816/.

Cragg GM, Newman DJ. 2013. Natural products: a continuing source of novel drug leads. Biochim Biophys Acta. 1830(6):3670–3695. doi:10.1016/J.BBAGEN.2013.02.008. [accessed 2023 Apr 7]. https://pubmed.ncbi.nlm.nih.gov/23428572/.

Crozier A, Clifford MN, Ashihara H, editors. 2006. Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet. Blackwell Publishing.

Cuff JA, Barton GJ. 2000. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins. 40(3):502–11. doi:10.1002/1097-0134(20000815)40:3<502::aid-prot170>3.0.co;2-q.

Cui L, Yao S, Dai X, Yin Q, Liu Y, Jiang X, Wu Y, Qian Y, Pang Y, Gao L, Xia T. 2016. Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). J Exp Bot. 67(8):2285–2297. doi:10.1093/JXB/ERW053. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/26941235/.

Cura V, Gangloff M, Eiler S, Moras D, Ruff M. 2008. Cleaved thioredoxin fusion protein enables the crystallization of poorly soluble ERalpha in complex with synthetic ligands. Acta Crystallogr Sect F Struct Biol Cryst Commun. 64(Pt 1):54–57. doi:10.1107/S1744309107066444. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/18097104/.

Dai J, Mumper RJ. 2010. Plant phenolics: extraction, analysis and their antioxidant and anticancer properties. Molecules. 15(10):7313–52. doi:10.3390/molecules15107313.

Daniel JJ, Owens DK, McIntosh CA. 2011. Secondary product glucosyltransferase and putative glucosyltransferase expression during *Citrus paradisi* (c.v. Duncan) growth and development. Phytochemistry. 72(14–15):1732–1738. doi:10.1016/j.phytochem.2011.04.024. http://dx.doi.org/10.1016/j.phytochem.2011.04.024.

Dao TTH, Linthorst HJM, Verpoorte R. 2011. Chalcone synthase and its functions in plant resistance. Phytochemistry Reviews. 10(3):397–412. doi:10.1007/s11101-011-9211-7.

Daryanavard H, Postiglione AE, Mühlemann JK, Muday GK. 2023. Flavonols modulate plant development, signaling, and stress responses. Curr Opin Plant Biol. 72:102350. doi:10.1016/J.PBI.2023.102350.

Dassault Systèmes. 2021. BIOVIA Discovery Studio v21.2.0.20298. San Diego: Dassault Systèmes.

David AVA, Arulmoli R, Parasuraman S. 2016. Overviews of Biological Importance of Quercetin: A Bioactive Flavonoid. Pharmacogn Rev. 10(20):84. doi:10.4103/0973-7847.194044. /pmc/articles/PMC5214562/.

Dehghan E, Reed DW, Covello PS, Hasanpour Z, Palazon J, Oksman-Caldentey KM, Ahmadi FS. 2017. Genetically engineered hairy root cultures of *Hyoscyamus senecionis* and *H. muticus*: ploidy as a promising parameter in the metabolic engineering of tropane alkaloids. Plant Cell Rep. 36(10):1615–1626. doi:10.1007/s00299-017-2178-0.

Denish PR, Fenger JA, Powers R, Sigurdson GT, Grisanti L, Guggenheim KG, Laporte S, Li J, Kondo T, Magistrato A, Moloney MP, Riley M, Rusishvili M, Ahmadiani N, Baroni S, Dangles O, Giusti M, Collins TM, Didzbalis J, Yoshida K, Siegel JB, Robbins RJ. 2021. Discovery of a natural cyan blue: A unique food-sourced anthocyanin could replace synthetic brilliant blue. Sci Adv. 7(15). doi:10.1126/SCIADV.ABE7871/SUPPL_FILE/ABE7871_SM.PDF. [accessed 2023 Aug 22]. https://www.science.org/doi/10.1126/sciadv.abe7871.

Derewenda U, Tarricone C, Choi WC, Cooper DR, Lukasik S, Perrina F, Tripathy A, Kim MH, Cafiso DS, Musacchio A, Derewenda ZS. 2007. The structure of the coiled-coil domain of Ndel1 and the basis of its interaction with Lis1, the causal protein of Miller-Dieker lissencephaly. Structure. 15(11):1467–1481. doi:10.1016/J.STR.2007.09.015. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/17997972/.

Devaiah SP, Owens DK, Sibhatu MB, Sarkar TR, Strong CL, Mallampalli VKPS, Asiago J, Cooke J, Kiser S, Lin Z, Wamucho A, Hayford D, Williams BE, Loftis P, Berhow M, Pike LM, McIntosh CA. 2016. Identification, Recombinant Expression, and Biochemical Analysis of Putative Secondary Product Glucosyltransferases from *Citrus paradisi*. J Agric Food Chem. 64(9):1957–1969. doi:10.1021/acs.jafc.5b05430.

Devaiah SP, Tolliver BM, Zhang C, Owens DK, McIntosh CA. 2018. Mutational analysis of substrate specificity in a *Citrus paradisi* flavonol 3-O-glucosyltransferase. J Plant Biochem Biotechnol. 27(1):13–27. doi:10.1007/s13562-017-0411-0.

Devi R, Kaur T, Guleria G, Rana KL, Kour D, Yadav N, Yadav AN, Saxena AK. 2020. Fungal secondary metabolites and their biotechnological applications for human health. New and Future Developments in Microbial Biotechnology and Bioengineering: Trends of Microbial Biotechnology for Sustainable Agriculture and Biomedicine Systems: Perspectives for Human Health.:147–161. doi:10.1016/B978-0-12-820528-0.00010-7.

Devika MS, Vysakh A, Vijeesh V, Jisha N, Ruby PM. 2022. Flavonoid Glycoside Diosmin Induces Apoptosis and Cell Cycle Arrest in DLD-1 Human Colon Cancer Cell Line. Journal of Biologically Active Products from Nature. 12(3):232–242. doi:10.1080/22311866.2022.2081608. [accessed 2023 Aug 18]. https://www.tandfonline.com/doi/abs/10.1080/22311866.2022.2081608.

Dey P. 2019. Gut microbiota in phytopharmacology: A comprehensive overview of concepts, reciprocal interactions, biotransformations and mode of actions. Pharmacol Res. 147:104367. doi:10.1016/J.PHRS.2019.104367.

Diamond A, Desgagné-Penix I. 2016. Metabolic engineering for the production of plant isoquinoline alkaloids. Plant Biotechnol J. 14(6):1319–1328. doi:10.1111/pbi.12494. http://doi.wiley.com/10.1111/pbi.12494.

Dias MC, Pinto DCGA, Silva AMS, Giovinazzo G, Gerardi C, Mosca L. 2021. Plant Flavonoids: Chemical Characteristics and Biological Activity. Molecules 2021, Vol 26, Page 5377.

26(17):5377. doi:10.3390/MOLECULES26175377. https://www.mdpi.com/1420-3049/26/17/5377/htm.

Dickson JMJ, Lee WJ, Shepherd PR, Buchanan CM. 2013. Enzyme activity effects of N-terminal His-tag attached to catalytic sub-unit of phosphoinositide-3-kinase. Biosci Rep. 33(6):857–863. doi:10.1042/BSR20130075. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/23968392/.

Ding M, Yan H, Li L, Zhai F, Shang L, Yin Z, Yuan Y. 2014. Biosynthesis of Taxadiene in *Saccharomyces cerevisiae*: Selection of Geranylgeranyl Diphosphate Synthase Directed by a Computer-Aided Docking Strategy. PLoS One. 9(10):e109348. doi:10.1371/journal.pone.0109348.

Divekar PA, Narayana S, Divekar BA, Kumar R, Gadratagi BG, Ray A, Singh Achuit Kumar, Rani V, Singh V, Singh Akhilesh Kumar, Kumar A, Singh RP, Meena RS, Behera TK. 2022. Plant Secondary Metabolites as Defense Tools against Herbivores for Sustainable Crop Protection. Int J Mol Sci. 23(5):2690. doi:10.3390/IJMS23052690/S1. https://www.mdpi.com/1422-0067/23/5/2690/htm.

Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 50(D1):D571–D577. doi:10.1093/NAR/GKAB1045. [accessed 2023 Jul 7]. https://doi.org/10.1093/nar/gkab1045.

Duffey SS, Isman MB. 1981. Inhibition of insect larval growth by phenolics in glandular trichomes of tomato leaves. Experientia. 37(6):574–576. doi:10.1007/BF01990057.

Ehrenworth AM, Peralta-Yahya P. 2017. Accelerating the semisynthesis of alkaloid-based drugs through metabolic engineering. Nat Chem Biol. 13(3):249–258. doi:10.1038/nchembio.2308.

Elliger CA, Chan BC, Waiss AC. 1980. Flavonoids as larval growth inhibitors - Structural factors governing toxicity. Naturwissenschaften. 67(7):358–360. doi:10.1007/BF01106595.

Erb M, Kliebenstein DJ. 2020. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy1[OPEN]. Plant Physiol. 184(1):39–52. doi:10.1104/PP.20.00433.

Esquivel P. 2016. Handbook on Natural Pigments in Food and Beverages. Chapter 4: Betalains. Schweiggert RCR, editor. Elsevier. https://www.sciencedirect.com/science/article/pii/B9780081003718000038.

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2021. Protein complex prediction with AlphaFold-Multimer. bioRxiv.:2021.10.04.463034. doi:10.1101/2021.10.04.463034. https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1.

Facchini Peter J. 2001. Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. Annu Rev Plant Physiol Plant Mol Biol. 52:29–66. doi:10.1146/ANNUREV.ARPLANT.52.1.29. [accessed 2023 Apr 7]. https://pubmed.ncbi.nlm.nih.gov/11337391/.

Facchini Peter J. 2001. Alkaloid Biosynthesis in Plants: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. Annu Rev Plant Physiol Plant Mol Biol. 52(1):29–66. doi:10.1146/annurev.arplant.52.1.29.

Falcone Ferreyra ML, Rius SP, Casati P. 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front Plant Sci. 3:222. doi:10.3389/fpls.2012.00222.

FAO. 2021. Citrus Fruit Statistical Compendium 2020. Rome. [accessed 2023 Oct 7]. https://www.fao.org/3/cb6492en/cb6492en.pdf.

Ferreyra MLF, Rius SP, Casati P. 2012. Flavonoids: Biosynthesis, biological functions, and biotechnological applications. Front Plant Sci. 3(SEP):34352. doi:10.3389/FPLS.2012.00222/BIBTEX.

Fini A, Guidi L, Ferrini F, Brunetti C, Di Ferdinando M, Biricolti S, Pollastri S, Calamai L, Tattini M. 2012. Drought stress has contrasting effects on antioxidant enzymes activity and phenylpropanoid biosynthesis in *Fraxinus ornus* leaves: An excess light stress affair? J Plant Physiol. 169(10):929–939. doi:10.1016/J.JPLPH.2012.02.014.

Fitzgerald K, Holliger P, Winter G. 1997. Improved tumor targeting by disulphide stabilized diabodies expressed in *Pichia pastoris*. Protein Eng. 10(10):1221–5. doi:10.1093/protein/10.10.1221.

Ford CM, Boss PK, Hæj PB. 1998. Cloning and characterization of *Vitis vinifera* UDP-glucose. Flavonoid 3- O-Glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze- 1 locus that may primarily serve to glucosylate anthocyanidins in vivo. Journal of Biological Chemistry. 273(15):9224–9233. doi:10.1074/jbc.273.15.9224.

Forkmann G, Eidenberger T, Wurst F. 1992. Flavonol synthase activity and the regulation of flavonol and anthocyanin biosynthesis during flower development in *Dianthus caryophyllus* L. (carnation). Zeitschrift fur Naturforschung - Section C Journal of Biosciences. 47(7–8):553–560. doi:10.1515/ZNC-1992-7-811/MACHINEREADABLECITATION/RIS. [accessed 2023 Aug 5]. https://www.degruyter.com/document/doi/10.1515/znc-1992-7-811/html?lang=en.

Furqan BRN, Akhmaloka. 2020. Heterologous expression and characterization of thermostable lipase (Lk1) in *Pichia pastoris* GS115. Biocatal Agric Biotechnol. 23:101448. doi:10.1016/J.BCAB.2019.101448.

Gachon CMM, Langlois-Meurinne M, Saindrenan P. 2005. Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends Plant Sci. 10(11):542–9. doi:10.1016/j.tplants.2005.09.007.

Galanie S, Thodey K, Trenchard IJ, Interrante MF, Smolke CD. 2015. Complete biosynthesis of opioids in yeast. Science (1979). 349(6252):1095–1100. doi:10.1126/science.aac9373.

Garnier J, Osguthorpe DJ, Robson B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol. 120(1):97–120. doi:10.1016/0022-2836(78)90297-8.

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino M, Mattanovich D, Villaverde A. 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact. 7.

doi:10.1186/1475-2859-7-11. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/18394160/.

Gharabli H, Della Gala V, Welner DH. 2023. The function of UDP-glycosyltransferases in plants and their possible use in crop protection. Biotechnol Adv. 67:108182. doi:10.1016/J.BIOTECHADV.2023.108182.

Grewal PS, Modavi C, Russ ZN, Harris NC, Dueber JE. 2018. Bioproduction of a betalain color palette in *Saccharomyces cerevisiae*. Metab Eng. 45:180–188. doi:10.1016/j.ymben.2017.12.008.

Grimsey E, Collis DWP, Mikut R, Hilpert K. 2020. The effect of lipidation and glycosylation on short cationic antimicrobial peptides. Biochimica et Biophysica Acta (BBA) - Biomembranes. 1862(8):183195. doi:10.1016/J.BBAMEM.2020.183195.

Grubb CD, Abel S. 2006. Glucosinolate metabolism and its control. Trends Plant Sci. 11(2):89–100. doi:10.1016/j.tplants.2005.12.006.

Gruyer S, Legin E, Bliard C, Ball S, Duchiron F. 2002. The Endopolysaccharide Metabolism of the Hyperthermophilic Archeon *Thermococcus hydrothermalis*: Polymer Structure and Biosynthesis. Current Microbiology 2002 44:3. 44(3):206–211. doi:10.1007/S00284-001-0029-1. [accessed 2023 Aug 6]. https://link.springer.com/article/10.1007/s00284-001-0029-1.

Guidi C, Biarnés X, Planas A, De Mey M. 2023. Controlled processivity in glycosyltransferases: A way to expand the enzymatic toolbox. Biotechnol Adv. 63:108081. doi:10.1016/J.BIOTECHADV.2022.108081.

Guirimand G, Courdavault V, Lanoue A, Mahroug S, Guihur A, Blanc N, Giglioli-Guivarc'h N, St-Pierre B, Burlat V. 2010. Strictosidine activation in Apocynaceae: Towards a "nuclear time bomb"? BMC Plant Biol. 10(1):1–20. doi:10.1186/1471-2229-10-182.

Guirimand G, Guihur A, Ginis O, Poutrain P, Héricourt F, Oudin A, Lanoue A, St-Pierre B, Burlat V, Courdavault V. 2011. The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. FEBS Journal. 278(5):749–763. doi:10.1111/j.1742-4658.2010.07994.x.

Guo J, Meng Y, Zhao Y, Hu Y, Ren D, Yang X. 2015. Myricetin derived from *Hovenia dulcis* Thunb. ameliorates vascular endothelial dysfunction and liver injury in high choline-fed mice. Food Funct. 6(5):1620–1634. doi:10.1039/C4FO01073F. [accessed 2023 Aug 3]. https://pubs.rsc.org/en/content/articlehtml/2015/fo/c4fo01073f.

Guzzo A V. 1965. The influence of amino-acid sequence on protein structure. Biophys J. 5(6):809–22. doi:10.1016/S0006-3495(65)86753-4.

Haeuptle MA, Hennet T. 2009. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. Hum Mutat. 30(12):1628–1641. doi:10.1002/HUMU.21126. [accessed 2023 Jul 7]. https://onlinelibrary.wiley.com/doi/full/10.1002/humu.21126.

Hagel JM, Facchini PJ. 2013. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. Plant Cell Physiol. 54(5):647–672. doi:10.1093/pcp/pct020.

Hahlbrock K, Grisebach H. 1970. Formation of coenzyme a esters of cinnamic acids with an enzyme preparation from cell suspension cultures of parsley. FEBS Lett. 11(1):62–64.

doi:10.1016/0014-5793(70)80492-6. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/11945450/.

Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. Annu Rev Plant Biol. 57(1):303–333. doi:10.1146/annurev.arplant.57.032905.105228.

Hamberger B, Hahlbrock K. 2004. The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc Natl Acad Sci U S A. 101(7):2209–2214. doi:10.1073/PNAS.0307307101. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/14769935/.

Harborne JB. 1995. The Flavonoids: Advances in Research Since 1986. J Chem Educ. 72(3):A73. doi:10.1021/ED072PA73.11. [accessed 2023 Aug 5]. https://pubs.acs.org/doi/epdf/10.1021/ed072pA73.11.

Harborne JB, Williams CA. 2000. Advances in flavonoid research since 1992. Phytochemistry. 55(6):481–504. doi:10.1016/S0031-9422(00)00235-1. [accessed 2023 Apr 8]. https://pubmed.ncbi.nlm.nih.gov/11130659/.

Haroth S, Feussner K, Kelly AA, Zienkiewicz K, Shaikhqasem A, Herrfurth C, Feussner I. 2019. The glycosyltransferase UGT76E1 significantly contributes to 12-O-glucopyranosyl-jasmonic acid formation in wounded *Arabidopsis thaliana* leaves. J Biol Chem. 294(25):9858–9872. doi:10.1074/JBC.RA119.007600. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/31072871/.

Hartwig UA, Joseph CM, Phillips DA. 1991. Flavonoids released naturally from alfalfa seeds enhance growth rate of *Rhizobium meliloti*. Plant Physiol. 95(3):797–803. doi:10.1104/pp.95.3.797.

Heinrich M, Mah J, Amirkia V. 2021a. Alkaloids Used as Medicines: Structural Phytochemistry Meets Biodiversity—An Update and Forward Look. Molecules. 26(7). doi:10.3390/MOLECULES26071836. [accessed 2023 Apr 7]. /pmc/articles/PMC8036335/.

Heinrich M, Mah J, Amirkia V. 2021b. Alkaloids Used as Medicines: Structural Phytochemistry Meets Biodiversity—An Update and Forward Look. Molecules 2021, Vol 26, Page 1836. 26(7):1836. doi:10.3390/MOLECULES26071836. [accessed 2023 Jul 7]. https://www.mdpi.com/1420-3049/26/7/1836/htm.

Heinrich Michael, Barnes Joanne, Prieto-Garcia Jose, Gibbons Simon, Williamson EM. 2018. Fundamentals of Pharmacognosy and Phytotherapy. Elsevier.

Heller Werner, Britsch L, Forkmann G, Grisebach H. 1985. Leucoanthocyanidins as intermediates in anthocyanidin biosynthesis in flowers of *Matthiola incana* R. Br. Planta. 163(2):191–196. doi:10.1007/BF00393505. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/24249337/.

Heller W., Forkmann G, Britsch L, Grisebach H. 1985. Enzymatic reduction of (+)-dihydroflavonols to flavan-3,4-cis-diols with flower extracts from *Matthiola incana* and its role in anthocyanin biosynthesis. Planta. 165(2):284–287. doi:10.1007/BF00395052. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/24241054/.

Hijaz FM, Manthey JA, Folimonova SY, Davis CL, Jones SE, Reyes-De-Corcuera JI. 2013. An HPLC-MS characterization of the changes in sweet orange leaf metabolite profile following infection by the bacterial pathogen Candidatus *Liberibacter asiaticus* . PLoS One. 8(11). doi:10.1371/JOURNAL.PONE.0079485.

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-glucosyltransferase from *Clitoria ternatea*. Protein Science. 24(3):395–407. doi:10.1002/pro.2630.

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*. J Synchrotron Radiat. 20(6):894–898. doi:10.1107/S0909049513020712.

Hochuli E, Bannwarth W, Dobeli H, Gentzi R, Stuber D. 1988. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. Bio/Technology 1988 6:11. 6(11):1321–1325. doi:10.1038/nbt1188-1321. [accessed 2023 Aug 9]. https://www.nature.com/articles/nbt1188-1321.

Hong R, Sun Y, Su L, Gu L, Wang F, Wu J. 2019. High-level expression of *Humicola insolens* cutinase in *Pichia pastoris* without carbon starvation and its use in cotton fabric bioscouring. J Biotechnol. 304:10–15. doi:10.1016/J.JBIOTEC.2019.07.011. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/31400343/.

Honjo H, Iwasaki K, Soma Y, Tsuruno K, Hamada H, Hanai T. 2019. Synthetic microbial consortium with specific roles designated by genetic circuits for cooperative chemical production. Metab Eng. 55:268–275. doi:10.1016/j.ymben.2019.08.007.

Horcajada MN, Habauzit V, Trzeciakiewicz A, Morand C, Gil-Izquierdo A, Mardon J, Lebecque P, Davicco MJ, Chee WSS, Coxam V, Offord E. 2008. Hesperidin inhibits ovariectomized-induced osteopenia and shows differential effects on bone mass and strength in young and adult intact rats. J Appl Physiol (1985). 104(3):648–654. doi:10.1152/JAPPLPHYSIOL.00441.2007. [accessed 2023 Aug 4]. https://pubmed.ncbi.nlm.nih.gov/18174393/.

Hoshino A, Mizuno T, Shimizu K, Mori S, Fukada-Tanaka S, Furukawa K, Ishiguro K, Tanaka Y, Iida S. 2019. Generation of Yellow Flowers of the Japanese Morning Glory by Engineering Its Flavonoid Biosynthetic Pathway toward Aurones. Plant Cell Physiol. 60(8):1871–1879. doi:10.1093/pcp/pcz101.

Hotti H, Rischer H. 2017. The killer of Socrates: Coniine and Related Alkaloids in the Plant Kingdom. Molecules: A Journal of Synthetic Chemistry and Natural Product Chemistry. 22(11). doi:10.3390/MOLECULES22111962. /pmc/articles/PMC6150177/.

Hou L, Zhou B, Yang L, Liu ZL. 2004. Inhibition of human low density lipoprotein oxidation by flavonols and their glycosides. Chem Phys Lipids. 129(2):209–219. doi:10.1016/j.chemphyslip.2004.02.001. [accessed 2023 Aug 3]. https://pubmed.ncbi.nlm.nih.gov/15081861/.

Hu J, Dong M, Tang Y-X, Zhang G-J. 2023. Improving protein-protein interaction site prediction using deep residual neural network. Anal Biochem. 670:115132. doi:10.1016/j.ab.2023.115132.

Huang J, Zhou Y, Zhong X, Su F, Xu L. 2022. Effects of Vitexin, a Natural Flavonoid Glycoside, on the Proliferation, Invasion, and Apoptosis of Human U251 Glioblastoma Cells. Oxid Med Cell Longev. 2022. doi:10.1155/2022/3129155.

Huang Y, Li D, Wang J, Cai Y, Dai Z, Jiang D, Liu C. 2019. GuUGT, a glycosyltransferase from *Glycyrrhiza uralensis*, exhibits glycyrrhetinic acid 3- and 30-O-glycosylation. R Soc Open Sci. 6(10). doi:10.1098/RSOS.191121. [accessed 2023 Aug 18]. https://royalsocietypublishing.org/doi/10.1098/rsos.191121.

Hughes J, Hughes MA. 1994. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons. DNA Seq. 5(1):41–9. doi:10.3109/10425179409039703.

Hyung Ko J, Gyu Kim B, Joong-Hoon A. 2006. Glycosylation of flavonoids with a glycosyltransferase from *Bacillus cereus*. FEMS Microbiol Lett. 258(2):263–268. doi:10.1111/J.1574-6968.2006.00226.X. [accessed 2023 Aug 5]. https://dx.doi.org/10.1111/j.1574-6968.2006.00226.x.

Ichiyanagi T, Kashiwada Y, Nashimoto M, Terahara N. 2022. Addition of Malonyl Groups Enhances Intestinal Absorption of Anthocyanins Derived from Edible Red Chrysanthemum (*Dendranthema grandiflorum*) in Rats. BPB Reports. 5(1):9–15. https://doi.org/10.1248/bpbreports.5.1_9

Idrees M, Hassan IU, Naeem M, Ali A, Aftab T, Masroor MAK. 2017. The accumulation and degradation of alkaloids in *Catharanthus roseus* supported by various external agents under different environmental conditions. In: Catharanthus Roseus: Current Research and Future Prospects. Cham: Springer International Publishing. p. 321–329.

Isah T, Umar S, Mujib A, Sharma MP, Rajasekharan PE, Zafar N, Frukh A. 2018. Secondary metabolism of pharmaceuticals in the plant in vitro cultures: strategies, approaches, and limitations to achieving higher yield. Plant Cell, Tissue and Organ Culture (PCTOC). 132(2):239–265. doi:10.1007/s11240-017-1332-2.

Jahic M, Gustavsson M, Jansen AK, Martinelle M, Enfors SO. 2003. Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes. J Biotechnol. 102(1):45–53. doi:10.1016/S0168-1656(03)00003-8.

Jan R, Khan M, Asaf S, Lubna, Asif S, Kim KM. 2022. Bioactivity and Therapeutic Potential of Kaempferol and Quercetin: New Insights for Plant and Human Health. Plants 2022, Vol 11, Page 2623. 11(19):2623. doi:10.3390/PLANTS11192623. https://www.mdpi.com/2223-7747/11/19/2623/htm.

Jeandet P, Sobarzo-Sánchez E, Clément C, Nabavi SF, Habtemariam S, Nabavi SM, Cordelier S. 2018. Engineering stilbene metabolic pathways in microbial cells. Biotechnol Adv. 36(8):2264–2283. doi:10.1016/j.biotechadv.2018.11.002.

Jenny RJ, Mann KG, Lundblad RL. 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expr Purif. 31(1):1–11. doi:10.1016/S1046-5928(03)00168-2.

Jones JA, Vernacchio VR, Collins SM, Shirke AN, Xiu Y, Englaender JA, Cress BF, McCutcheon CC, Linhardt RJ, Gross RA, Koffas MAG. 2017. Complete Biosynthesis of Anthocyanins Using *E. coli* Polycultures. mBio. 8(3). doi:10.1128/mBio.00621-17.

Jones JA, Vernacchio VR, Sinkoe AL, Collins SM, Ibrahim MHA, Lachance DM, Hahn J, Koffas MAG. 2016. Experimental and computational optimization of an *Escherichia coli*

co-culture for the efficient production of flavonoids. Metab Eng. 35:55–63. doi:10.1016/j.ymben.2016.01.006.

Josiah SS, Famusiwa CD, Crown OO, Lawal AO, Olaleye MT, Akindahunsi AA, Akinmoladun AC. 2022. Neuroprotective effects of catechin and quercetin in experimental Parkinsonism through modulation of dopamine metabolism and expression of IL-1β, TNF-α, NF-κB, IκKB, and p53 genes in male Wistar rats. Neurotoxicology. 90:158–171. doi:10.1016/J.NEURO.2022.03.004.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature. 596(7873):583–589. doi:10.1038/s41586-021-03819-2. [accessed 2023 Jul 10]. https://www.nature.com/articles/s41586-021-03819-2.

Kabera JN, Semana E, Mussa AR, He X. 2014. Plant Secondary Metabolites: Biosynthesis, Classification, Function and Pharmacological Properties.

Kandel S. 2016. Structural and Functional Analysis of Grapefruit Flavonol-Specific-3-O-GT Mutant P145T. [Johnson City]: East Tennessee State University. [accessed 2023 Aug 9]. https://dc.etsu.edu/etd/3150.

Karak P. 2019. Biological activities of flavonoids: an overview. Int J Pharm Sci Res. 10(4):1567–1574. doi:10.13040/IJPSR.0975-8232.10(4).1567-74. http://ijpsr.com/bft-article/biological-activities-of-flavonoids-an-overview/?view=fulltext.

Karbalaei M, Rezaee SA, Farsiani H. 2020. *Pichia pastoris*: A highly successful expression system for optimal synthesis of heterologous proteins. J Cell Physiol. 235(9):5867–5881. doi:10.1002/JCP.29583. [accessed 2023 Aug 9]. https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.29583.

Katsuyama Y, Funa N, Miyahisa I, Horinouchi S. 2007. Synthesis of Unnatural Flavonoids and Stilbenes by Exploiting the Plant Biosynthetic Pathway in *Escherichia coli*. Chem Biol. 14(6):613–621. doi:10.1016/j.chembiol.2007.05.004.

Kennedy DO, Wightman EL. 2011. Herbal Extracts and Phytochemicals: Plant Secondary Metabolites and the Enhancement of Human Brain function. Advances in Nutrition. 2(1):32–50. doi:10.3945/an.110.000117.

Kielkopf CL, Bauer W, Urbatsch IL. 2020. Purification of Polyhistidine-Tagged Proteins by Immobilized Metal Affinity Chromatography. Cold Spring Harb Protoc. 2020(6):102194–102194. doi:10.1101/PDB.PROT102194. [accessed 2023 Aug 9]. https://europepmc.org/article/med/32482902.

Kim BG, Lee H, Ahn JH. 2014. Biosynthesis of pinocembrin from glucose using engineered *Escherichia coli*. J Microbiol Biotechnol. 24(11):1536–1541. doi:10.4014/jmb.1406.06011. http://www.ncbi.nlm.nih.gov/pubmed/25085569.

Kim BG, Yang SM, Kim SY, Cha MN, Ahn J-H. 2015. Biosynthesis and production of glycosylated flavonoids in *Escherichia coli*: current state and perspectives. Appl Microbiol Biotechnol. 99(7):2979–2988. doi:10.1007/s00253-015-6504-6.

Kim IS. 2022. Current perspectives on the beneficial effects of soybean isoflavones and their metabolites on plants. Food Sci Biotechnol. 31(5):515–526. doi:10.1007/S10068-022-01070-7/METRICS. [accessed 2023 Jun 26]. https://link.springer.com/article/10.1007/s10068-022-01070-7.

Klein AP, Sattely ES. 2017. Biosynthesis of cabbage phytoalexins from indole glucosinolate. Proc Natl Acad Sci U S A. 114(8):1910–1915. doi:10.1073/pnas.1615625114.

Kogawa K, Kato N, Kazuma K, Noda N, Suzuki M. 2007. Purification and characterization of UDP-glucose: anthocyanin 3',5'-O-glucosyltransferase from *Clitoria ternatea*. Planta. 226(6):1501–1509. doi:10.1007/S00425-007-0584-1. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/17668234/.

Koirala N, Thuan NH, Ghimire GP, Thang D Van, Sohng JK. 2016. Methylation of flavonoids: Chemical structures, bioactivities, progress and perspectives for biotechnological production. Enzyme Microb Technol. 86:103–116. doi:10.1016/J.ENZMICTEC.2016.02.003.

Konrad M, Nieman DC. 2015. Evaluation of Quercetin as a Countermeasure to Exercise-Induced Physiological Stress. Antioxidants in Sport Nutrition.:155–170. doi:10.1201/b17442-10. [accessed 2023 Aug 4]. https://www.ncbi.nlm.nih.gov/books/NBK299055/.

Kootstra A. 1994. Protection from UV-B-induced DNA damage by flavonoids. Plant Mol Biol. 26(2):771–774. doi:10.1007/BF00013762.

Kosobokova EN, Skrypnik KA, Kosorukov VS. 2016. Overview of fusion tags for recombinant proteins. Biochemistry (Moscow). 81(3):187–200. doi:10.1134/S0006297916030019/METRICS. [accessed 2023 Aug 9]. https://link.springer.com/article/10.1134/S0006297916030019.

Kotelchuck D, Scheraga HA. 1969. The influence of short-range interactions on protein conformation, II. A model for predicting the alpha-helical regions of proteins. Proc Natl Acad Sci U S A. 62(1):14–21. doi:10.1073/pnas.62.1.14.

Koukol J, Conn EE. 1961. The Metabolism of Aromatic Compounds in Higher Plants. Journal of Biological Chemistry. 236(10):2692–2698. doi:10.1016/s0021-9258(19)61721-7.

Kozłowska A, Szostak-Wegierek D. 2014. Flavonoids--food sources and health benefits. Rocz Panstw Zakl Hig. 65(2):79–85.

Kozłowska A, Szostak-Węgierek D. 2022. Targeting Cardiovascular Diseases by Flavonols: An Update. Nutrients. 14(7):1439. doi:10.3390/NU14071439/S1. [accessed 2023 Jul 7]. https://www.mdpi.com/2072-6643/14/7/1439/htm.

Kreuzaler F, Hahlbrock K. 1972. Enzymatic synthesis of aromatic compounds in higher plants: Formation of naringenin (5,7,4′-trihydroxyflavanone) from p-coumaroyl coenzyme A and malonyl coenzyme A. FEBS Lett. 28(1):69–72. doi:10.1016/0014-5793(72)80679-3. [accessed 2023 Aug 5]. https://onlinelibrary.wiley.com/doi/full/10.1016/0014-5793%2872%2980679-3.

Kreuzaler F, Hahlbrock K. 1975. Enzymic synthesis of an aromatic ring from acetate units. Partial purification and some properties of flavanone synthase from cell-suspension cultures of *Petroselinum hortense*. Eur J Biochem. 56(1):205–213. doi:10.1111/J.1432-

1033.1975.TB02223.X. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/240705/.

Krishnamurthy P, Tsukamoto C, Ishimoto M. 2020. Reconstruction of the Evolutionary Histories of UGT Gene Superfamily in Legumes Clarifies the Functional Divergence of Duplicates in Specialized Metabolism. Int J Mol Sci. 21(5). doi:10.3390/IJMS21051855. [accessed 2023 Jul 7]. /pmc/articles/PMC7084467/.

Kröner A, Marnet N, Andrivon D, Val F. 2012. Nicotiflorin, rutin and chlorogenic acid: Phenylpropanoids involved differently in quantitative resistance of potato tubers to biotrophic and necrotrophic pathogens. Plant Physiology and Biochemistry. 57:23–31. doi:10.1016/J.PLAPHY.2012.05.006.

Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. 2021. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. Proteins: Structure, Function and Bioinformatics. 89(12):1607–1617. doi:10.1002/PROT.26237.

Kubo A, Arai Y, Nagashima S, Yoshikawa T. 2004. Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys. 429(2):198–203. doi:10.1016/J.ABB.2004.06.021.

Kumar A, Mahajan A, Begum Z. 2022. In-vitro Estimation of Phytochemicals & Flavonoids in *Citrus sinensis* & Study their Antimicrobial & Antioxidant Activity. Asian Journal of Advanced Research and Reports. doi:10.9734/ajarr/2022/v16i930503.

Kumar S, Pandey AK. 2013. Chemistry and Biological Activities of Flavonoids: An Overview. The Scientific World Journal. 2013:13. doi:https://doi.org/10.1155/2013/162750. https://www.ncbi.nlm.nih.gov/pubmed/24470791.

Kutchan TM. 1995. Alkaloid biosynthesis - The basis of metabolic engineering of medicinal plants. Plant Cell. 7(7):1059–1070. doi:10.1105/tpc.7.7.1059.

Kwak S, Jin Y-S. 2017. Production of fuels and chemicals from xylose by engineered *Saccharomyces cerevisiae*: a review and perspective. Microb Cell Fact. 16(1):82. doi:10.1186/s12934-017-0694-9.

Lairson L, Henrissat B, Davies G, Withers S. 2008. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. 77(1):521–555. doi:10.1146/annurev.biochem.76.061005.092322.

Lauber J, Handrick R, Leptihn S, Dürre P, Gaisser S. 2015. Expression of the functional recombinant human glycosyltransferase GalNAcT2 in *Escherichia coli*. Microb Cell Fact. 14(1). doi:10.1186/s12934-014-0186-0.

LaVallie ER, McCoy JM, Smith DB, Riggs P. 1994. Enzymatic and Chemical Cleavage of Fusion Proteins. Curr Protoc Mol Biol. 28(1):16.4.5-16.4.17. doi:10.1002/0471142727.MB1604BS28. [accessed 2023 Aug 9]. https://onlinelibrary.wiley.com/doi/full/10.1002/0471142727.mb1604bs28.

Lee YS, Woo J Bin, Ryu SI, Moon SK, Han NS, Lee SB. 2017. Glucosylation of flavonol and flavanones by *Bacillus* cyclodextrin glucosyltransferase to enhance their solubility and stability. Food Chem. 229:75–83. doi:10.1016/j.foodchem.2017.02.057. http://dx.doi.org/10.1016/j.foodchem.2017.02.057.

Lehka BJ, Eichenberger M, Bjørn-Yoshimoto WE, Vanegas KG, Buijs N, Jensen NB, Dyekjær JD, Jenssen H, Simon E, Naesby M. 2017. Improving heterologous production of phenylpropanoids in *Saccharomyces cerevisiae* by tackling an unwanted side reaction of

Tsc13, an endogenous double bond reductase. FEMS Yeast Res.:fox004. doi:10.1093/femsyr/fox004.

Lei J, Jayaprakasha GK, Singh J, Uckoo R, Borrego EJ, Finlayson S, Kolomiets M, Patil BS, Braam J, Zhu-Salzman K. 2019. CIRCADIAN CLOCK-ASSOCIATED1 controls resistance to aphids by altering indole glucosinolate production. Plant Physiol. 181(3):1344–1359. doi:10.1104/pp.19.00676.

Lewis PN, Go N, Go M, Kotelchuck D, Scheraga HA. 1970. Helix probability profiles of denatured proteins and their correlation with native structures. Proc Natl Acad Sci U S A. 65(4):810–5. doi:10.1073/pnas.65.4.810.

Li F, Dong S, Leier A, Han M, Guo X, Xu J, Wang X, Pan S, Jia C, Zhang Y, Webb GI, Coin LJM, Li C, Song J. 2022. Positive-unlabeled learning in bioinformatics and computational biology: a brief review. Brief Bioinform. 23(1). doi:10.1093/bib/bbab461.

Li G, Meng X, Zhu M, Li Z. 2019. Research Progress of Betalain in Response to Adverse Stresses and Evolutionary Relationship Compared with Anthocyanin. Molecules. 24(17):3078. doi:10.3390/molecules24173078.

Li J, Ou-Lee TM, Raba R, Amundson RG, Last RL. 1993. *Arabidopsis* Flavonoid Mutants Are Hypersensitive to UV-B Irradiation. Plant Cell. 5(2):171–179. doi:10.1105/TPC.5.2.171. [accessed 2023 Aug 3]. https://pubmed.ncbi.nlm.nih.gov/12271060/.

Li L, Modolo L V, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. 2007. Crystal Structure of *Medicago truncatula* UGT85H2 - Insights into the Structural Basis of a Multifunctional (Iso)flavonoid Glycosyltransferase. J Mol Biol. 370(5):951–963. doi:10.1016/j.jmb.2007.05.036.

Li M, Schneider K, Kristensen M, Borodina I, Nielsen J. 2016. Engineering yeast for high-level production of stilbenoid antioxidants. Sci Rep. 6(1):36827. doi:10.1038/srep36827.

Li R, Yu H, Muthana SM, Freedberg DI, Chen X. 2020. Size-Controlled Chemoenzymatic Synthesis of Homogeneous Oligosaccharides of *Neisseria meningitidis* W Capsular Polysaccharide. ACS Catal. 10(4):2791–2798. doi:10.1021/ACSCATAL.9B05597/ASSET/IMAGES/LARGE/CS9B05597_0001.JPEG. [accessed 2023 Aug 18]. https://pubs.acs.org/doi/full/10.1021/acscatal.9b05597.

Li Y, Li S, Thodey K, Trenchard I, Cravens A, Smolke CD. 2018. Complete biosynthesis of noscapine and halogenated alkaloids in yeast. Proceedings of the National Academy of Sciences. 115(17). doi:10.1073/pnas.1721469115.

Li Y, Li XL, Lai CJS, Wang RS, Kang LP, Ma T, Zhao ZH, Gao W, Huang LQ. 2019. Functional characterization of three flavonoid glycosyltransferases from *Andrographis paniculata*. R Soc Open Sci. 6(6). doi:10.1098/RSOS.190150. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/31312486/.

Li Y, Liu F, Li P, Wang T, Zheng C, Hou B. 2020. An *Arabidopsis* Cytokinin-Modifying Glycosyltransferase UGT76C2 Improves Drought and Salt Tolerance in Rice. Front Plant Sci. 11:560696. doi:10.3389/FPLS.2020.560696/BIBTEX.

Li Y, Zhang C, Yu D-J, Zhang Y. 2022. Deep learning geometrical potential for high-accuracy ab initio protein structure prediction. iScience. 25(6):104425. doi:10.1016/j.isci.2022.104425.

Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG. 2011. High-Yield Resveratrol Production in Engineered *Escherichia coli*. Appl Environ Microbiol. 77(10):3451–3460. doi:10.1128/AEM.02186-10.

Lim EK, Jackson RG, Bowles DJ. 2005. Identification and characterization of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. FEBS Lett. 579(13):2802–2806. doi:10.1016/J.FEBSLET.2005.04.016. [accessed 2023 Jul 7]. https://pubmed.ncbi.nlm.nih.gov/15907484/.

Lin J, Rexrode KM, Hu F, Albert CM, Chae CU, Rimm EB, Stampfer MJ, Manson JE. 2007. Dietary intakes of flavonols and flavones and coronary heart disease in US women. Am J Epidemiol. 165(11):1305–1313. doi:10.1093/AJE/KWM016. [accessed 2023 Aug 3]. https://pubmed.ncbi.nlm.nih.gov/17379619/.

Linton MF, Yancey PG, Davies SS, Jerome WG, Linton EF, Song WL, Doran AC, Vickers KC. 2019. The Role of Lipids and Lipoproteins in Atherosclerosis. Science (1979). 111(2877):166–186. [accessed 2023 Aug 3]. https://www.ncbi.nlm.nih.gov/sites/books/NBK343489/.

Liou CS, Sirk SJ, Diaz CAC, Klein AP, Fischer CR, Higginbottom SK, Erez A, Donia MS, Sonnenburg JL, Sattely ES. 2020. A Metabolic Pathway for Activation of Dietary Glucosinolates by a Human Gut Symbiont. Cell. 180(4):717-728.e19. doi:10.1016/j.cell.2020.01.023.

Liu CW, Murray JD. 2016. The Role of Flavonoids in Nodulation Host-Range Specificity: An Update. Plants 2016, Vol 5, Page 33. 5(3):33. doi:10.3390/PLANTS5030033. [accessed 2023 Aug 29]. https://www.mdpi.com/2223-7747/5/3/33/htm.

Logemann E, Wu S-C, Schröder J, Schmelzer E, Somssich IE, Hahlbrock K. 1995. Gene activation by UV light, fungal elicitor or fungal infection in *Petroselinum crispum* is correlated with repression of cell cycle-related genes. The Plant Journal. 8(6):865–876. doi:10.1046/j.1365-313X.1995.8060865.x.

Lopez-Lazaro M. 2009. Distribution and biological activities of the flavonoid luteolin. Mini Rev Med Chem. 9(1):31–59. doi:10.2174/138955709787001712. [accessed 2023 Apr 8]. https://pubmed.ncbi.nlm.nih.gov/19149659/.

Luo X, Wang R, Wang J, Li Y, Luo H, Chen S, Zeng X, Han Z. 2022. Acylation of Anthocyanins and Their Applications in the Food Industry: Mechanisms and Recent Research Advances. Foods 2022, Vol 11, Page 2166. 11(14):2166. doi:10.3390/FOODS11142166. [accessed 2023 Apr 16]. https://www.mdpi.com/2304-8158/11/14/2166/htm.

Lv Y, Marsafari M, Koffas M, Zhou J, Xu P. 2019. Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis. ACS Synth Biol. 8(11):2514–2523. doi:10.1021/acssynbio.9b00193.

M. Calderon-Montano J, Burgos-Moron E, Perez-Guerrero C, Lopez-Lazaro M. 2011. A review on the dietary flavonoid kaempferol. Mini Rev Med Chem. 11(4):298–344. doi:10.2174/138955711795305335. [accessed 2023 Apr 8]. https://pubmed.ncbi.nlm.nih.gov/21428901/.

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Bélanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P, Magdalou J, Chowdhury JR, Ritter JK, Schachter H, Tephly TR, Tipton KF, Nebert DW. 1997. The UDP glycosyltransferase

gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics. 7(4):255–69. doi:10.1097/00008571-199708000-00001.

Maeda H, Chatani E, Koyama T, Sugiura M, Izumi H, Hayashi R. 2004. Indiscriminate glycosylation of procarboxypeptidase Y expressed in *Pichia pastoris*. Carbohydr Res. 339(6):1041–1045. doi:10.1016/j.carres.2004.02.003. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/15063190/.

Mahmoud AM, Hernández Bautista RJ, Sandhu MA, Hussein OE. 2019. Beneficial effects of citrus flavonoids on cardiovascular and metabolic health. Oxid Med Cell Longev. 2019. doi:10.1155/2019/5484138.

Malka SK, Cheng Y. 2017. Possible interactions between the biosynthetic pathways of indole glucosinolate and Auxin. Front Plant Sci. 8. doi:10.3389/fpls.2017.02131.

Mark R, Lyu X, Ng KR, Chen WN. 2019. Gene Source Screening as a Tool for Naringenin Production in Engineered *Saccharomyces cerevisiae*. ACS Omega. 4(7):12872–12879. doi:10.1021/acsomega.9b00364.

Martens S, Forkmann G, Matern U, Lukačin R. 2001. Cloning of parsley flavone synthase I. Phytochemistry. 58(1):43–46. doi:10.1016/S0031-9422(01)00191-1. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/11524111/.

De Masi L, Bontempo P, Rigano D, Stiuso P, Carafa V, Nebbioso A, Piacente S, Montoro P, Aversano R, D'Amelia V, Carputo D, Altucci L. 2020. Comparative Phytochemical Characterization, Genetic Profile, and Antiproliferative Activity of Polyphenol-Rich Extracts from Pigmented Tubers of Different *Solanum tuberosum* Varieties. Molecules. 25(1). doi:10.3390/molecules25010233.

Mathesius U. 2018. Flavonoid Functions in Plants and Their Interactions with Other Organisms. Plants (Basel). 7(2). doi:10.3390/plants7020030.

Matsuura HN, Fett-Neto AG. 2015. Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action. Plant Toxins.:1–15. doi:10.1007/978-94-007-6728-7_2-1.

McArthur JB, Chen X. 2016. Glycosyltransferase engineering for carbohydrate synthesis. Biochem Soc Trans. 44(1):129. doi:10.1042/BST20150200. [accessed 2023 Aug 6]. /pmc/articles/PMC5764748/.

McIntosh CA, Latchinian L, Mansell RL. 1990. Flavanone-specific 7-O-glucosyltransferase activity in *Citrus paradisi* seedlings: purification and characterization. Arch Biochem Biophys. 282(1):50–7. doi:10.1016/0003-9861(90)90085-d.

McIntosh CA, Mansell RL. 1990. Biosynthesis of naringin in *Citrus paradisi*: UDP-glucosyl-transferase activity in grapefruit seedlings. Phytochemistry. 29(5):1533–1538. doi:10.1016/0031-9422(90)80115-W.

McIntosh CA, Owens DK. 2016. Advances in flavonoid glycosyltransferase research: integrating recent findings with long-term citrus studies. Phytochemistry Reviews. 15(6):1075–1091. doi:10.1007/s11101-016-9460-6.

Menezes JCJMDS, Orlikova B, Morceau F, Diederich M. 2016. Natural and Synthetic Flavonoids: Structure–Activity Relationship and Chemotherapeutic Potential for the Treatment of Leukemia. Crit Rev Food Sci Nutr. 56(sup1):S4–S28. doi:10.1080/10408398.2015.1074532.

Meng N, Chen K, Wang Y, Hou J, Chu W, Xie S, Yang F, Sun C. 2022. Dihydrohomoplantagin and Homoplantaginin, Major Flavonoid Glycosides from *Salvia plebeia* R. Br. Inhibit oxLDL-Induced Endothelial Cell Injury and Restrict Atherosclerosis via Activating Nrf2 Anti-Oxidation Signal Pathway. Molecules. 27(6):1990. doi:10.3390/MOLECULES27061990/S1. [accessed 2023 Aug 18]. https://www.mdpi.com/1420-3049/27/6/1990/htm.

Miguel MG. 2018. Betalains in some species of the amaranthaceae family: A review. Antioxidants. 7(4). doi:10.3390/antiox7040053.

Mikkelsen MD, Petersen BL, Olsen CE, Halkier BA. 2002. Biosynthesis and metabolic engineering of glucosinolates. Amino Acids. 22(3):279–295. doi:10.1007/s007260200014.

Milke L, Ferreira P, Kallscheuer N, Braga A, Vogt M, Kappelmann J, Oliveira J, Silva AR, Rocha I, Bott M, Noack S, Faria N, Marienhagen J. 2019. Modulation of the central carbon metabolism of *Corynebacterium glutamicum* improves malonyl-CoA availability and increases plant polyphenol synthesis. Biotechnol Bioeng. 116(6):1380–1391. doi:10.1002/bit.26939.

Mital S, Christie G, Dikicioglu D. 2021. Recombinant expression of insoluble enzymes in *Escherichia coli*: a systematic review of experimental design and its manufacturing implications. Microbial Cell Factories 2021 20:1. 20(1):1–20. doi:10.1186/S12934-021-01698-W. [accessed 2023 Aug 9]. https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-021-01698-w.

Modolo L V, Blount JW, Achnine L, Naoumkina MA, Wang X, Dixon RA. 2007. A functional genomics approach to (iso)flavonoid glycosylation in the model legume *Medicago truncatula*. Plant Mol Biol. 64(5):499–518. doi:10.1007/s11103-007-9167-6.

Modolo L V, Li L, Pan H, Blount JW, Dixon RA, Wang X. 2009. Crystal Structures of Glycosyltransferase UGT78G1 Reveal the Molecular Basis for Glycosylation and Deglycosylation of (Iso)flavonoids. J Mol Biol. 392(5):1292–1302. doi:10.1016/j.jmb.2009.08.017. http://dx.doi.org/10.1016/j.jmb.2009.08.017.

Mohideen FI, Nguyen LH, Richard JD, Ouadhi S, Kwan DH. 2022. *In Vitro* Reconstitution of the dTDP- l -Daunosamine Biosynthetic Pathway Provides Insights into Anthracycline Glycosylation. ACS Chem Biol. 17(12):3331–3340. doi:10.1021/ACSCHEMBIO.1C00646/ASSET/IMAGES/LARGE/CB1C00646_0006.JPEG. [accessed 2023 Jun 26]. https://pubs.acs.org/doi/full/10.1021/acschembio.1c00646.

Moritzer AC, Prior T, Niemann HH. 2020. Not Cleaving the His-tag of Thal Results in More Tightly Packed and Better-Diffracting Crystals. Crystals 2020, Vol 10, Page 1135. 10(12):1135. doi:10.3390/CRYST10121135. [accessed 2023 Aug 9]. https://www.mdpi.com/2073-4352/10/12/1135/htm.

Moustafa E, Wong E. 1967. Purification and properties of chalcone-flavanone isomerase from soya bean seed. Phytochemistry. 6(5):625–632. doi:10.1016/S0031-9422(00)86001-X.

Nagatomo Y, Usui S, Ito T, Kato A, Shimosaka M, Taguchi G. 2014. Purification, molecular cloning and functional characterization of flavonoid *C* -glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. The Plant Journal. 80(3):437–448. doi:10.1111/tpj.12645.

Nakagawa A, Matsumura E, Koyanagi T, Katayama T, Kawano N, Yoshimatsu K, Yamamoto K, Kumagai H, Sato F, Minami H. 2016. Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. Nat Commun. 7. doi:10.1038/ncomms10390.

Nakai S. 1983. Structure-Function Relationships of Food Proteins with an Emphasis on the Importance of Protein Hydrophobicity. J Agric Food Chem. 31(4):676–683. doi:10.1021/JF00118A001/ASSET/JF00118A001.FP.PNG_V03. [accessed 2023 Sep 21]. https://pubs.acs.org/doi/abs/10.1021/jf00118a001.

Neish AC. 1960. Biosynthetic Pathways of Aromatic Compounds. Annu Rev Plant Physiol. 11(1):55–80. doi:10.1146/ANNUREV.PP.11.060160.000415. [accessed 2023 Aug 5]. https://www.annualreviews.org/doi/abs/10.1146/annurev.pp.11.060160.000415.

Nguyen HM, Putterill J, Dare AP, Plunkett BJ, Cooney J, Peng Y, Souleyre EJF, Albert NW, Espley R V., Günther CS. 2023. Two genes, ANS and UFGT2, from *Vaccinium* spp. are key steps for modulating anthocyanin production. Front Plant Sci. 14:1082246. doi:https://doi.org/10.3389/fpls.2023.1082246.

Nidetzky B, Gutmann A, Zhong C. 2018. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. ACS Catal. 8(7):6283–6300. doi:https://doi.org/10.1021/acscatal.8b00710. [accessed 2023 Aug 18]. https://pubs.acs.org/doi/full/10.1021/acscatal.8b00710.

Nieman DC. 2007. Marathon training and immune function. Sports Med. 37(4–5):412–415. doi:10.2165/00007256-200737040-00036. [accessed 2023 Aug 4]. https://pubmed.ncbi.nlm.nih.gov/17465622/.

NIH National Library of Medicine. 3-Hydroxyflavone | C15H10O3 | CID 11349. PubChem. [accessed 2023 Aug 15]. https://pubchem.ncbi.nlm.nih.gov/compound/3-Hydroxyflavone.

Nix A, Paull C, Colgrave M. 2017. Flavonoid profile of the cotton plant, *Gossypium hirsutum*: A review. Plants. 6(4). doi:10.3390/plants6040043.

Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T. 2007. A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (*Glycine max*) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. J Biol Chem. 282(32):23581–90. doi:10.1074/jbc.M702651200.

Noirclerc-Savoye M, Flayhan A, Pereira C, Gallet B, Gans P, Ebel C, Breyton C. 2015. Tail proteins of phage T5: investigation of the effect of the His6-tag position, from expression to crystallization. Protein Expr Purif. 109:70–78. doi:10.1016/J.PEP.2015.02.003. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/25676818/.

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. 2006. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO Journal. 25(6):1396–1405. doi:10.1038/sj.emboj.7600970.

Ono E, Fukuchi-Mizutani M, Nakamura N, Fukui Y, Yonekura-Sakakibara K, Yamaguchi M, Nakayama T, Tanaka T, Kusumi T, Tanaka Y. 2006. Yellow flowers generated by expression of the aurone biosynthetic pathway. Proceedings of the National Academy of Sciences. 103(29):11075–11080. doi:10.1073/pnas.0604246103.

Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y, Ishiguro M, Fukui Y, Nakayama T. 2010. Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant Cell. 22(8):2856–2871. doi:10.1105/TPC.110.074625. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/20693356/.

Ooi KE, Zhang XW, Kuo CY, Liu YJ, Yu CC. 2022. Chemoenzymatic Synthesis of Asymmetrically Branched Human Milk Oligosaccharide Lacto-N-Hexaose. Front Chem. 10. doi:10.3389/fchem.2022.905105. [accessed 2023 Nov 1]. https://pubmed.ncbi.nlm.nih.gov/35711960/.

Ortiz R, Geleta M, Gustafsson C, Lager I, Hofvander P, Löfstedt C, Cahoon EB, Minina E, Bozhkov P, Stymne S. 2020. Oil crops for the future. Curr Opin Plant Biol. 56:181–189. doi:10.1016/j.pbi.2019.12.003.

Owens DK, Crosby KC, Runac J, Howard BA, Winkel BSJ. 2008. Biochemical and genetic characterization of *Arabidopsis* flavanone 3beta-hydroxylase. Plant Physiol Biochem. 46(10):833–843. doi:10.1016/J.PLAPHY.2008.06.004. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/18657430/.

Owens DK, McIntosh CA. 2009. Identification, recombinant expression, and biochemical characterization of a flavonol 3-O-glucosyltransferase clone from *Citrus paradisi*. Phytochemistry. 70(11–12):1382–1391. doi:10.1016/j.phytochem.2009.07.027. [accessed 2023 Nov 1]. https://www.sciencedirect.com/science/article/pii/S0031942209003033?via%3Dihub.

Owens DK, McIntosh CA. 2011. Biosynthesis and Function of Citrus Glycosylated Flavonoids. In: Gang D, editor. Recent Advances in Phytochemistry: The Biological Activity of Phytochemicals. Vol. 41. Springer New York. p. 67–96. https://doi.org/10.1007/978-1-4419-7299-6_6.

Panche AN, Diwan AD, Chandra SR. 2016. Flavonoids: An overview. J Nutr Sci. 5:e47. doi:10.1017/jns.2016.41.

Pandey RK, Ojha R, Prajapati VK. 2020. Wet-Lab Approaches to Determine Three-Dimensional Structures of Proteins. Frontiers in Protein Structure, Function, and Dynamics.:57–70. doi:10.1007/978-981-15-5530-5_3. [accessed 2023 Sep 21]. https://link.springer.com/chapter/10.1007/978-981-15-5530-5_3.

Pandey RP, Parajuli P, Koffas MAG, Sohng JK. 2016. Microbial production of natural and non-natural flavonoids: Pathway engineering, directed evolution and systems/synthetic biology. Biotechnol Adv. 34(5):634–662. doi:10.1016/j.biotechadv.2016.02.012.

Park S, Kim B-K, Park S-K. 2022. Effects of Fisetin, a Plant-Derived Flavonoid, on Response to Oxidative Stress, Aging, and Age-Related Diseases in *Caenorhabditis elegans*. Pharmaceuticals (Basel). 15(12). doi:10.3390/ph15121528.

Parray JA, Yaseen Mir M, Shameem N. 2019. Plant Genetic Engineering and GM Crops: Merits and Demerits. In: Sustainable Agriculture: Biotechniques in Plant Biology. Springer, Singapore. p. 155–229. [accessed 2023 Nov 1]. https://link.springer.com/chapter/10.1007/978-981-13-8840-8_4.

Petersen A, Crocoll C, Halkier BA. 2019. *De novo* production of benzyl glucosinolate in *Escherichia coli*. Metab Eng. 54:24–34. doi:10.1016/j.ymben.2019.02.004.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 25(13):1605–12. doi:10.1002/jcc.20084.

Pfalz M, Mikkelsen MD, Bednarek P, Olsen CE, Halkier BA, Kroymann J. 2011. Metabolic Engineering in *Nicotiana benthamiana* Reveals Key Enzyme Functions in *Arabidopsis* Indole Glucosinolate Modification. Plant Cell. 23(2):716–729. doi:10.1105/tpc.110.081711.

Pichersky E, Raguso RA. 2018. Why do plants produce so many terpenoid compounds? New Phytologist. 220(3):692–702. doi:10.1111/NPH.14178. [accessed 2023 Apr 7]. https://onlinelibrary.wiley.com/doi/full/10.1111/nph.14178.

Pinheiro PF, Justino GC, Pinheiro PF, Justino GC. 2012. Structural Analysis of Flavonoids and Related Compounds - A Review of Spectroscopic Applications. Phytochemicals - A Global Perspective of Their Role in Nutrition and Health. doi:10.5772/29152. [accessed 2023 Apr 7]. https://www.intechopen.com/chapters/32937.

Pixley K V., Falck-Zepeda JB, Giller KE, Glenna LL, Gould F, Mallory-Smith CA, Stelly DM, Stewart CN. 2019. Genome Editing, Gene Drives, and Synthetic Biology: Will They Contribute to Disease-Resistant Crops, and Who Will Benefit? Annu Rev Phytopathol. 57(1):165–188. doi:10.1146/annurev-phyto-080417-045954.

Polturak G, Aharoni A. 2018. “La Vie en Rose”: Biosynthesis, Sources, and Applications of Betalain Pigments. Mol Plant. 11(1):7–22. doi:10.1016/j.molp.2017.10.008. http://www.ncbi.nlm.nih.gov/pubmed/29081360.

Polturak G, Aharoni A. 2019. Advances and future directions in betalain metabolic engineering. New Phytologist. 224(4):1472–1478. doi:10.1111/nph.15973. https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.15973.

Polturak G, Grossman N, Vela-Corcia D, Dong Y, Nudel A, Pliner M, Levy M, Rogachev I, Aharoni A. 2017. Engineered gray mold resistance, antioxidant capacity, and pigmentation in betalain-producing crops and ornamentals. Proc Natl Acad Sci U S A. 114(34):9062–9067. doi:10.1073/pnas.1707176114.

Pontrelli S, Chiu TY, Lan EI, Chen FYH, Chang P, Liao JC. 2018. *Escherichia coli* as a host for metabolic engineering. Metab Eng. 50(April):16–46. doi:10.1016/j.ymben.2018.04.008. http://dx.doi.org/10.1016/j.ymben.2018.04.008.

Prevarskaya N, Zhang L, Barritt G. 2007. TRP channels in cancer. Biochim Biophys Acta Mol Basis Dis. 1772(8):937–946. doi:10.1016/j.bbadis.2007.05.006.

Procházková D, Boušová I, Wilhelmová N. 2011. Antioxidant and prooxidant properties of flavonoids. Fitoterapia. 82(4):513–523. doi:10.1016/J.FITOTE.2011.01.018.

Prothero JW. 1966. Correlation between the distribution of amino acids and alpha helices. Biophys J. 6(3):367–70. doi:10.1016/S0006-3495(66)86662-6.

Puxbaum V, Mattanovich D, Gasser B. 2015. Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia Pastoris*. Appl Microbiol Biotechnol. 99(7):2925–2938. doi:10.1007/S00253-015-6470-Z/METRICS. [accessed 2023 Aug 9]. https://link.springer.com/article/10.1007/s00253-015-6470-z.

Pyne ME, Narcross L, Martin VJJ. 2019. Engineering Plant Secondary Metabolism in Microbial Systems. Plant Physiol. 179(3):844–861. doi:10.1104/pp.18.01291.

Qian H, Sun B, Miao H, Cai C, Xu C, Wang Q. 2015. Variation of glucosinolates and quinone reductase activity among different varieties of Chinese kale and improvement of glucoraphanin by metabolic engineering. Food Chem. 168:321–326. doi:10.1016/j.foodchem.2014.07.073.

Qu G, Li A, Acevedo-Rocha CG, Sun Z, Reetz MT. 2020. The Crucial Role of Methodology Development in Directed Evolution of Selective Enzymes. Angewandte Chemie International Edition. 59(32):13204–13231. doi:10.1002/ANIE.201901491. [accessed 2023 Aug 9]. https://onlinelibrary.wiley.com/doi/full/10.1002/anie.201901491.

Rafińska K, Pomastowski P, Wrona O, Górecki R, Buszewski B. 2017. *Medicago sativa* as a source of secondary metabolites for agriculture and pharmaceutical industry. Phytochem Lett. 20:520–539. doi:10.1016/j.phytol.2016.12.006.

Rahimi P, Abedimanesh S, Mesbah-Namin SA, Ostadrahimi A. 2019. Betalains, the nature-inspired pigments, in health and diseases. Crit Rev Food Sci Nutr. 59(18):2949–2978. doi:10.1080/10408398.2018.1479830. http://www.ncbi.nlm.nih.gov/pubmed/29846082.

Rashidi S, Yousefi AR, Pouryousef M, Goicoechea N. 2022. Effect of arbuscular mycorrhizal fungi on the accumulation of secondary metabolites in roots and reproductive organs of *Solanum nigrum*, *Digitaria sanguinalis* and *Ipomoea purpurea*. Chemical and Biological Technologies in Agriculture. 9(1):1–11. doi:10.1186/S40538-022-00288-1/TABLES/4. [accessed 2023 Aug 15]. https://chembioagro.springeropen.com/articles/10.1186/s40538-022-00288-1.

Ratnayake R, Covell D, Ransom TT, Gustafson KR, Beutler JA. 2009. Englerin a, a selective inhibitor of renal cancer cell growth, from *Phyllanthus engleri*. Org Lett. 11(1):57–60. doi:10.1021/ol802339w.

Ratzka A, Vogel H, Kliebenstein DJ, Mitchell-Olds T, Kroymann J. 2002. Disarming the mustard oil bomb. Proc Natl Acad Sci U S A. 99(17):11223–11228. doi:10.1073/pnas.172112899.

Ravishankar D, Rajora AK, Greco F, Osborn HelenMI. 2013. Flavonoids as prospective compounds for anti-cancer therapy. Int J Biochem Cell Biol. 45(12):2821–2831. doi:10.1016/j.biocel.2013.10.004. http://www.ncbi.nlm.nih.gov/pubmed/24128857.

Reimer JJ, Shaaban B, Drummen N, Ambady SS, Genzel F, Poschet G, Wiese-Klinkenberg A, Usadel B, Wormit A. 2022. *Capsicum* Leaves under Stress: Using Multi-Omics Analysis to Detect Abiotic Stress Network of Secondary Metabolism in Two Species. Antioxidants. 11(4):671. doi:10.3390/ANTIOX11040671/S1. [accessed 2023 Aug 15]. https://www.mdpi.com/2076-3921/11/4/671/htm.

Richardson LL, Adler LS, Leonard AS, Andicoechea J, Regan KH, Anthony WE, Manson JS, Irwin RE. 2015. Secondary metabolites in floral nectar reduce parasite infections in bumblebees. Proceedings of the Royal Society B: Biological Sciences. 282(1803). doi:10.1098/RSPB.2014.2471. [accessed 2023 Aug 15]. https://royalsocietypublishing.org/doi/10.1098/rspb.2014.2471.

Richardson M. 1978. Flavonols and C-Glycosylflavonoids of the caryophyllales. Biochem Syst Ecol. 6(4):283–286. doi:10.1016/0305-1978(78)90046-7.

Robert F, Bierau H, Rossi M, Agugiaro D, Soranzo T, Broly H, Mitchell-Logean C. 2009. Degradation of an Fc-fusion recombinant protein by host cell proteases: Identification of

a CHO cathepsin D protease. Biotechnol Bioeng. 104(6):1132–1141. doi:10.1002/BIT.22494. [accessed 2023 Aug 9]. https://onlinelibrary.wiley.com/doi/full/10.1002/bit.22494.

Roell GW, Zha J, Carr RR, Koffas MA, Fong SS, Tang YJ. 2019. Engineering microbial consortia by division of labor. Microb Cell Fact. 18(1):35. doi:10.1186/s12934-019-1083-3.

Rosa R da, Schenkel EP, Bernardes LSC. 2020. Semisynthetic and newly designed derivatives based on natural chemical scaffolds: moving beyond natural products to fight *Trypanosoma cruzi*. Phytochemistry Reviews. 19(1):105–122. doi:10.1007/s11101-020-09659-8.

Le Roy J, Huss B, Creach A, Hawkins S, Neutelings G. 2016. Glycosylation Is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants. Front Plant Sci. 7:735. doi:10.3389/fpls.2016.00735.

Ruggiero A, Masullo M, Ruocco MR, Grimaldi P, Lanzotti MA, Arcari P, Zagari A, Vitagliano L. 2009. Structure and stability of a thioredoxin reductase from *Sulfolobus solfataricus*: a thermostable protein with two functions. Biochim Biophys Acta. 1794(3):554–562. doi:10.1016/J.BBAPAP.2008.11.011. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/19110078/.

Ruprecht C, Bönisch F, Ilmberger N, Heyer T V., Haupt ETK, Streit WR, Rabausch U. 2019. High level production of flavonoid rhamnosides by metagenome-derived Glycosyltransferase C in *Escherichia coli* utilizing dextrins of starch as a single carbon source. Metab Eng. 55:212–219. doi:10.1016/j.ymben.2019.07.002.

Russell DW. 1971. The Metabolism of Aromatic Compounds in Higher Plants: X. Properties Of The Cinnamic Acid 4-hydroxylase Of Pea Seedlings And Some Aspects Of Its Metabolic And Developmental Control. Journal of Biological Chemistry. 246(12):3870–3878. doi:10.1016/S0021-9258(18)62115-5. [accessed 2023 Aug 5]. https://linkinghub.elsevier.com/retrieve/pii/S0021925818621155.

Russell DW, Conn EE. 1967. The cinnamic acid 4-hydroxylase of pea seedlings. Arch Biochem Biophys. 122(1):256–258. doi:10.1016/0003-9861(67)90150-6. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/4383827/.

Sabaty M, Grosse S, Adryanczyk G, Boiry S, Biaso F, Arnoux P, Pignol D. 2013. Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. BMC Biochem. 14(1). doi:10.1186/1471-2091-14-28. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/24180491/.

Saigo T, Wang T, Watanabe M, Tohge T. 2020. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. Curr Opin Plant Biol. 55:93–99. doi:10.1016/J.PBI.2020.04.001.

Sakai Y, Koller A, Rangell LK, Keller GA, Subramani S. 1998. Peroxisome Degradation by Microautophagy in *Pichia pastoris*: Identification of Specific Steps and Morphological Intermediates. Journal of Cell Biology. 141(3):625–636. doi:10.1083/JCB.141.3.625. [accessed 2023 Aug 9]. http://www.jcb.org.

Salawu KM, Oyerinde AA, Bello RH. 2022. Antiproliferative and Antimicrobial Activities of *Citrus limon* (L.) Burm. f. Stem Bark Extract. Nigerian Journal of Basic and Applied Sciences. 29(1). doi:10.4314/njbas.v29i1.6.

Salehi B, Machin L, Monzote L, Sharifi-Rad J, Ezzat SM, Salem MA, Merghany RM, El Mahdy NM, Klllç CS, Sytar O, Sharifi-Rad M, Sharopov F, Martins N, Martorell M, Cho WC. 2020. Therapeutic Potential of Quercetin: New Insights and Perspectives for Human Health. ACS Omega. 5(20):11849–11872. doi:10.1021/ACSOMEGA.0C01818/ASSET/IMAGES/LARGE/AO0C01818_0002.JPEG. [accessed 2023 Aug 3]. https://pubs.acs.org/doi/full/10.1021/acsomega.0c01818.

Salehi B, Upadhyay S, Orhan IE, Jugran AK, Jayaweera SLD, Dias DA, Sharopov F, Taheri Y, Martins N, Baghalpour N, Cho WC, Sharifi-Rad J. 2019. Therapeutic Potential of α- and β-Pinene: A Miracle Gift of Nature. Biomolecules. 9(11). doi:10.3390/BIOM9110738. [accessed 2023 Apr 7]. /pmc/articles/PMC6920849/.

Sanlier N, Salsan MG. 2018. The Benefits of Brassica Vegetables on Human Health. J Human Health Res. 1(1):104. https://www.ncbi.nlm.nih.gov/pubmed/23631258.

Santoni G, Farfariello V. 2012. TRP Channels and Cancer: New Targets for Diagnosis and Chemotherapy. Endocr Metab Immune Disord Drug Targets. 11(1):54–67. doi:10.2174/187153011794982068.

Saravanan KM, Selvaraj S. 2017. Dihedral angle preferences of amino acid residues forming various non-local interactions in proteins. J Biol Phys. 43(2):265–278. doi:10.1007/s10867-017-9451-x.

Sasaki N, Nemoto K, Nishizaki Y, Sugimoto N, Tasaki K, Watanabe A, Goto F, Higuchi A, Morgan E, Hikage T, Nishihara M. 2021. Identification and characterization of xanthone biosynthetic genes contributing to the vivid red coloration of red-flowered gentian. The

Plant Journal. 107(6):1711–1723. doi:10.1111/TPJ.15412. [accessed 2023 Apr 16]. https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.15412.

Sathanantham P. 2015. Impact of Mutations of Targeted Serine, Histidine, and Glutamine Residues in *Citrus paradisi* Flavonol Specific Glucosyltransferase Activity. [Johnson City]: East Tennessee State University. [accessed 2023 Aug 9]. https://dc.etsu.edu/etd/2560.

Schenck CA, Maeda HA. 2018. Tyrosine biosynthesis, metabolism, and catabolism in plants. Phytochemistry. 149:82–102. doi:10.1016/j.phytochem.2018.02.003.

Schiffer M, Edmundson AB. 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys J. 7(2):121–35. doi:10.1016/S0006-3495(67)86579-2.

Schmidt A, Li C, Shi F, Jones AD, Pichersky E. 2011. Polymethylated Myricetin in Trichomes of the Wild Tomato Species *Solanum habrochaites* and Characterization of Trichome-Specific 3′/5′- and 7/4′-Myricetin O-Methyltransferases. Plant Physiol. 155(4):1999. doi:10.1104/PP.110.169961. [accessed 2023 Aug 3]. /pmc/articles/PMC3091102/.

Schrödinger Release 2023-3. 2023. Maestro. New York: Schrödinger, LLC.

Schutte-Smith M, Roodt A, Alberto R, Twigge L, Visser HG, Kirsten L, Koen R. 2019. Structures of rhenium(I) complexes with 3-hydroxyflavone and benzhydroxamic acid as O,O′-bidentate ligands and confirmation of π-stacking by solid-state NMR spectroscopy. Acta Crystallogr C Struct Chem. 75(4):378–387. doi:10.1107/S2053229619002717.

Schweiger R, Heise AM, Persicke M, Müller C. 2014. Interactions between the jasmonic and salicylic acid pathway modulate the plant metabolome and affect herbivores of different feeding types. Plant Cell Environ. 37(7):1574–1585. doi:10.1111/PCE.12257/SUPPINFO. [accessed 2023 Aug 15]. https://onlinelibrary.wiley.com/doi/full/10.1111/pce.12257.

Schweizer F, Colinas M, Pollier J, Moerkercke A Van, Bossche R Vanden, de Clercq R, Goossens A. 2018. An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in *Catharanthus roseus*. Metab Eng. 48:150–162. doi:10.1016/j.ymben.2018.05.016.

Sebghatollahi Z, Ghanadian M, Agarwal P, Ghaheh HS, Mahato N, Yogesh R, Hejazi SH. 2022. Citrus Flavonoids: Biological Activities, Implementation in Skin Health, and Topical Applications: A Review. ACS Food Science and Technology. 2(9):1417–1432. doi:10.1021/ACSFOODSCITECH.2C00165/ASSET/IMAGES/MEDIUM/FS2C00165_0009.GIF. https://pubs.acs.org/doi/full/10.1021/acsfoodscitech.2c00165.

Seca A, Pinto D. 2018. Plant Secondary Metabolites as Anticancer Agents: Successes in Clinical Trials and Therapeutic Application. Int J Mol Sci. 19(1):263. doi:10.3390/ijms19010263.

Semwal DK, Semwal RB, Combrinck S, Viljoen A. 2016. Myricetin: A Dietary Molecule with Diverse Biological Activities. Nutrients. 8(2). doi:10.3390/NU8020090. https://pubmed.ncbi.nlm.nih.gov/26891321/.

Sezer ED, Oktay LM, Karadadaş E, Memmedov H, Gunel NS, Sözmen E. 2019. Assessing Anticancer Potential of Blueberry Flavonoids, Quercetin, Kaempferol, and Gentisic Acid,

Through Oxidative Stress and Apoptosis Parameters on HCT-116 Cells. J Med Food. 22(11):1118–1126. doi:10.1089/jmf.2019.0098.

Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X. 2005. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. Plant Cell. 17(11):3141–3154. doi:10.1105/tpc.105.035055.

Sharma A, Tuli HS, Sharma AK. 2019. Chemistry and Synthetic Overview of Flavonoids. Current Aspects of Flavonoids: Their Role in Cancer Treatment.:23–38. doi:10.1007/978-981-13-5874-6_3.

Sharma A, Verma P, Mathur A, Mathur AK. 2018. Genetic engineering approach using early Vinca alkaloid biosynthesis genes led to increased tryptamine and terpenoid indole alkaloids biosynthesis in differentiating cultures of *Catharanthus roseus*. Protoplasma. 255(1):425–435. doi:10.1007/s00709-017-1151-7.

Shen N, Wang T, Gan Q, Liu S, Wang L, Jin B. 2022. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. Food Chem. 383:132531. doi:10.1016/J.FOODCHEM.2022.132531.

Shi Q, Chen W, Huang S, Wang Y, Xue Z. 2021. Deep learning for mining protein data. Brief Bioinform. 22(1):194–218. doi:10.1093/bib/bbz156.

Sholikhah EN. 2016. Indonesian medicinal plants as sources of secondary metabolites for pharmaceutical industry. Journal of thee Medical Sciences (Berkala Ilmu Kedokteran). 48(04):226–239. doi:10.19106/JMedSci004804201606.

Shrestha A, Pandey RP, Dhakal D, Parajuli P, Sohng JK. 2018. Biosynthesis of flavone C-glucosides in engineered *Escherichia coli*. Appl Microbiol Biotechnol. 102(3):1251–1267. doi:10.1007/s00253-017-8694-6.

Siemon T, Wang Z, Bian G, Seitz T, Ye Z, Lu Y, Cheng S, Ding Y, Huang Y, Deng Z, Liu T, Christmann M. 2020. Semisynthesis of Plant-Derived Englerin A Enabled by Microbe Engineering of Guaia-6,10(14)-diene as Building Block. J Am Chem Soc. 142(6):2760–2765. doi:10.1021/jacs.9b12940.

Simlat M, Stobiecki M, Szklarczyk M. 2013. Accumulation of selected phenolics and expression of PAL genes in carrots differing in their susceptibility to carrot fly (*Psila rosae* F.). Euphytica. 190(2):253–266. doi:10.1007/S10681-012-0836-4.

Simmonds MSJ. 2001. Importance of flavonoids in insect–plant interactions: feeding and oviposition. Phytochemistry. 56(3):245–252. doi:10.1016/S0031-9422(00)00453-2.

Singh A, Menéndez-Perdomo IM, Facchini PJ. 2019. Benzylisoquinoline alkaloid biosynthesis in opium poppy: an update. Phytochemistry Reviews. 18(6):1457–1482. doi:10.1007/s11101-019-09644-w.

Singh P, Arif Y, Bajguz A, Hayat S. 2021. The role of quercetin in plants. Plant Physiol Biochem. 166:10–19. doi:10.1016/J.PLAPHY.2021.05.023. [accessed 2023 Aug 3]. https://pubmed.ncbi.nlm.nih.gov/34087741/.

Siva R, Mayes S, Behera SK, Rajasekaran C. 2012. Anthraquinones dye production using root cultures of *Oldenlandia umbellata* L. Ind Crops Prod. 37(1):415–419. doi:10.1016/j.indcrop.2011.12.027.

Slámová K, Kapešová J, Valentová K. 2018. “Sweet Flavonoids”: Glycosidase-Catalyzed Modifications. International Journal of Molecular Sciences 2018, Vol 19, Page 2126. 19(7):2126. doi:10.3390/IJMS19072126. https://www.mdpi.com/1422-0067/19/7/2126/htm.

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. 2003. Crystal structures of fusion proteins with large-affinity tags. Protein Sci. 12(7):1313–1322. doi:10.1110/PS.0243403. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/12824478/.

Sola RJ, Griebenow K. 2009. Effects of glycosylation on the stability of protein pharmaceuticals. J Pharm Sci. 98(4):1223–1245. doi:10.1002/JPS.21504.

Song Q, Wu Z, Fan Y, Song W, Zhang P, Wang L, Wang F, Xu Y, Wang PG, Cheng J. 2017. Production of homogeneous glycoprotein with multisite modifications by an engineered N-glycosyltransferase mutant. Journal of Biological Chemistry. 292(21):8856–8863. doi:10.1074/jbc.M117.777383. [accessed 2023 Aug 18]. http://www.jbc.org/article/S0021925820427802/fulltext.

Sorensen SO, Van Den Hazel HB, Kielland-Brandt MC, Winther JR. 1994. pH-dependent processing of yeast procarboxypeptidase Y by proteinase A *in vivo* and *in vitro*. Eur J Biochem. 220(1):19–27. doi:10.1111/J.1432-1033.1994.TB18594.X. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/8119286/.

Sosa T, Chaves N, Alias JC, Escudero JC, Henao F, Gutiérrez-Merino C. 2004. Inhibition of Mouth Skeletal Muscle Relaxation by Flavonoids of *Cistus ladanifer* L.: A Plant Defense

Mechanism Against Herbivores. J Chem Ecol. 30(6):1087–1101. doi:10.1023/B:JOEC.0000030265.45127.08.

De Souza Farias SA, Da Costa KS, Martins JBL. 2021. Analysis of Conformational, Structural, Magnetic, and Electronic Properties Related to Antioxidant Activity: Revisiting Flavan, Anthocyanidin, Flavanone, Flavonol, Isoflavone, Flavone, and Flavan-3-ol. ACS Omega. 6(13):8908–8918. doi:https://doi.org/10.1021/acsomega.0c06156. [accessed 2023 Jul 7]. https://pubs.acs.org/doi/full/10.1021/acsomega.0c06156.

Sozer O, Komenda J, Ughy B, Domonkos I, Laczk-Dobos H, Malec P, Gombos Z, Kis M. 2010. Involvement of Carotenoids in the Synthesis and Assembly of Protein Subunits of Photosynthetic Reaction Centers of *Synechocystis* sp. PCC 6803. Plant Cell Physiol. 51(5):823–835. doi:10.1093/PCP/PCQ031. [accessed 2023 Aug 15]. https://dx.doi.org/10.1093/pcp/pcq031.

Spribille R, Forkmann G. 1984. Conversion of dihydroflavonols to flavonols with enzyme extracts from flower buds of *Matthiola incana* R. Br. Zeitschrift fur Naturforschung - Section C Journal of Biosciences. 39(7–8):714–719. doi:https://doi.org/10.1515/znc-1984-7-805. [accessed 2023 Aug 5]. https://www.degruyter.com/document/doi/10.1515/znc-1984-7-805/html.

Stevenson PC, Nicolson SW, Wright GA. 2017. Plant secondary metabolites in nectar: impacts on pollinators and ecological functions. Funct Ecol. 31(1):65–75. doi:10.1111/1365-2435.12761.

Stotz G, Forkmann G. 1982. Hydroxylation of the B-Ring of Flavonoids in the 3′-and 5′-Position with Enzyme Extracts from Flowers of *Verbena hybrida*. Zeitschrift fur Naturforschung -

Section C Journal of Biosciences. 37(1–2):19–23. doi:10.1515/ZNC-1982-1-205/MACHINEREADABLECITATION/RIS. [accessed 2023 Aug 5]. https://www.degruyter.com/document/doi/10.1515/znc-1982-1-205/html.

Suástegui M, Shao Z. 2016. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. J Ind Microbiol Biotechnol. 43(11):1611–1624. doi:10.1007/s10295-016-1824-9.

Sudheeran PK, Ovadia R, Galsarker O, Maoz I, Sela N, Maurer D, Feygenberg O, Shamir MO, Alkan N. 2020. Glycosylated flavonoids: fruit's concealed antifungal arsenal. New Phytologist. 225(4):1788–1798. doi:10.1111/nph.16251. https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.16251.

Sue M, Ishihara A, Iwamura H. 2000. Occurrence and characterization of a UDP-glucose:hydroxamic acid glucosyltransferase isolated from wheat (*Triticum aestivum*) seedlings. Z Naturforsch C J Biosci. 55(9–10):701–7. doi:10.1515/znc-2000-9-1006.

Sukprasansap M, Chanvorachote P, Tencomnao T. 2020. Cyanidin-3-glucoside activates Nrf2-antioxidant response element and protects against glutamate-induced oxidative and endoplasmic reticulum stress in HT22 hippocampal neuronal cells. BMC Complement Med Ther. 20(1):1–12. doi:https://doi.org/10.1186/s12906-020-2819-7. [accessed 2023 Aug 18]. https://bmccomplementmedtherapies.biomedcentral.com/articles/10.1186/s12906-020-2819-7.

Sun W, Sun S, Xu H, Wang Y, Chen Y, Xu X, Yi Y, Ju Z. 2022. Characterization of Two Key Flavonoid 3-O-Glycosyltransferases Involved in the Formation of Flower Color in *Rhododendron delavayi*. Front Plant Sci. 13:863482. doi:10.3389/fpls.2022.863482.

Sutter A, Poulton J, Grisebach H. 1975. Oxidation of flavanone to flavone with cell-free extracts from young parsley leaves. Arch Biochem Biophys. 170(2):547–556. doi:10.1016/0003-9861(75)90150-2. [accessed 2023 Aug 5]. https://pubmed.ncbi.nlm.nih.gov/242261/.

Swallah MS, Sun H, Affoh R, Fu H, Yu H. 2020. Antioxidant Potential Overviews of Secondary Metabolites (Polyphenols) in Fruits. Int J Food Sci. 2020. doi:10.1155/2020/9081686.

Taguchi G, Ubukata T, Hayashida N, Yamamoto H, Okazaki M. 2003. Cloning and characterization of a glucosyltransferase that reacts on 7-hydroxyl group of flavonol and 3-hydroxyl group of coumarin from tobacco cells. Arch Biochem Biophys. 420(1):95–102. doi:10.1016/j.abb.2003.09.027.

Tang K, Pan Q. 2017. Strategies for enhancing alkaloids yield in *Catharanthus roseus* via metabolic engineering approaches. In: Naeem M, Aftab T, Khan M, editors. Catharanthus Roseus: Current Research and Future Prospects. Springer, Cham. p. 1–16. [accessed 2023 Nov 1]. https://link.springer.com/chapter/10.1007/978-3-319-51620-2_1.

Tanifuji R, Minami A, Oguri H, Oikawa H. 2020. Total synthesis of alkaloids using both chemical and biochemical methods. Nat Prod Rep. doi:10.1039/c9np00073a.

Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR. 2003. Proanthocyanidin biosynthesis in plants. Purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. Journal of Biological Chemistry. 278(34):P31647-31656.

doi:https://doi.org/10.1074/jbc.M302783200. [accessed 2023 Aug 5]. http://www.jbc.org/article/S0021925820839568/fulltext.

Tapas AR, Sakarkar DM, Kakde RB. 2008. Flavonoids as Nutraceuticals: A Review. Tropical Journal of Pharmaceutical Research. 7(3):1089–1099. doi:10.4314/tjpr.v7i3.14693.

Tatsis EC, O'Connor SE. 2016. New developments in engineering plant metabolic pathways. Curr Opin Biotechnol. 42:126–132. doi:10.1016/j.copbio.2016.04.012.

Tattini M, Gravano E, Pinelli P, Mulinacci N, Romani A. 2000. Flavonoids accumulate in leaves and glandular trichomes of *Phillyrea latifolia* exposed to excess solar radiation. New Phytol. 148(1):69–77. doi:10.1046/j.1469-8137.2000.00743.x. [accessed 2023 Aug 18]. https://www.cambridge.org/core/journals/new-phytologist/article/flavonoids-accumulate-in-leaves-and-glandular-trichomes-of-phillyrea-latifolia-exposed-to-excess-solar-radiation/0204DA44658D90C525B410B070832830.

Taujale R, Venkat A, Huang LC, Zhou Z, Yeung W, Rasheed KM, Li S, Edison AS, Moremen KW, Kannan N. 2020. Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. Elife. 9. doi:10.7554/ELIFE.54532. [accessed 2023 Jul 7]. /pmc/articles/PMC7185993/.

Tegl G, Nidetzky B. 2020. Leloir glycosyltransferases of natural product C-glycosylation: structure, mechanism and specificity. Biochem Soc Trans. 48(4):1583–1598. doi:10.1042/BST20191140. /biochemsoctrans/article/48/4/1583/225770/Leloir-glycosyltransferases-of-natural-product-C.

Terpe K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 60(5):523–33. doi:10.1007/s00253-002-1158-6.

Tetali SD. 2019. Terpenes and isoprenoids: a wealth of compounds for global use. Planta. 249(1):1–8. doi:https://doi.org/10.1007/s00425-018-3056-x. [accessed 2023 Jul 7]. https://link.springer.com/article/10.1007/s00425-018-3056-x.

Tholl D. 2015. Biosynthesis and biological functions of terpenoids in plants. Adv Biochem Eng Biotechnol. 148:63–106. doi:10.1007/10_2014_295. [accessed 2023 Apr 7]. https://pubmed.ncbi.nlm.nih.gov/25583224/.

Ti H, Zhuang Z, Yu Q, Wang S. 2021. Progress of plant medicine derived extracts and alkaloids on modulating viral infections and inflammation. Drug Des Devel Ther. 15:1385–1408. doi:10.2147/DDDT.S299120. [accessed 2023 Jul 7]. https://doi.org/10.2147/DDDT.S299120.

Tian Y, Fu X, Yang Z, Wang B, Gao J, Wang M, Xu J, Han H, Li Z, Yao Q, Peng R. 2020. Metabolic engineering of rice endosperm for betanin biosynthesis. New Phytologist. 225(5):1915–1922. doi:10.1111/nph.16323. https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.16323.

Timoneda A, Feng T, Sheehan H, Walker-Hale N, Pucker B, Lopez-Nieves S, Guo R, Brockington S. 2019. The evolution of betalain biosynthesis in Caryophyllales. New Phytologist. 224(1):71–85. doi:10.1111/nph.15980. https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.15980.

Tippmann S, Scalcinati G, Siewers V, Nielsen J. 2016. Production of farnesene and santalene by *Saccharomyces cerevisiae* using fed-batch cultivations with *RQ* -controlled feed. Biotechnol Bioeng. 113(1):72–81. doi:10.1002/bit.25683.

Traka M, Mithen R. 2009. Glucosinolates, isothiocyanates and human health. Phytochemistry Reviews. 8(1):269–282. doi:10.1007/s11101-008-9103-7.

Trantas EA, Koffas MAG, Xu P, Ververidis F. 2015. When plants produce not enough or at all: metabolic engineering of flavonoids in microbial hosts. Front Plant Sci. 6. doi:10.3389/fpls.2015.00007.

Turanlı-Yıldız B, Hacısalihoğlu B, Çakar ZP. 2017. Advances in Metabolic Engineering of *Saccharomyces cerevisiae* for the Production of Industrially and Clinically Important Chemicals. In: Old Yeasts - New Questions. InTech.

Twaij BM, Hasan MN. 2022. Bioactive Secondary Metabolites from Plant Sources: Types, Synthesis, and Their Therapeutic Uses. International Journal of Plant Biology 2022, Vol 13, Pages 4-14. 13(1):4–14. doi:10.3390/IJPB13010003. [accessed 2023 Aug 3]. https://www.mdpi.com/2037-0164/13/1/3/htm.

Vasudevan UM, Lee EY. 2020. Flavonoids, terpenoids, and polyketide antibiotics: Role of glycosylation and biocatalytic tactics in engineering glycosylation. Biotechnol Adv. 41:107550. doi:10.1016/j.biotechadv.2020.107550.

Vavitsas K, Fabris M, Vickers C. 2018. Terpenoid Metabolic Engineering in Photosynthetic Microorganisms. Genes (Basel). 9(11):520. doi:10.3390/genes9110520.

Vera C, Illanes A, Guerrero C. 2021. Enzymatic production of prebiotic oligosaccharides. Curr Opin Food Sci. 37:160–170. doi:10.1016/J.COFS.2020.10.013.

Verpoorte R, Memelink J. 2002. Engineering secondary metabolite production in plants. Curr Opin Biotechnol. 13(2):181–187. doi:10.1016/S0958-1669(02)00308-7.

Viñas P, Martínez-Castillo N, Campillo N, Hernández-Córdoba M. 2011. Directly suspended droplet microextraction with in injection-port derivatization coupled to gas chromatography–mass spectrometry for the analysis of polyphenols in herbal infusions, fruits and functional foods. J Chromatogr A. 1218(5):639–646. doi:10.1016/J.CHROMA.2010.12.026.

Wallace RJ. 2004. Antimicrobial properties of plant secondary metabolites. Proceedings of the Nutrition Society. 63(4):621–629. doi:10.1079/PNS2004393.

Waller GR, Nowacki EK. 1978. Alkaloid Biology and Metabolism in Plants. Springer New York, NY. [accessed 2023 Nov 1]. https://link.springer.com/book/10.1007/978-1-4684-0772-3.

Waller GR, Nowacki EK. 1978. The Role of Alkaloids in Plants. In: Alkaloid Biology and Metabolism in Plants. Boston, MA: Springer . p. 143–181. [accessed 2023 Nov 1]. https://link.springer.com/chapter/10.1007/978-1-4684-0772-3_5#:~:text=The%20proposed%20roles%20of%20alkaloids,of%20some%20of%20them%20resemble.

Wang C, Dissing MM, Agerbirk N, Crocoll C, Halkier BA. 2020. Characterization of *Arabidopsis* CYP79C1 and CYP79C2 by Glucosinolate Pathway Engineering in *Nicotiana benthamiana* Shows Substrate Specificity Toward a Range of Aliphatic and

Aromatic Amino Acids. Front Plant Sci. 11:57. doi:10.3389/fpls.2020.00057. https://www.frontiersin.org/article/10.3389/fpls.2020.00057/full.

Wang H, Yang Y, Lin L, Zhou W, Liu M, Cheng K, Wang W. 2016. Engineering *Saccharomyces cerevisiae* with the deletion of endogenous glucosidases for the production of flavonoid glucosides. Microb Cell Fact. 15(1):134. doi:10.1186/s12934-016-0535-2. http://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-016-0535-2.

Wang H, Zhao T, Liu Z, Danzengquzhen, Cisangzhuoma, Ma J, Li X, Huang X, Li B. 2023. The neuromodulatory effects of flavonoids and gut Microbiota through the gut-brain axis. Front Cell Infect Microbiol. 13:1197646. doi:10.3389/FCIMB.2023.1197646/BIBTEX.

Wang J, Qiu Y, Wang X, Yue Z, Yang X, Chen X, Zhang X, Shen D, Wang H, Song J, He H, Li X. 2017. Insights into the species-specific metabolic engineering of glucosinolates in radish (*Raphanus sativus* L.) based on comparative genomic analysis. Sci Rep. 7(1):16040. doi:10.1038/s41598-017-16306-4.

Wang P, Zhao F-J, Kopittke PM. 2019. Engineering Crops without Genome Integration Using Nanotechnology. Trends Plant Sci. 24(7):574–577. doi:10.1016/j.tplants.2019.05.004.

Wang S, Yang C, Tu H, Zhou J, Liu X, Cheng Y, Luo J, Deng X, Zhang H, Xu J. 2017. Characterization and Metabolic Diversity of Flavonoids in Citrus Species. Scientific Reports 2017 7:1. 7(1):1–10. doi:10.1038/s41598-017-10970-2. [accessed 2023 Aug 5]. https://www.nature.com/articles/s41598-017-10970-2.

Wang Z, Dordick JS, Linhardt RJ. 2011. *Escherichia coli* K5 heparosan fermentation and improvement by genetic engineering. Bioeng Bugs. 2(1):63. doi:10.4161/BBUG.2.1.14201. [accessed 2023 Aug 23]. /pmc/articles/PMC3127022/.

Wasson AP, Pellerone FI, Mathesius U. 2006. Silencing the Flavonoid Pathway in *Medicago truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. Plant Cell. 18(7):1617–1629. doi:10.1105/tpc.105.038232.

Waugh DS. 2005. Making the most of affinity tags. Trends Biotechnol. 23(6):316–320. doi:10.1016/J.TIBTECH.2005.03.012. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/15922084/.

Wei W, Wang P, Wei Y, Liu Q, Yang C, Zhao G, Yue J, Yan X, Zhou Z. 2015. Characterization of *Panax ginseng* UDP-Glycosyltransferases Catalyzing Protopanaxatriol and Biosyntheses of Bioactive Ginsenosides F1 and Rh1 in Metabolically Engineered Yeasts. Mol Plant. 8(9):1412–1424. doi:10.1016/J.MOLP.2015.05.010.

Wen W, Alseekh S, Fernie AR. 2020. Conservation and diversification of flavonoid metabolism in the plant kingdom. Curr Opin Plant Biol. 55:100–108. doi:10.1016/J.PBI.2020.04.004.

Wen Z, Zhang ZM, Zhong L, Fan J, Li M, Ma Y, Zhou Y, Zhang W, Guo B, Chen B, Wang JB. 2021. Directed Evolution of a Plant Glycosyltransferase for Chemo- And Regioselective Glycosylation of Pharmaceutically Significant Flavonoids. ACS Catal. 11(24):14781–14790. https://doi.org/10.1021/acscatal.1c04191

Weseler AR, Bast A. 2017. Masquelier's grape seed extract: from basic flavonoid research to a well-characterized food supplement with health benefits. Nutr J. 16(1):1–19. doi:10.1186/s12937-016-0218-1. http://nutritionj.biomedcentral.com/articles/10.1186/s12937-016-0218-1.

Willson ML, Burke L, Ferguson T, Ghersi D, Nowak AK, Wilcken N. 2019. Taxanes for adjuvant treatment of early breast cancer. Cochrane Database of Systematic Reviews. doi:10.1002/14651858.CD004421.pub3.

Wilson SA, Roberts SC. 2014. Metabolic engineering approaches for production of biochemicals in food and medicinal plants. Curr Opin Biotechnol. 26:174–182. doi:10.1016/j.copbio.2014.01.006.

Wink M. 2018. Plant Secondary Metabolites Modulate Insect Behavior-Steps Toward Addiction? Front Physiol. 9. doi:10.3389/fphys.2018.00364.

Winnacker M. 2022. The Terpenes Limonene, Pinene(s), and Related Compounds: Advances in Their Utilization for Sustainable Polymers and Materials. :1–30. doi:10.1007/12_2022_123. [accessed 2023 Jun 23]. https://link.springer.com/chapter/10.1007/12_2022_123.

Witherup KM, Look SA, Stasko MW, Ghiorzi TJ, Muschik GM, Cragg GM. 1990. *Taxus* spp. Needles contain amounts of taxol comparable to the bark of *Taxus brevifolia*: Analysis and isolation. J Nat Prod. 53(5):1249–1255. doi:10.1021/np50071a017.

Woestenenk EA, Hammarström M, Van Den Berg S, Härd T, Berglund H. 2004. His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. J Struct Funct Genomics. 5(3):217–229. doi:10.1023/B:JSFG.0000031965.37625.0E. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/15503425/.

Wu J, Du G, Chen J, Zhou J. 2015. Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*. Sci Rep. 5(13477). doi:10.1038/srep13477.

Wu M, Shen Q, Yang Y, Zhang S, Qu W, Chen J, Sun H, Chen S. 2013. Disruption of YPS1 and PEP4 genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115. J Ind Microbiol Biotechnol. 40(6):589–599. doi:10.1007/S10295-013-1264-8. [accessed 2023 Aug 9]. https://academic.oup.com/jimb/article/40/6/589/5994886.

Xi A, Li C, Zhang F. 2008. Tobacco plants can use nitrogen taken up before mechanical wounding to synthesize nicotine afterwards. Plant Signal Behav. 3(2):87–90. doi:10.4161/psb.3.2.5121.

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. 2003. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science (1979). 299(5605):396–399. doi:10.1126/science.1078540. [accessed 2023 Aug 5]. https://www.science.org/doi/10.1126/science.1078540.

Xie K, Chen R, Li J, Wang R, Chen D, Dou X, Dai J. 2014. Exploring the catalytic promiscuity of a new glycosyltransferase from *Carthamus tinctorius*. Org Lett. 16(18):4874–7. doi:10.1021/ol502380p.

Xie L, Deng Z, Zhang J, Dong H, Wang W, Xing B, Liu X. 2022. Comparison of Flavonoid O-Glycoside, C-Glycoside and Their Aglycones on Antioxidant Capacity and Metabolism during *In Vitro* Digestion and *In Vivo*. Foods. 11(6). doi:10.3390/foods11060882.

Xu Y, Masuko S, Takieddin M, Xu H, Liu R, Jing J, Mousa SA, Linhardt RJ, Liu J. 2011. Chemoenzymatic Synthesis of Homogeneous Ultralow Molecular Weight Heparins.

Science . 334(6055):498–501. doi:10.1126/SCIENCE.1207478. [accessed 2023 Aug 23]. https://www.science.org/doi/10.1126/science.1207478.

Xu Z, Pu X, Gao R, Demurtas OC, Fleck SJ, Richter M, He C, Ji A, Sun W, Kong J, Hu K, Ren F, Song Jiejie, Wang Z, Gao T, Xiong C, Yu H, Xin T, Albert VA, Giuliano G, Chen S, Song Jingyuan. 2020. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biol. 18(1):1–14. doi:10.1186/s12915-020-00795-3. [accessed 2023 Jul 7]. https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00795-3.

Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Saito K. 1999. Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. J Biol Chem. 274(11):7405–11. doi:10.1074/jbc.274.11.7405.

Yang D, Peng M, Yang H, Yang Q, Xu J. 2009. Expression, purification and characterization of *Gloydius shedaoensis* venom gloshedobin as Hsp70 fusion protein in *Pichia pastoris*. Protein Expr Purif. 66(2):138–142. doi:10.1016/J.PEP.2009.03.003.

Yang H, Liu F, Li Y, Yu B. 2018. Reconstructing Biosynthetic Pathway of the Plant-Derived Cancer Chemopreventive-Precursor Glucoraphanin in *Escherichia coli*. ACS Synth Biol. 7(1):121–131. doi:10.1021/acssynbio.7b00256.

Yang J, Roy A, Zhang Y. 2013. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res. 41(Database issue):D1096-103. doi:10.1093/nar/gks966.

Yang J, Zhang Y. 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43(W1):W174-81. doi:10.1093/nar/gkv342.

Yang L, Wen K-S, Ruan X, Zhao Y-X, Wei F, Wang Q. 2018. Response of Plant Secondary Metabolites to Environmental Factors. Molecules. 23(4):762. doi:10.3390/molecules23040762.

Yang W, Chen X, Li Y, Guo S, Wang Z, Yu X. 2020. Advances in Pharmacological Activities of Terpenoids. Nat Prod Commun. 15(3). doi:https://doi.org/10.1177/1934578X20903555. [accessed 2023 Jul 7]. https://journals.sagepub.com/doi/full/10.1177/1934578X20903555.

Yang Y, Lin Y, Li L, Linhardt RJ, Yan Y. 2015. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. Metab Eng. 29:217–226. doi:10.1016/j.ymben.2015.03.018.

Yang Z, Li N, Kitano T, Li P, Spindel JE, Wang L, Bai G, Xiao Y, McCouch SR, Ishihara A, Zhang J, Yang X, Chen Z, Wei J, Ge H, Jander G, Yan J. 2021. Genetic mapping identifies a rice naringenin O-glucosyltransferase that influences insect resistance. The Plant Journal. 106(5):1401–1413. doi:10.1111/TPJ.15244. [accessed 2023 Aug 6]. https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.15244.

Yeon YJ, Park HJ, Park HY, Yoo YJ. 2014. Effect of His-tag location on the catalytic activity of 3-hydroxybutyrate dehydrogenase. Biotechnology and Bioprocess Engineering. 19(5):798–802. doi:https://doi.org/10.1007/s12257-014-0089-2. [accessed 2023 Aug 29]. https://link.springer.com/article/10.1007/s12257-014-0089-2.

Yin R, Han K, Heller W, Albert A, Dobrev PI, Zažímalová E, Schäffner AR. 2013. Kaempferol 3-O-rhamnoside-7-O-rhamnoside is an endogenous flavonol inhibitor of polar auxin transport in *Arabidopsis* shoots. New Phytologist. 201(2):466–475. doi:10.1111/nph.12558. [accessed 2023 Aug 18]. https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.12558.

Yonekura-Sakakibara K, Hanada K. 2011. An evolutionary view of functional diversity in family 1 glycosyltransferases. The Plant Journal. 66(1):182–193. doi:10.1111/J.1365-313X.2011.04493.X. [accessed 2023 Jul 7]. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-313X.2011.04493.x.

Yonekura-Sakakibara K, Higashi Y, Nakabayashi R. 2019. The Origin and Evolution of Plant Flavonoid Metabolism. Front Plant Sci. 10:943. doi:10.3389/fpls.2019.00943. [accessed 2023 Nov 1]. https://www.frontiersin.org/articles/10.3389/fpls.2019.00943/full.

Young CL, Britton ZT, Robinson AS. 2012. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J. 7(5):620–634. doi:10.1002/BIOT.201100155. [accessed 2023 Aug 9]. https://pubmed.ncbi.nlm.nih.gov/22442034/.

Yousefzadeh MJ, Zhu Y, McGowan SJ, Angelini L, Fuhrmann-Stroissnigg H, Xu M, Ling YY, Melos KI, Pirtskhalava T, Inman CL, McGuckian C, Wade EA, Kato JI, Grassi D, Wentworth M, Burd CE, Arriaga EA, Ladiges WL, Tchkonia T, Kirkland JL, Robbins PD, Niedernhofer LJ. 2018. Fisetin is a senotherapeutic that extends health and lifespan. EBioMedicine. 36:18–28. doi:10.1016/j.ebiom.2018.09.015. [accessed 2023 Oct 4]. http://www.thelancet.com/article/S2352396418303736/fulltext.

Yu L, Zhang SD, Zhao XL, Ni HY, Song XR, Wang W, Yao LP, Zhao XH, Fu YJ. 2020. Cyanidin-3-glucoside protects liver from oxidative damage through AMPK/Nrf2 mediated signaling pathway *in vivo* and *in vitro*. J Funct Foods. 73:104148. doi:10.1016/J.JFF.2020.104148.

Zang YX, Kim DH, Park BS, Hong SB. 2009. Metabolic engineering of indole glucosinolates in Chinese cabbage hairy roots expressing *Arabidopsis* CYP79B2, CYP79B3, and CYP83B1. Biotechnology and Bioprocess Engineering. 14(4):467–473. doi:10.1007/s12257-008-0294-y.

Záveská Drábková L, Honys D, Motyka V. 2021. Evolutionary diversification of cytokinin-specific glucosyltransferases in angiosperms and enigma of missing cis-zeatin O-glucosyltransferase gene in Brassicaceae. Scientific Reports 2021 11:1. 11(1):1–12. doi:10.1038/s41598-021-87047-8. [accessed 2023 Jul 7]. https://www.nature.com/articles/s41598-021-87047-8.

Zea CJ, Camci-Unal G, Pohl NL. 2008. Thermodynamics of binding of divalent magnesium and manganese to uridine phosphates: implications for diabetes-related hypomagnesaemia and carbohydrate biocatalysis. Chem Cent J. 2(1). doi:10.1186/1752-153X-2-15. [accessed 2023 Aug 29]. https://pubmed.ncbi.nlm.nih.gov/18627619/.

Zhang C, Freddolino PL, Zhang Y. 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. 45(W1):W291–W299. doi:10.1093/nar/gkx366.

Zhang C, Zhang X, Freddolino PL, Zhang Y. 2023. BioLiP2: an updated structure database for biologically relevant ligand-protein interactions. Nucleic Acids Res. doi:10.1093/nar/gkad630.

Zhang H, Hassan YI, Liu R, Mats L, Yang C, Liu C, Tsao R. 2020. Molecular Mechanisms Underlying the Absorption of Aglycone and Glycosidic Flavonoids in a Caco-2 BBe1 Cell Model. ACS Omega. 5(19):10782–10793. doi:10.1021/acsomega.0c00379.

Zhang M, Yi Y, Gao BH, Su HF, Bao YO, Shi XM, Wang HD, Li FD, Ye M, Qiao X. 2022. Functional Characterization and Protein Engineering of a Triterpene 3-/6-/2′-O-Glycosyltransferase Reveal a Conserved Residue Critical for the Regiospecificity. Angewandte Chemie International Edition. 61(8):e202113587. doi:10.1002/ANIE.202113587. [accessed 2023 Aug 18]. https://onlinelibrary.wiley.com/doi/full/10.1002/anie.202113587.

Zhang P, Li Y, Chong S, Yan S, Yu R, Chen R, Si J, Zhang X. 2022. Identification and quantitative analysis of anthocyanins composition and their stability from different strains of *Hibiscus syriacus* L. flowers. Ind Crops Prod. 177:114457. doi:10.1016/j.indcrop.2021.114457.

Zhang Q, Lin S, Zhang Qingxu, Wang P, Wang H, He H. 2022. Comparative analysis on metabolites and gene expression difference of various allelopathic potential rice under weed stress. Weed Res. 62(5):340–346. doi:10.1111/WRE.12548. [accessed 2023 Aug 15]. https://onlinelibrary.wiley.com/doi/full/10.1111/wre.12548.

Zhang Y, Liu R, Wu X. 2007. The proteolytic systems and heterologous proteins degradation in the methylotrophic yeast *Pichia pastoris*. Ann Microbiol. 57(4):553–560.

doi:10.1007/BF03175354. [accessed 2023 Jan 8]. https://annalsmicrobiology.biomedcentral.com/articles/10.1007/BF03175354.

Zhao X, Li G, Liang S. 2013. Several Affinity Tags Commonly Used in Chromatographic Purification. J Anal Methods Chem. 2013(17). doi:10.1155/2013/581093.

Zheng W, Wuyun Q, Freddolino PL, Zhang Y. 2023. Integrating deep learning, threading alignments, and a multi-MSA strategy for high-quality protein monomer and complex structure prediction in CASP15. Proteins. doi:10.1002/prot.26585.

Zheng W, Wuyun Q, Zhou X, Li Y, Freddolino PL, Zhang Y. 2022. LOMETS3: integrating deep learning and profile alignment for advanced protein template recognition and function annotation. Nucleic Acids Res. 50(W1):W454–W464. doi:10.1093/nar/gkac248.

Zhou K, Qiao K, Edgar S, Stephanopoulos G. 2015. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. Nat Biotechnol. 33(4):377–383. doi:10.1038/nbt.3095.

Zhou S, Jander G. 2021. Engineering insect resistance using plant specialized metabolites. Curr Opin Biotechnol. 70:115–121. doi:10.1016/J.COPBIO.2021.03.005.

Zhu J, Chen C, Zhang B, Huang Q. 2020. The inhibitory effects of flavonoids on α-amylase and α-glucosidase. Crit Rev Food Sci Nutr. 60(4):695–708. doi:10.1080/10408398.2018.1548428.

Zong G, Li J, Gao Y, Fei S, Liu X, Wang X, Shen Y. 2019. Overexpression, purification, biochemical and structural characterization of rhamnosyltransferase UGT89C1 from *Arabidopsis thaliana*. Protein Expr Purif. 156:44–49. doi:10.1016/j.pep.2018.12.007.

APPENDICES

Appendix A: Recipes

Quercetin

3.4 mg Quercetin

Dissolve in 1mL of ethylene glycol monomethyl ether (EGME)

Vortex to dissolve completely

Store at -20°C

Sugar donor (UDP- glucose)

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 -20° C

UDP-14C-glucose (radioactive) – 45,000cpm, 100nmol/10µL, 400µL

6.78mg of UDP-glucose dissolve in 1mL of dH2O = 100nmol/9uL (Stock solution)

40uL of UDP-^{14}C-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

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

<u>YPD-Agar Plate (200mL)</u>

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

<u>Zeocin (100mg/mL) stock solution (1mL)</u>

100mg zeocin

Dissolve in 1mL dH2O

Filter sterilize

Store at -20°C in dark

<u>20% Dextrose (100 mL)</u>

20 g dextrose (α-D-glucose)

Dissolve in 100 mL 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 4° C

10X YNB (1000 mL)

17 g yeast nitrogen base

50 g ammonium sulfate

Dissolve in 1000 mL dH2O

Filter sterilize the solution and store at 4°C

500X Biotin (0.02% Biotin) 100ml

20 mg biotin

Dissolve in100 mL 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

<u>10X - SDS-PAGE buffer (1000mL)</u>

30.28g Tris base

144.14g glycine

10g SDS

Dissolve in 1000ml DH2O

Store at room temperature

<u>1X – SDS-PAGE buffer</u>

100mL of 10X stock

Add 900mL of dH2O

<u>4X SDS-PAGE Sample Dye (10mL)</u>

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)

10X Western buffer (1000mL)

30.28g Tris base

144.14g glycine

Dissolve ingredients in 1000 mL dH2O

Store at 4°C

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

1X TBS buffer (1L)

Add 200mL of 5X TBS to 800mL dH2O

Add 1mL Tween-20

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° C

5-bromo-4-chloro-3-indolyl-phosphate (BCIP) solution (1mL)

42mg BCIP dissolve in 1mL of N, N-dimethylformamide

Store at -20°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

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

<u>200mM PMSF (1mL)</u>

35mg of PMSF dissolved in 1mL methanol

Store at -20°C

<u>50% glycerol (500ml)</u>

250mL of 100% Glycerol

Make up volume to 500mL adding dH2O

Mix well and store at 4° C

<u>Equilibration/Wash buffer (pH=7.5) (500mL)</u>

50mM Sodium phosphate buffer (pH 7.5)

300mM NaCl

5mM BME

10mM Imidazole

1% Triton X-100

0.65 g monobasic sodium phosphate

5.5 g dibasic sodium phosphate

8.75 g sodium chloride

0.34g Imidazole

0.5mL Triton X-100

Dissolve in 450mL dH2O

Adjust pH to 7.5

Make up the volume using dH2O and store at 4°C

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

<u>Wash buffer (pH=7.5) (500mL)</u>

50mM Sodium phosphate buffer (pH 7.5)

300mM NaCl

5mM BME

20mM Imidazole

1% Triton X-100

0.65 g monobasic sodium phosphate

5.5 g dibasic sodium phosphate

8.75 g sodium chloride

0.68g Imidazole

0.5mL Triton X-100

Dissolve in 450mL dH2O

Adjust pH to 7.5

Make up the volume using dH2O and store at 4 °C

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

Elution buffer, pH 7.5 (500 ml)

50mM Sodium phosphate buffer (pH 7.5)

5mM BME

250mM Imidazole

0.65 g monobasic sodium phosphate

5.5 g dibasic sodium phosphate

8.51 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 buffer prior to use

Assay 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

HisPure Resin Cleaning Solution (100ml)

20mM MES, 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 NaOH

Store at 4 °C

<u>HisPur Resin Storing Solution (100 ml)</u>

20% ethanol

20ml of 100X ethanol

Make volume to 100ml using dH2O

Mix well and Store at 4 °C

<u>Buffer A (1000 ml)</u>

25mM Bicine

4.08g Bicine

Dissolve in 900mL dH20

Adjust pH to 8.5 using 1-3M NaOH

Add dH20 to 1000mL

Filter/sterilize into an autoclaved bottle.

Store at 4 °C

<u>Buffer B (1000 ml)</u>

25mM Bicine

1M NaCl

4.08g Bicine

58.44g NaCl

Dissolve in 900mL dH20

Adjust pH to 8.5 using 1-3M NaOH

Add dH20 to 1000mL

Filter/sterilize into an autoclaved bottle.

Appendix B: Methods

SDS-PAGE

A separating gel mixture sufficient to pour two gels was prepared as described in appendix A. The gel electrophoresis chamber was filled with 1X Trix-glycine-SDS buffer and the gels were submerged in the chamber after housing them in the appropriate apparatus. 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 10-180kDa PageRuler Pre-Stained Molecular Weight Marker (ThermoFisher CA #26616) was added to the first well, followed by the protein samples. 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 placed 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. Air bubbles were removed by gently rolling the top 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 the 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.

Appendix C: Supplementary Materials

Table S1 Crystallization Specifications for 7 Plant Glucosyltransferases. Summary of crystallization conditions, substrate specificity, and sequence similarity/identity for seven plant glucosyltransferases, as reported in various studies.

Title – Author – Year	**Name (PDB ID)**	**Substrate**	**Conditions**	**Similarity**	**Identity**
Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from *Medicago truncatula* (Shao et al. 2005)	UGT71G1 (2ACV)	Quercetin Medicagenic Acid Hederagenin	**Method:** Hanging Drop **Protein Concentration:** 5mg/mL in 10mM NaCl, 10mM Tris-HCl, pH 7.5, 5mM BME **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 **Dimensions:** 0.1 x 0.1 x 0.2 mm **Resolution:** 2.0 Å **Notes:** Protein mixed with 5mM UDP-galactose and 5mM quercetin in 2:1 (v/v) ratio, then mixed with equal	65.6%	32.5%

			concentrations reservoir solution. Mixture equilibrated over reservoir solution.		
Structure of a Flavonoid Glucosyltransferase Reveals the Basis for Plant Natural Product Modification (Offen et al. 2006)	VvGT1 (2C1Z)	Cyanidin Kaempferol Quercetin	**Method:** Hanging drop **Protein Concentration :** 4.8mg/mL in 5mM Tris (2-carboxyethyl) phosphine HCl, 10mM Tris–Cl, pH 8.3 diluted 1:1 with mother liquor. **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:** n/a **Time:** 24h **Dimensions:** n/a **Resolution:** 1.9 Å **Notes**: Crystallized in the presence of 20mM UDP-Glc and 0.5mM kaempferol with 4.8mg/mL protein	82.3%	56.7%

Crystal Structure of *Medicago truncatula* UGT85H2 - Insights into the Structural Basis of a Multifunctional (Iso)flavonoid Glycosyltransferase (Li et al. 2007)	UGT85H2 (2PQ6)	Kaempferol Quercetin Genistein Biochanin A Isoliquiritigenin	**Method:** Hanging Drop **Protein Concentration:** 5mg/mL **Reservoir solution:** 20% (w/v) PEG 3350, 0.2M magnesium formate (pH 5.9) **Temperature:** 20°C **Time:** 1-2 weeks **Dimensions:** 0.1 x 0.1 x 0.2 mm **Resolution:** 2.1 Å **Notes:** Crystallized with equal volume reservoir solution	61.7%	25.5%
Characterization and Engineering of the Bifunctional N- and O-glucosyltransferase Involved in Xenobiotic Metabolism in Plants (Brazier-Hicks et al. 2007)	UGT72B1 (2VCE)	Quercetin Kaempferol Myricetin 2,4,5-Trichlorophenol (TCP) 3,4-dichloroaniline (DCA)	**Method:** Hanging drop **Protein Concentration:** 5mg/mL in 5 mM Tris (2-carboxyethyl) phosphine HCl, 10 mM Tris-HCl, pH 8.1 **Reservoir solution:** 23% (w/v) PEG 3350, 0.1M MMT buffer (malic acid, MES, Tris buffer as made up in the PACT	58.7%	28.3%

			screen, Molecular Dimensions Limited), pH 8.0. **Temperature:** n/a **Time:** n/a **Dimensions:** n/a **Resolution**: 1.4 Å **Notes**: "…the observation that bis-Tris propane was also determined in the active site of VvUG reiterates **that Tris-based buffers should be avoided** when determining UGT kinetics and for complex formation (Brazier-Hicks et al. 2007)."		
Crystal Structures of Glycosyltransferase UGT78G1 Reveal the Molecular Basis for Glycosylation and Deglycosylation of (Iso)flavonoids (Modolo et al. 2009)	UGT78G1 (3HBJ)	Quercetin Kaempferol Myricetin Cyanidin Pelargonidin Genistein Biochanin A	**Method**: Hanging Drop **Protein Concentration**: 15mg/mL in 10 mM Tris–HCl (pH 8.0), 20 mM NaCl, and 0.17 mM dodecyl-β-D-maltoside. **Reservoir solution**: 25% (w/v) polyethylene	78.2%	47.5%

			glycol 3350, 0.1 M Bis-Tris, pH 5.5 **Temperature**: 20°C **Time**: n/a **Dimensions**: n/a **Resolution:** 2.1 Å **Notes**: Protein solution contained either 5mM UDP-glucose for co-crystallization with UDP-glucose, or 5mM UDP-glucose, 2.5mM myricetin, and 6µM n-hexadecyl-β-D-maltoside for co-crystallization with UDP-glucose and myricetin, mixed with an equal volume of reservoir solution		
Crystal Structure of UDP-glucose:anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea* (Hiromoto et al. 2013) Structural basis for Acceptor-substrate Recognition of UDP-	UGT78K6 (4REN)	Delphinidin Malvidin Peonidin Pelargonidin Flavonols (trace activity)	**Method:** Hanging Drop **Protein Concentration:** 20mg/mL in 20 mM Tris-HCl (pH 7.4), 200 mM NaCl, and 2 mM CaCl2. **Reservoir solution:** 26%	74.4%	43%

glucose: Anthocyanidin 3-O-glucosyltransferase from *Clitoria ternatea* (Hiromoto et al. 2015)			(w/v) PEG 4000, 0.1M Sodium Citrate Tribasic Dihydrate (pH 5.6), 0.2M Ammonium Acetate **Temperature:** 20°C **Time:** n/a **Dimensions:** 0.05 x 0.05 x 0.5 mm **Resolution:** 1.85 Å **Notes:** After growing crystals to dimensions specified delphinidin, petunidin, and kaempferol in dimethyl sulfoxide or UDP-Glc in the reservoir solution were added to each drop up to concentrations of 0.5 or 20 mM.		
Directed Evolution of a Plant Glycosyltransferase for Chemo- And Regioselective Glycosylation of Pharmaceutically Significant Flavonoids (Wen et al. 2021)	MiCGT-VFAH (7VA8)	Quercetin Kaempferol	**Method:** Hanging Drop **Protein Concentration :** 35mg/mL in 25 mM Tris pH 8.0, 200 mM NaCl, and 5 mM DTT. **Reservoir solution:** 20% (v/v) PEG	39.7%	24.%

			3350, 0.1M Bis-Tris (pH 7.0), 0.2M Calcium Acetate, 4% (v/v) formamide **Temperature:** 20°C **Time:** n/a **Dimensions:** 0.05 x 0.05 x 0.5 mm **Resolution:** **Notes:** WT complex formed by incubating with 10 mM UDP-Glc on ice for 1 hour in the presence of 12% (v/v) ethylene glycol. Equal volumes of WT and reservoir solution were mixed and diffused against 200 μL reservoir solution. VFAH mutant complex was formed by incubating with 10 mM UDP and 2.5 mM quercetin on ice for 1 hour in the presence of 12% (v/v) ethylene glycol. Equal volumes		

			of WT and reservoir solution were mixed and diffused against 200 µL reservoir solution.		

VITA

AARON S. BIRCFIELD

Education: Ph.D. Biomedical Sciences, East Tennessee State University, Johnson City, Tennessee, 2023

M.S. Biology, East Tennessee State University, Johnson City, Tennessee, 2019

B.S. Chemistry, East Tennessee State University, Johnson City, Tennessee, 2016

Professional Experience: Graduate Research Assistant, East Tennessee State University; Johnson City, Tennessee, 2019 – 2023

Biochemistry Lab Coordinator and Instructor, East Tennessee State University, College of Arts and Sciences, 2019 – 2023

Religious Affairs Specialist / Cavalry Scout, Sergeant, Army National Guard, Tennessee, 2017 – Present

Exam Administrator, Uexams, Tri-Cities, TN, 2017 - Present

Publications: Birchfield, Aaron S.; McIntosh, Cecilia A. (2020). "The Effect of Recombinant Tags on *Citrus paradisi* Flavonol-Specific 3-O-Glucosyltransferase Activity." *Plants*, *9*(3): 402. https://doi.org/10.3390/plants9030402

Birchfield, A. S., & McIntosh, C. A. (2020a). Metabolic engineering and synthetic biology of plant natural products – A minireview. *Current Plant Biology*, *24*, 100163. https://doi.org/10.1016/j.cpb.2020.100163

Honors and Awards: Excellence in Teaching Award, School of Graduate and Continuing Studies, East Tennessee State University, 2023

Flash Talk Award, Phytochemical Society of North America, 2022

Best Oral Presentation in Biosciences, Appalachian Student Research Forum, 2022

Best Poster Presentation in Natural Sciences, Appalachian Student Research Forum, 2017 and 2018