In vitro characterization of chalcone isomerase-like (CHIL) in flavanone biosynthesis

BRANDON COREY SALTZMAN

Department of Plant Science

Faculty of Agricultural and Environmental Sciences

McGill University, Macdonald Campus

Montréal, QC

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Supervisor: Mehran Dastmalchi

Committee Member: Jean-Benoit Charron

© Brandon Corey Saltzman August 2023

Table of Contents

ABSTRACT

Isoflavonoids are a class of legume (Fabaceae)-restricted specialized metabolites with roles in plant physiology, including signalling for symbiosis with nitrogen-fixing rhizobia and deterring microbial pathogens. In addition, they are noted for their phytoestrogenic utility in human health, particularly in reducing hormone-dependent changes and diseases. Certain desirable or biologically active isoflavonoid derivatives occur in low abundance physiologically, in complex cocktails, or only under stress conditions. Furthermore, the levels of these compounds *in planta* can be by affected environmental conditions, rendering them highly unpredictable. Therefore, cultivation of plants and extraction thereof can be highly costly and environmental harmful. Sustainable, scalable, and dynamic bioproduction of isoflavonoids can be achieved by reassembling the plant pathway in engineered microbes, such as *Escherichia coli* or *Saccharomyces cerevisiae*. However, when the isoflavonoid pathway is reconstituted *ex planta*, the absence of auxiliary or regulatory players can derail the preferred canonical pathway, reducing the yield of the desired product(s). One such player is chalcone isomerase-like (CHIL), a non-catalytic isoform of the larger chalcone isomerase (CHI) family. CHIL is not involved enzymatically but significantly increases carbon flux through the pathway by rectifying chalcone synthase (CHS) activity. CHS catalyzes a crucial four-step reaction, generating the chalcone scaffold from condensation of *p*-coumaroyl-CoA and three units of malonyl-CoA. Thus, CHS produces either naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone) or, when coupled with chalcone reductase (CHR), the legume-specific isoliquiritigenin (2′,4,4′-trihydroxychalcone). However, CHS is catalytically promiscuous and can, alternatively, produce two lactone by-products, bisnoryangonin (BNY) and *p*-coumaroyltriacetic acid lactone (CTAL), which diverge

due to aberrant cyclization reactions. These reactions supersede the preferred chalcone products *in vitro* or *ex planta*, leading to "carbon loss" from a synthetic biology perspective. CHIL-assisted chalcone biosynthesis is conserved in land plants; however, it is unclear if CHIL has a similar corrective impact on the production of legume-characteristic scaffolds, such as isoliquiritigenin and the downstream isoflavonoids. My research investigates the ability of four CHIL isoforms from a bryophyte (*Marchantia polymorpha*), a lycophyte (*Selaginella moellendorffii*), an angiosperm monocot (*Oryza sativa*), and a legume (*Glycine max*) to facilitate the carbon flux leading to flavanone production. When CHIL isoforms were coupled with CHS *in vitro*, the percentage of CTAL produced was reduced in an isoform-specific manner, with soybean CHIL generating the best titers of the preferred chalcone or flavanone product, depending on the reaction constituents. The exact mechanism of CHIL-associated corrective action on CHS has yet to be elucidated; however, the binding affinity of the former to pathway intermediates has been suggested. CHILs and other CHIs collectively evolved from the fatty acid-binding proteins (FAPs). Soybean CHIL shares only 25.46 % identity with the most recent FAP ancestor. Nevertheless, the overall crystal structures of FAPs, CHILs, and *bona fide* CHIs overlap significantly. The theoretical modelling of structure-function relationships suggests that CHIL activity may be established by mutating six active site residues within soybean FAP3, proximal to the substrate (7 Å distance). Overall, the *in vitro* characterization of CHIL shows that the inclusion of this non-catalytic auxiliary protein can improve yields of valuable chalcone and flavanone products, bringing the commercial bioproduction of downstream health and agrochemicals one step closer. These results also highlight the importance of auxiliary

components in specialized metabolism, especially in engineering heterologous systems for natural product synthesis.

RÉSUMÉ

Les isoflavonoïdes sont une classe de métabolites spécialisés jouant un rôle dans la physiologie végétale, y compris la signalisation de la symbiose avec la rhizobie fixatrice d'azote et la protection face aux pathogènes microbiens. Ils se limitent notamment aux légumineuses (Fabaceae). Leurs propriétés phytoestrogénique peuvent traiter les maladies hormonodépendantes. Quelques isoflavonoïdes sont produits en faible abondance, dans des cocktails complexes, ou des conditions de stress. Par conséquent, des étapes de purification coûteuses et complexes limitent l'efficacité d'extraction, et leurs niveaux *in planta* sont imprévisibles ou impactés par les conditions environnementales. Reconstituer la voie métabolique dans des microbes, comme l'*E. coli* ou la *Saccharomyces cerevisiae,* permettrait d'obtenir une bioproduction fiable et évolutive d'isoflavonoïdes. Cependant, lorsque la voie isoflavonoïde est reconstituée *ex planta*, l'absence de protéines auxiliaires peut faire dérailler la voie canonique privilégiée, réduisant le rendement du produit(s) désiré(s). Un de ces joueurs est chalcone isomérase-*like* (CHIL), un isoforme non catalytique de la famille de chalcone isomérase (CHI). CHIL augmente considérablement le flux de carbone dans la voie en corrigeant l'activité de la chalcone synthase (CHS). CHS catalyse une réaction cruciale en quatre étapes, générant l'échafaudage chalcone du *p*-coumaroyl-CoA et trois unités de malonyl-CoA. Ainsi, CHS produit soit naringenin chalcone (2′,4,4,6′ -tétrahydroxychalcone), soit isoliquiritigénine (2′,4,4′ -trihydroxychalcone) lorsqu'il est associé à chalcone réductase (CHR). Cependant, CHS est catalytiquement promiscueux et peut produire deux sous-produits de lactone, bisnoryangonine (BNY) et l'acide p-coumaroyltriacétique lactone (CTAL), dérivant en raison de réactions de cyclisation aberrantes *ex planta*. Ces réactions mènent à une "perte de carbone'' du

point de vue de la biologie synthétique. La biosynthèse du chalcone, assistée par le CHIL, est conservée dans les plantes terrestres; toutefois, on ne sait pas si CHIL peut promouvoir la production d'échafaudages en aval, caractéristiques des légumineuses, comme les isoflavonoïdes. Ma recherche investigue l'habileté de quatre isoformes CHIL, provenant d'un bryophyte (*Marchantia polymorpha*), d'un lycophyte (*Selaginella moellendorffii*), d'un monocot d'angiosperme (*Oryza sativa*) et d'une légumineuse (*Glycine max*), à faciliter le flux de carbone menant à la production d'isoflavonoïdes. Lorsque les isoformes du CHIL ont été couplés avec CHS *in vitro*, une réduction du pourcentage de CTAL a été observée, fluctuant dépendant des isoformes introduits. En particulier, CHIL du soja a généré les meilleurs titres du produit privilégié (chalcone ou flavanone), selon les constituants de la réaction. Le mécanisme exact des CHIL et CHS n'a pas encore été élucidé; cependant, l'affinité de liaison de ce dernier avec les intermédiaires de voie a été soulevée comme une possibilité. Les CHILs appartiennent à la famille de protéine chalcone isomérase (CHI), qui ont évolué collectivement à partir de protéines de liaison aux acides gras (FAP). CHIL de soja partage une identité d'acide aminé de 25,46 % avec son ancêtre FAP le plus récent. Néanmoins, les structures cristallines globales des FAP, des CHIL, et des CHI *bona fide* se chevauchent considérablement. L'activité du CHIL peut être établie par la mutation de six résidus de site actifs dans soja FAP3, à proximité du substrat (7 Å distance). Ces résultats soulignent l'importance des composants auxiliaires dans le métabolisme spécialisé, en particulier dans l'ingénierie des systèmes hétérologues pour la synthèse de produits naturels. Dans l'ensemble, la caractérisation *in vitro* de CHIL désigne que l'inclusion de cette protéine auxiliaire non catalytique peut améliorer le rendement des produits précieux de chalcone

et de flavanone, stimulant l'éventuelle bioproduction commerciale des produits sanitaires et agrochimiques.

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
2-HID	2-Hydroxyisoflavanone dehydratase
4CL	4-Coumarate-coenzyme A lyase
ACCase	Acetyl-CoA carboxylase
AKR	Aldo-keto reductase
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
BiFC	Bimolecular fluorescence complementation
BNY	Bisnoryangonin
bp	Nucleic acid base pairs
BSA	Bovine serum albumin
C4H	Cinnamate 4-hydroxylase
CHI	Chalcone isomerase
CHIL	Chalcone isomerase-like
CHR	Chalcone reductase
CHS	Chalcone synthase
CoA	Coenzyme-A
COR	Codeinone reductase
CTAL	*p*-Coumaroyltriacetic acid lactone
CTAS	*p*-Coumaroyltriacetic acid synthase
DDC	Duplication, degeneration, and complementation
DIMBOA	2,4-Dihydroxy-7-methoxy-1,4-benzoaxazin-3-one
DMX	Desmethylxanthohumol
DNA	Deoxyribonucleic acid
DSF	Differential scanning fluorimetry
ECL	Enhanced chemiluminescence
EFP	Enhancer of flavonoid production
ER	Endoplasmic reticulum
FAP	Fatty acid-binding protein
FNS	Flavone synthase
HCl	Hydrochloric acid
HCT	Hydroxycinnamoyl transferase
hER	Human estrogen receptor
His	Histidine
HRP	Horseradish peroxidase
HSD	Honestly significant difference
IFS	Isoflavone synthase
IPTG	Isopropyl-β-D-1-thiogalactopryranoside

iTOL	Interactive tree of life
JTT	Jones-Taylor-Thornton method
kDa	Kilodalton
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
MIA	Monoterpenoid indole alkaloid
MUSCLE	Multiple sequence comparison by log-expectation
MYB	Myeloblastosis
NADPH	Nicotinamide adenine dinucleotide phosphate
NISO	Neopinone isomerase
Nod	Nodulation genes/factors
NP	Natural product
OD	Optical density
P450	Cytochrome P450 monooxygenase
PA	Proanthocyanidins
PAL	Phenylalanine ammonia-lyase
PDB	Protein data bank
PKS	Polyketide synthase
PKS III	Polyketide synthase III
PMSF	Phenylmethylsulphonyl fluoride
PR10	Pathogenesis-related 10 proteins
PVDF	Poly(vinylidene) difluoride
RAS	Rosmarinic acid synthase
RNA	Ribonucleic acid
RNAi	RNA interference
ROS	Reactive oxygen species
RPM	Rotations per minute
SDS-PAGE	Sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SPS	Styrylpyrone synthase
STS	Stilbene synthase
TAL	Tyrosine ammonia-lyase
TBST	Tris-buffered saline with Tween-20
TCA	Tricarboxylic acid cycle
T-DNA	Transfer DNA
TT5	Transparent testa 5
UV	Ultraviolet
WGD	Whole-genome duplication
XN	Xanthohumol

ACKNOWLEDGMENTS

First and foremost, I must acknowledge all of the support from my family. To Mum, Dad, Bayla and the more recent additions Chris (and Elyse), thank you for helping, encouraging, guiding and supporting me during my academic career and beyond. If it was not always clear, I love all of you. I would also like to thank my tremendous girlfriend, Evening, with whom I've shared all my struggles and successes from close and from afar. Thank you for everything.

I would not have reached this point without the direction of my incredible supervisor, Dr. Mehran Dastmalchi. During my Master's, Mehran has always strived for excellence and supported me and my project during prosperous and challenging times. Thank you, Mehran, for taking a chance on me as one of your first students. I hope our partnership sets the tone for the rest of our careers.

Thank you to my committee member, Dr. Jean-Benoit Charron, for his valuable insights and comments. To my hilarious lab, Fanfan Li, Audrey Cote, Lee-Marie Raytek, Meha Sharma, Xavier Brasseur-Trottier, and Defne Helvacioglu, each of you not only assisted in aspects of my project but also have optimism and a fresh prospective during the hours (days? weeks? months?) of troubleshooting experiments. Thank you all.

Lastly, I want to extend my gratitude to the Department of Plant Science at McGill University, stay sturdy, and the Natural Sciences and Engineering Research Council of Canada for supporting the Dastmalchi lab.

CONTRIBUTION OF AUTHORS

This thesis is composed of original work in its entirety. It is organized and presented here in the traditional monograph style. The contributions towards each section are as follows: All chapters were written by Brandon Saltzman and Mehran Dastmalchi. Brandon Saltzman completed the experiments outlined in Chapter III except for LC-MS set-up, which Fanfan Li completed. Specific constructs were gifted to this study from Agriculture and Agri-food Canada, the University of New Brunswick, and the University of North Carolina-Wilmington. Brandon Saltzman performed the data analysis.

CHAPTER I: INTRODUCTION

1.1 Introduction

Isoflavonoids are a class of specialized metabolites typically restricted to legumes (Fabaceae). They are derivatives of the phenylpropanoid pathway, whereby phenylalanine is enzymatically converted into *p*-coumaroyl-CoA (Hrazdina & Wagner, 1985). Chalcone synthase (CHS; EC 2.3.1.74) is an enzyme ubiquitous in higher plants within the polyketide synthase (PKS) protein family. The CHS-catalyzed reaction is the first committed step in (iso)flavonoid biosynthesis as the substrates *p*-coumaroyl-CoA and malonyl-CoA are condensed into the product naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone) (Welle & Grisebach, 1987) (**Figure 1**). In legumes, CHS activity can be coupled with chalcone reductase (CHR; EC 2.3.1.170), producing the chalcone intermediate isoliquiritigenin (2′,4,4′-trihydroxychalcone) (Welle & Grisebach, 1987, Welle & Grisebach, 1988). Chalcone scaffolds possess an open-ring configuration which can be closed either spontaneously or via a chalcone isomerase (CHI; EC 5.5.1.6)-catalyzed reaction into flavanone products (2*S*)-naringenin ((2*S*)-4′,5,7-trihydroxyflavanone) and (2*S*)-liquiritigenin ((2*S*)-4′,7-dihydroxyflavanone) (Dixon & Whitehead, 1982). Naringenin is a common precursor for all land plants, initiating many diverging pathways, while the legume-characteristic liquiritigenin can be transformed into isoflavonoids and downstream metabolites.

Isoflavonoids have demonstrated potential in the agricultural and health sectors, including antimicrobials, such as the pterocarpan pisatin, and nutraceuticals, such as genistein (Dixon & Steele, 1999). Unfortunately, metabolite extraction of some isoflavonoid-derived compounds can be complicated due to their occurrence in complex cocktails or in low abundance; alternatively,

de novo chemical synthesis suffers from challenges, such as the potential for complex stereochemistries of products and the need for harsh and costly chemicals or conditions (Nabavi *et al.*, 2020). Therefore, establishing scalable and sustainable strategies for producing plant specialized metabolites is appealing both commercially and environmentally.

Synthetic biology is a burgeoning field, using next-generation-sequencing, automation, gene-editing, and microbial engineering to re-envision life and metabolic pathways as a series of gene-circuits and tunable technologies. Furthermore, engineering plant pathways into heterologous hosts can circumvent costly plant extraction and limited chemical synthesis. Entire plant pathways can be reassembled in hosts, such as *Saccharomyces cerevisiae* (Baker's yeast) and *Escherichia coli.* However, when the canonical pathway for isoflavone biosynthesis is reconstituted *ex planta*, as in many other cases, CHS demonstrates catalytic promiscuity, leading to the production of lactone byproducts bisnoryangonin (BNY) and *p*-coumaroyltriacetic acid lactone (CTAL). The production of these lactone byproducts occurs at the expense of the desired canonical products, thus representing "carbon loss" (Waki *et al.*, 2020).

Chalcone isomerase-like (CHIL) has demonstrated a species-specific ability to rectify the catalytic promiscuity of CHS (Ban *et al.*, 2018, Waki *et al*., 2020). CHIL orthologs across the plant kingdom have demonstrated a conserved strategy to direct metabolic flux towards naringenin and away from lactone byproducts. Thus, CHIL activity may be putatively applied to reconstitute the isoflavonoid biosynthetic pathway in microbial hosts efficiently. However, the utility of CHIL when coupled with CHS and CHR has not yet been reported. CHIL, like other CHI family members, has evolved from fatty acid-binding proteins (FAPs), and it is believed to

affect or guide CHS activity through protein-protein interactions, although the exact mechanism is unknown.

1.2 Research hypotheses

1. CHIL proteins from across the four major phyla of the plant kingdom can equally rectify the catalytic activity of CHS away from byproduct formation toward the desired chalcone and flavanone products. Therefore, each CHIL isoform will reduce the accumulation of lactone byproducts (BNY and CTAL) when coupled with soybean CHS *in vitro.*
2. CHIL activity is conserved when coupled with CHS and CHR, or with CHS, CHR, and CHI; the metabolic flux towards legume-specific products, isoliquiritigenin (chalcone) and liquiritigenin (flavanone) will be enhanced.
3. CHIL function was established through subtle mutations in the active site. Therefore, starting from a FAP protein sequence, "minimal" amino acid substitutions can be identified through *in silico* predictive tools to achieve CHIL chalcone-binding activity.

1.3 Objectives

1. The purification of recombinant proteins for CHS, CHR, and CHI from *Glycine max* (soybean), in addition to CHIL orthologs from *Marchantia polymorpha* (MpCHIL), *Selaginella moellendorffii* (SmoCHIL), *Oryza sativa* (rice) (OsCHIL), and soybean (GmCHIL) using an *E. coli* expression strain (BL21 (DE3)).
2. The characterization of CHIL function with catalytic partners in coupled *in vitro* assays:

a. CHIL and CHS assays to measure the distribution of CTAL, naringenin chalcone, and naringenin.
b. CHIL, CHS, and CHR to assess the production of legume-specific isoliquiritigenin.
c. CHIL, CHS, CHR, and CHI to determine the distribution of flavanones, naringenin and liquiritigenin.

3. Identification of the minimal amino acid substitutions required to establish CHIL function in a FAP ancestor protein, using theoretical modelling based on resolved structures.

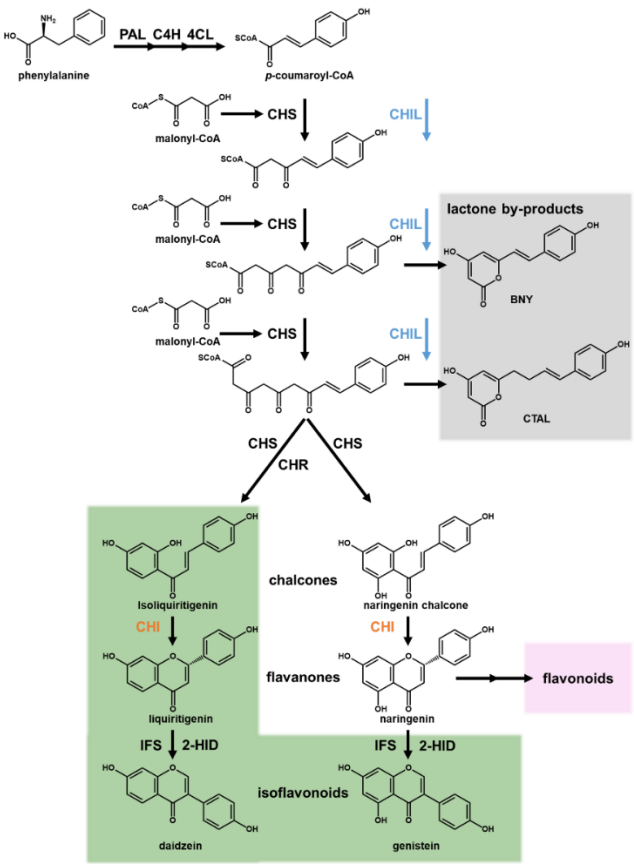


Figure 1. Isoflavonoid scaffold biosynthesis. Chalcone synthase (CHS) catalyzes a four-step condensation reaction using one *p*-coumaroyl-CoA and three malonyl-CoA molecules, yielding a chalcone scaffold. There are two potential chalcone fates, either the canonical naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone), present in all land plants, or the legume-specific isoliquiritigenin (2′,4,4′-trihydroxychalcone), synthesized with chalcone reductase (CHR). Chalcone isomerase (CHI) produces the flavanones (2*S*)-naringenin (2′,5,7-trihydroxyflavanone) or (2*S*)-liquiritigenin (2′,7-dihydroxyflavanone) through a stereospecific ring closure of the chalcone scaffold. Isoflavonoids are produced by a C-2 to C-3 aryl-ring migration and dehydration catalyzed by isoflavone synthase (IFS) and 2-hydroxyisoflavanone dehydrogenase (2-HID). *In vitro,* CHS produces lactone byproducts bisnoryangonin (BNY) or *p*-coumaroyltriacetic acid lactone (CTAL). Chemical structures were drawn using ChemDraw® (20.1.1) software.

CHAPTER II: LITERATURE REVIEW

2.1 Introduction to plant metabolism

Plant metabolism is the complex of biochemical networks responsible for synthesizing and cycling organic compounds. Central (primary) metabolism includes biochemical networks required to maintain organism survival and homeostasis. Large portions of central metabolism are conserved across all eukaryotes and beyond to produce amino acids, nucleic acids, and lipids, in addition to generating free energy, typically as adenosine triphosphate (ATP) (Scossa & Fernie, 2020). The enzymes that catalyze central metabolic pathways were likely established in the earliest life forms and underwent purifying selection, persisting with minimal modifications throughout eukaryotic history (Fani, 2009). As the kingdoms of life diverged, central metabolism developed lineage-specific pathways with unique strategies for maintaining life. For example, in plants, central metabolism is primarily responsible for maintaining cellular functions, such as cellular respiration, photosynthesis, carbon fixation reactions and cell wall biosynthesis. Specialized metabolism is an interconnected network of biochemical transformations through which plants produce compounds that enhance their ability to respond to ecological conditions. Specialized metabolites can be deployed to respond to abiotic and biotic stresses, used as cell signalling molecules, or to conduct plant-microbe interactions. Specialized metabolites were initially defined as “secondary” and, therefore, not required for survival, despite often attributing fitness and advantages (Dixon, 1999b).

Photosynthesis is a fundamental adaptation of plant physiology separate from other eukaryotes, enabling plants to capture light energy from the sun. Photosynthesis itself results

from an endosymbiotic event with a cyanobacterium (Sánchez-Baracaldo & Cardona, 2020). Plants transitioned to terrestrial ecosystems to enhance photosynthetic reactions and thus were exposed to shifting environmental threats, such as desiccation, harmful ultraviolet radiation, and previously unencountered predators. A plant lineage-specific whole genome duplication event seemed to be a second-order effect after the colonization of land 500 million years ago (Weng, 2012). Multiple functional isoforms of critical players in central metabolism were a foundation for functional innovation through natural selection. For example, the primary metabolite shikimate and the specialized metabolite quinate are synthesized by shikimate and quinate dehydrogenases, respectively, with the latter having putatively evolved from a duplicate of the former.

Interestingly, both enzymes belong to the same gene family and can supply similar metabolites for central and specialized metabolic pathways (Carrington *et al.*, 2018). Therefore, central metabolism has enabled plants to survive and thrive in previously uninhabited environments.

2.1.1 The diversity of plant specialized metabolism

Plants produce an extensive catalogue of specialized metabolites that fall into three major classes – alkaloids, phenylpropanoids, and terpenoids, based on common core structures and the precursors used (Wang *et al.*, 2019, Garagounis *et al.*, 2021). All specialized metabolite classes are synthesized using precursors from central metabolism (Aharoni & Galili, 2011). There could be as many as one million specialized metabolites across the plant kingdom, and only a handful of types have been functionally characterized from an eco-physiological perspective (Maeda,

2019). Biological roles, such as defence, communication, and reproduction, have been assigned to specialized metabolites often in response to environmental cues, developmental stage, or in specific tissues.

Specialized metabolites play a vital role in defence against herbivores. For example, the alkaloid DIMBOA (2,4-dihydroxy-7-methoxy-1,4-benzoaxazin-3-one) is accumulated by certain grasses (Brassicaceae), including maize (*Zea mays*) (Wouters *et al.*, 2016) as a deterrent of herbivory. Chewing herbivores, such as the European corn borer (O*strinia nubilalis*), ingest DIMBOA, which inhibits digestion and reduces biomass consumption (Houseman *et al.*, 1992). Isoflavones play a communicative role in plant-microbe interactions, including the control of nodulation. Genistein (4′,5,7-trihydroxyisoflavone) is an isoflavone aglycone produced as part of the phenylpropanoid pathway in legumes and secreted into the rhizosphere to interact with nitrogen-fixing rhizobia (Subramanian *et al.*, 2006). Genistein induces the transcription of bacterial nodulation (*nod*) genes that encode for a lipo-chitin signal molecule (Nod signal). Thereafter, Nod signals are recognized by the roots of a plant host, inducing a signal cascade culminating in the formation of a nodule organ to encapsulate anaerobic rhizobia within the plants (Geurts, 2002, Van de Sande & Bisseling, 1997). Therefore, genistein is a pivotal contributor to initiating the mutualistic “handshake” between legumes and rhizobia, enabling legumes to thrive in nitrogen-depleted environments by indirectly assimilating atmospheric nitrogen. Thirdly, anthocyanins are another phenylpropanoid derivative, which have ubiquitous in the plant kingdom with reproductive and protective roles. Anthocyanins are water-soluble pigments accumulating in coloured organs, such as the flowers, fruits, and seeds their hue (Tanaka, 2008, Liu *et al.*, 2018b). Pollinators, typically insects or birds, are attracted to colourful

flowers and become a vehicle for sexual reproduction (Fenster *et al.*, 2004). The catalogue of diverse structures and functions of specialized metabolites enables plants to interact with and shape their environments.

Specialized metabolites are often limited or unique to a plant clade, tissue, cell type, developmental stage, or following an environmental cue to initiate production. Certain branches of specialized metabolism are restricted to specific taxa. For instance, within the phenylpropanoid class of metabolites, both taxa-specific and broadly conserved metabolites are produced in nature. For example, isoflavonoids are produced primarily in legumes, while flavonoids are generally conserved in land-plants (Dixon, 1999a). Isoflavonoid structures diverge from flavonoid scaffolds through an aryl-ring migration of the B-ring from the C-2 to C-3 and the subsequent dehydration, by a pair of enzymes, isoflavone synthase (IFS; EC 1.14.14.87; CYP93C) and 2-hydroxyisoflavaone dehydrase (2-HID; EC 2.1.1.150) (Dixon, 1999a, Akashi *et al.*, 2005) (**Figure 1**). Decorations or modifications of the scaffold expand the structure into a large legume-specific catalogue of specialized metabolites.

Plant specialized metabolism is tightly controlled spatially and temporally to minimize metabolic expenditure. Biosynthesis is often restricted to specific tissues and developmental stages. Tomato (*Solanum lycopersicum* L.) fruits are a strong example of this dynamic regulation. α-Tomatine is a toxic alkaloid found in tomato stems, leaves, and early stages of fruit development (Fontaine & Irving Jr, 1948). Immature tomato fruits accumulate α-tomatine to dissuade animals from eating fruits with underdeveloped seeds (Tohge, 2013). As the fruits ripen, α-tomatine is turned over in favour of animal attractants for colour and flavour.

Specifically, anthocyanins give the fruit a bright hue signifying to animals that the fruit is flavourful with hydroxycinnamates, most commonly 5-caffeoyl-quinic acid (Tohge, 2013). Thereby, tomatoes tightly control the accumulation of chemical deterrents or attractants to improve the dispersal of mature seeds and to ensure reproductive success.

Conversely, lignins are a phenylpropanoid that is constitutively produced in most tissues. Lignins can account for up to 30 % of organic content in the biosphere and are a primary component of plant cell walls used for plant growth and development (Ralph, 2004, Liu *et al.*, 2018a). Tissue-specific expression of specialized metabolites allows plants to allocate carbon efficiency in developing pathways where they are needed. Metabolites such as lignin are needed throughout most tissues and organs, while others can be more exclusive or restricted.

Subcellular compartmentalization of metabolites is another molecular layer of regulation within plant cells. Biosynthesis of specialized metabolites can occur in specific organelles or areas of the cell through compartmentalization. In specialized metabolism, metabolons commonly form around the endoplasmic reticulum (ER) anchored by cytochrome P450 monooxygenases (P450s) (Ralston & Yu, 2006). For example, the key enzymes involved in the biosynthesis of the isoflavonoid scaffold in soybean localize to the ER despite being soluble (Dastmalchi *et al.*, 2016). There are two distinct advantages of compartmentalization. First, the co-localization of the steps to a biochemical pathway may lead to the formation of multienzyme complexes or metabolons. Metabolons allow enzymes and structural or regulatory elements to aggregate and produce greater metabolic efficacy or output while saving space and energy (Srere, 1985, Ralston & Yu, 2006). Metabolons can improve the rate of turnover, also known as

metabolic flux, of molecules in the pathway by direct substrate and product channelling between enzymes. Within the isoflavonoid pathway, enzymes showed an *in planta* protein-protein interaction with the P450s IFS in addition to cinnamate 4-hydroxylase (C4H; EC 1.14.14.91; CYP73A) (Dastmalchi et al., 2016). The isoflavonoid-specific metabolon may limit competition with intermediates for the overlapping segments of flavonoid metabolism.

Secondly, compartmentalization can protect plant cells by accumulating toxic intermediates in distinct organelles where their cytotoxicity is mitigated or neutralized. In the earlier example, DIMBOA is only released by grasses following herbivory. The inactive form, DIMBOA-glucoside, is stored in the plant vacuole of undamaged cells. Herbivory can damage cell integrity, thus releasing the contents of the vacuole and exposing DIMBOA-glucoside to *β*-glucosidases, which cleave the glucosyl moiety. Toxic DIMBOA aglycones are then consumed by the insect and inhibit digestion (Wouters *et al*., 2016).

Specialized metabolites are often produced only after exposure to a specific environmental stimulus. Such compounds are labelled as phytoalexins or low-molecular-weight and host-induced compounds synthesized in response to pathogen attack or stress (Mansfield, 2000). Wounding stimuli can also induce the production of the phytohormone jasmonic acid (JA), leading to the production of phytoalexins after infection (Felton *et al.*, 1999, Wouters *et al*., 2016, Garagounis *et al*., 2021). There is a trade-off between storing bioactive conjugates or synthesizing compounds from upstream precursors. Plants have arrived at different solutions based on the evolutionary arms race between plants and predators.

2.1.2 Evolution of specialized metabolism through polyploidy

Specialized metabolic enzymes typically evolved from promiscuous catalysts involved in central metabolism. Catalytic promiscuity is the tendency of one enzyme to yield different product fates, either by accepting multiple substrates or by varying the reaction mechanism (Leveson-Gower *et al.*, 2019). Central metabolism is thought to demand constant metabolic flux through pathways that sustain life. As such, the involved enzymes are constrained by selective evolutionary pressure (Nam *et al.*, 2012, Weng & Noel, 2012). However, gene duplication can remove the pressure and allow for innovation in the protein and chemical space.

Plants have undergone several whole genome duplications (WGD) events wherein ancestral central metabolic players could be diverted towards specialized metabolism; at least 5 separate WGD events took in the evolution of *Arabidopsis thaliana* (Arabidopsis) (Jiao *et al.*, 2011). On average, 65 % of annotated plant genes have a duplicate copy, and most are derived from WGD (Panchy *et al.*, 2016). The expansion of plant genomes generated multiple functional copies of critical biosynthetic enzymes in central metabolism. Additionally, WGD events led to expanding regulatory protein families, such as the myeloblastosis (MYB) transcription factors, with 339 proteins documented in Arabidopsis (Stracke *et al.*, 2001). Isoforms of enzymatic players and regulatory proteins were retained in the Arabidopsis proteome to direct metabolic flux towards central metabolism, and additional copies could shift to specialized metabolism through subfunctionalization or neofunctionalization.

The process of enzyme subfunctionalization typically starts with a multifunctional ancestral enzyme that can specialize into one function following gene duplication (**Figure 2a**). For

example, the evolution of rosmarinic acid biosynthesis in *Lamiaceae* herbs (parsley) arose after a duplication event of the gene encoding hydroxycinnamoyl transferase (HCT; EC 2.3.1.133), a BAHD acyltransferase (Landmann *et al.*, 2011). BAHD acyltransferases transfer acyl moieties from a coenzyme A (CoA) thioester to an acceptor molecule in phenolic metabolism (St-Pierre & De Luca, 2000). HCT catalyzes the formation of hydroxycinnamoyl esters and amides; therefore, isoforms are conserved amongst land plants because they play a central role in lignin biosynthesis. Specifically, HCT catalyzes a reaction with shikimic acid and *p*-coumaroyl-CoA as substrates leading to *p*-coumaroyl-shikimate, an intermediate of monolignol biosynthesis (Matsuno *et al.*, 2009, Padmakshan *et al.*, 2022). HCT also weakly catalyzed an alternative reaction with 4-OH-phenylacetic acid and *p*-coumaroyl-CoA as substrates, forming *p*-coumaroyl-4′-hydroxyphenyllactic acid, an intermediate in rosmarinic acid biosynthesis (Landmann *et al*., 2011, Weng & Noel, 2012). Rosmarinic acid has multiple biological functions, including anti-herbivory effects against the genus *Salvia,* as well as antioxidant functions in the cell (Petersen *et al.*, 2009, Dostálek *et al.*, 2016). Following a WGD event, one copy of the HCT specialize into a rosmarinic acid synthase (RAS; EC 2.3.1.140), thus enhancing the first committed step to rosmarinic biosynthesis, while the other copy maintained metabolic flux towards monolignols (Weng, 2012).

Subfunctionalization of enzymes also occurs within large gene families of specialized metabolism. A benefit of multiple gene copies is that refined cellular signals can allow plants to fine-tune the production of metabolites. A significant example of the expansion of a specialized metabolite gene family through subfunctionalization is in the superfamily of *type III polyketide synthase* (*PKS III*) genes. PKS III orthologs broaden their metabolic output by accepting multiple

substrates or by conducting alternative reactions with the same substrate. The CHS group within PKS IIIs, can catalyze reactions using different starter coenzyme A molecules (*p*-coumaroyl-CoA, caffeoyl-CoA, etc.) as substrates (Austin & Noel, 2003). PKS IIIs act as metabolic hubs that lead to a variety of phenylpropanoid products, including flavonoids, isoflavonoids, and cannabinoids (Ferrer *et al.*, 1999, Flores-Sanchez & Verpoorte, 2008). Higher copy numbers also facilitated tissue or organelle differentiation.

Neofunctionalization occurs after gene duplications where alleles gain separate mutations, leading to distinct and sometimes novel functions (**Figure 2b**). Both homologs will persist in the plant genome if the novel function is beneficial. One example of neofunctionalization is the *CHI* gene family, which evolved from genes encoding the non-catalytic fatty acid-binding proteins (FAPs). Specifically, FAP proteins are involved in fatty acid biosynthesis in developing seeds, and ligand-binding assays show FAP affinity for medium chain fatty acids (Ngaki *et al.*, 2012). Conversely, CHI-like proteins developed polyketide binding while losing an affinity for fatty acids (Jez *et al.*, 2000b, Kaltenbach *et al.*, 2018).

Additionally, *bona fide* CHI proteins have evolved the ability to catalyze a chalcone-to-flavanone isomerization forming the flavanone scaffold (Bednar & Hadcock, 1988). Plants typically have retained both isoform types (FAPs and CHILs) (Chen *et al.*, 2009). Therefore, the *CHI* gene family developed additional functions across multiple isoforms through the neofunctionalization of an ancestral *FAP* gene.

Gene duplications leading to subfunctionalization and neofunctionalization are two evolutionary mechanisms that enabled the vast expansion of plant-specialized metabolism. Not all gene duplication events lead to the establishment of novel or fitness-attributing qualities. Pseudogenes with truncated structures or inactive catalytic domains are highly prevalent in plant genomes (Panchy *et al*., 2016) (**Figure 2c**). Nevertheless, the expansion of promiscuous loci gave plants the evolutionary building blocks to develop novel compounds that can facilitate interactions with the environment. From a human (user) perspective, such compounds are often called plant-natural products (NPs) and include examples of medicinal drugs, such as the genistein derivative and anticancer medication, phenoxodiol (Georgaki *et al.*, 2009), health supplements, such as vitamin B12 from *β*-carotene, and agrochemicals including pyrethrum extracted from *Tanacetum cinerariifolium* leaves that disrupts insect nervous systems (Souto *et al.*, 2021).

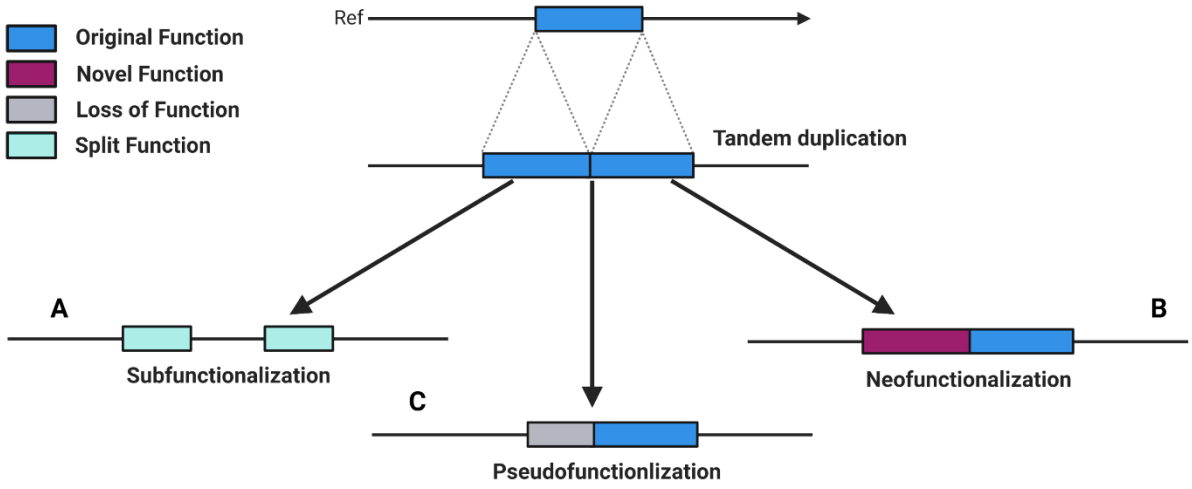


Figure 2. Schematic overview of gene duplication events leading to alternative functions of orthologous genes. A) Subfunctionalization occurs when a gene is duplicated, and the resulting orthologs encode for isoforms catalyzing separate but conserved reactions. **B**) Neofunctionalization occurs when one ortholog encodes for a novel activity. **C**) Pseudofunctionalization is when one gene copy has a loss-of-function mutation. Created with BioRender.com.

2.1.3 Limitations and strategies in the production of specialized metabolites

Natural products and their chemical scaffolds have long been exploited for their uses in medicine. For example, opiate alkaloids isolated from *Papaver somniferum* (opium poppy) are one of the oldest drugs in human history use traditionally for pain treatment (Booth, 2013). Newer discoveries, such as the monoterpenoid indole alkaloid (MIA) vinblastine, isolated from *Catharanthus rosesus*, is a crucial chemotherapy drug (O'Connor & Maresh, 2006). Challengingly, valuable specialized metabolites often serve narrow biological roles within plant tissues; thus, their biosynthesis is strictly regulated. As mentioned previously, tight regulation in specialized metabolite production and turnover maintains flexibility to respond to dynamic environmental or developmental cues, often leading to production of NPs at low abundance.

Traditional plant extraction of NPs is further limited by the availability of arable land, long growth periods, and the costs associated with labour and pesticides (O'Connor, 2012, Lv *et al.*, 2022). For example, in 2021, over 118 thousand hectares of land were used to exclusively grow opium poppy in the United States alone (Board, 2023). Additionally, industrial farming for NPs is susceptible to environmental factors such as pests, diseases, and climate (Bradsher, 2014). Limitations in the bioavailability of valuable compounds and the significant investment of time, energy, and capital make large-scale cultivation of plants for specialized metabolites challenging, often leading to unreliable supply and high costs that cannot meet growing demands.

Currently, many manufacturers produce specialized metabolites *de novo* or semi-synthetically; however, this approach is limited by byproduct formation and the complex stereochemistries of NPs, leading to racemic mixtures. For example, the complex

stereochemistry of vinblastine precludes total chemical synthesis commercially (O'Connor & Maresh, 2006, O'Connor, 2012). On the other hand, most poppy-derived opioids are semi-synthetic, as the plant NP thebaine is extracted from the plant chemically converted into higher-value compounds, such as codeine or hydrocodone (Galanie *et al.*, 2015). Nevertheless, semi-synthetic production too suffers from drawbacks including the need for toxic reagents. In the case of opioids, the conversion of morphine requires a toxic methylating agent, trimethylphenyl ammonium ethoxide, and produces carcinogenic byproducts (Symons *et al.*, 2009).

Metabolic engineering aims to circumvent the challenges associated with traditional plant extraction and partial or complete chemical synthesis of NPs. Cutting-edge 'omics and systems biology is use pathway elucidation, which are then reconstituted in tractable microbial hosts (Chemler & Koffas, 2008, Stovicek *et al.*, 2017, Nabavi *et al*., 2020, Sajid *et al.*, 2021). Microorganisms, such as *E. coli* and Baker's yeast, are excellent hosts for the heterologous expression of whole metabolic pathways through the introduction of enzymes, auxiliary proteins, the tuning of metabolic flux, and metabolic rewiring to deliver low-cost precursors for downstream chemical synthesis or biomanufacturing. *In silico* predictive tools are often utilized to gauge the outcome of gene expression, including predictive peptide localization, peptide anchors, and identification of previously uncharacterized non-catalytic auxiliary proteins through gene co-expression analyses. Once candidates are identified functionally characterized, pathways can be reassembled appropriately in heterologous hosts. The next hurdle is to optimize product yield or titre (Malla *et al.*, 2020, Liu *et al.*, 2021, Liu *et al.*, 2022).

There have been many attempts to reconstitute diverse metabolic pathways in heterologous hosts, each with limitations and challenges (Dixon, 1999b, Ververidis *et al.*, 2007, Dastmalchi *et al.*, 2019, Sajid et al., 2021, Utomo *et al.*, 2021a, Utomo *et al.*, 2021b). One example is the complete reconstitution of the MIA vindoline pathway in yeast which requires cytochrome P450 enzymes, each requiring co-factors, electron donors, and chaperones to catalyze the reactions or localize to distinct subcellular compartments (Gao *et al.*, 2022). Through 54 genetic modifications, including 34 heterologous plant genes and endogenous yeast metabolism rewiring to remove bottlenecks and increase the supply of vinblastine precursors catharanthine and vindoline (Zhang *et al.*, 2022). A second example is the complete production of the opioid compounds thebaine and hydrocodone in yeast, starting from sugars (Galanie *et al*., 2015). In both cases, the optimized titres are significantly less than the yield from commercial extraction. Currently, these efforts have not reached commercial viability.

Limitations to the capacity for producing NPs using synthetic biology extend beyond identifying the key enzymatic steps involved in a specific pathway. Introducing plant enzymes outside their native cellular context can lead to unexpected constraints due to uncharacterized dynamics within hosts, unknown players in endogenous metabolic flux, and other limitations for proper metabolic assembly (Dastmalchi, 2021). For example, when morphine biosynthesis is reconstituted in yeast, alternative alkaloid byproducts, neopine and neomorphine are produced at the expense of morphine and codeine (Thodey *et al.*, 2014). Codeinone reductase (COR; EC 1.1.1.247) catalyzes the reduction of codeinone to codeine during morphine biosynthesis in addition to the neopine byproduct (Bomati *et al.*, 2005, Dastmalchi *et al.*, 2018). However neopinone isomerase (NISO) is an auxiliary protein that binds to neopinone during opiate

alkaloid biosynthesis and lowers the activation energy required to form codeinone, while limiting neopine and neomorphine byproducts (Dastmalchi *et al*., 2019). This example highlights the role of auxiliary proteins in mitigating carbon loss when using synthetic biology platforms.

Once engineered stains a perfected, the large-scale fermentation or culture growth is relatively inexpensive and can be sustainable "bio-foundries" capable of supplying complex, specialized metabolites from cheap carbon sources such as glucose (Keasling, 2010, Sajid *et al*., 2021). Academically speaking, heterologous hosts are also valuable tools to characterize novel enzymes or reactions. For example, expression of the precursors of the taxol biosynthetic pathway in yeast allowed researchers to resolve the missing steps (Chemler & Koffas, 2008). Furthermore, engineered microbes can generate new-to-nature modifications of plant-derived metabolites by installing non-natural enzymes or reaction steps (Chemler *et al.*, 2010). To that end, a thorough understanding of all the factors affecting a metabolic network is required to reconstitute a pathway in a heterologous host. This thesis focuses on the effect of one auxiliary protein, chalcone isomerase-like (CHIL), on the synthesis of isoflavonoid precursors *in vitro*.

2.2 The phenylpropanoid pathway

Phenylpropanoids are the largest class of specialized metabolites, with more than 8000 structures, include lignins, flavonoids, anthocyanins, and isoflavonoids. Phenylpropanoids and other polyphenols perform multiple functions, from acting as or enhancing the production of photosynthetic pigments, improving nutrient uptake from the soil, regulating cellular growth, repelling herbivores, and maintaining REDOX homeostasis (Dixon *et al.*, 2002, Sharma *et al.*, 2019, Pratyusha & Sarada, 2022). Metabolic flux is directed towards the production of specific phenylpropanoid derivatives partly by regulating the gene expression of enzymes. MYB transcription factors partially regulate plant development and stress tolerance with phenylpropanoid metabolites. For example, anthocyanin biosynthesis is promoted in Arabidopsis during flower development and in response to oxidative stress by *AtMYB75* and *AtMYB90* (Mondal & Roy, 2018). Conversely, *AtMYB4* is a repressor of anthocyanin biosynthesis (Yan *et al.*, 2021).

In plants, the aromatic amino acids phenylalanine and tyrosine are the precursors for the phenylpropanoid pathway. Three enzymatic steps are required to transform phenylalanine into a CoA-conjugated molecule (Ferrer *et al.*, 2008). The first committed step is a spontaneous elimination reaction catalyzed by phenylalanine ammonia-lyase (PAL; EC 4.3.1.24), which cleaves the carbon-nitrogen bond of phenylalanine to form the phenylpropanoid skeleton, cinnamic acid (Koukol & Conn, 1961). Interestingly, some species of fungi, bacteria, and some plants (e.g., Sorghum) possess tyrosine ammonia-lyases (TAL; EC 4.3.1.23) that use tyrosine as a starting substrate (Ferrer *et al*., 2008). Next, C4H, a cytochrome P450-dependent hydroxylase,

introduces a hydroxyl group to the phenyl ring of cinnamic acid, producing to 4-coumaric acid from *trans*-cinnamic acid (Blount *et al.*, 2000). Subsequently, the carboxyl group of 4-coumaric acid is activated by the formation of a thioester bond with CoA by 4-coumarate-coenzyme A lyase (4CL; EC 6.2.1.12) to form hydroxycinnamates-CoA easters, such as *p*-coumaroyl-CoA (Ehlting *et al.*, 2001). The intermediate 4-hydroxycinnamoyl-CoA forms the common branch point for the previously mentioned pathways. Certain phenylpropanoid intermediates can be siphoned toward other biosynthetic pathways, including pathways for benzoic acids, salicylic acid, and coumarins (Ferrer *et al*., 2008).

2.2.1 Flavonoids and isoflavonoids

Flavonoids are the most common and structurally diverse class of phenylpropanoid metabolites present throughout land plants (Mouradov & Spangenberg, 2014). The 15-carbon backbone consists of two benzene ring structures (ring A and ring B) connected by a 3-carbon chain (C6-C3-C6) (Steele *et al.*, 1999). Flavonoid biosynthesis likely began in microalgae but flourished following the colonization of land by plants and the WGD events that followed 450 million years ago (Weng, 2012, Goiris *et al.*, 2014, Davies *et al.*, 2020). Flavonoids are predominant in angiosperm flowers and fruits because anthocyanin pigments attract pollinators (Fenster *et al*., 2004). Flavonoids are also valuable tools for chemical defence and stress response.

Modifications to the flavonoid skeleton can alter the biological function of flavonoids and can be species-dependent using methyltransferases, isomerases, reductases or cytochrome P450 enzymes. For example, flavone synthase I (FNS1; EC 1.14.20.5), a 2-oxoglutarate-dependent

dioxygenase, appears to be constrained to the Apiaceae family and catalyzes the formation of the flavone apigenin from the flavonoid skeleton (Martens *et al.*, 2001). Conversely, FNSII (CYP93B16; EC 1.14.19.76) is more broadly distributed amongst land plants and catalyzes the formation of 7,4′-dihydroxyflavanone (Martens & Mithöfer, 2005, Fliegmann *et al.*, 2010). Decorations to the flavonoid skeleton include hydroxylation, methylation, acetylation, or glycosylation. Flavonoids are often methylated to improve stability and bioactivity within plant cells (Koirala *et al.*, 2016). For example, *O*-methylation of naringenin ((2*S*)-4′,5,7-trihydroxyflavanone) in rice by naringenin 7-*O*-methyltransferase (EC 2.1.1.232) produces the defence compound and phytoalexin sakuranetin that accumulates following UV stress and fungal infection (Kodama *et al.*, 1992, Rakwal *et al.*, 2000). Decorated flavonoid compounds can also serve as valuable NPs from a user perspective. For instance, sakuranetin has shown the ability to inhibit tumorigenesis and cancer cell proliferation in mammalian models (Le Bail *et al.*, 1998).

Isoflavonoids are primarily synthesized by leguminous plants and structurally distinct from flavonoids, as they retain the common C6-C3-C6 flavonoid skeleton with an aryl-ring migration of the B-ring from the C-2 to C-3 position. It is noteworthy that there are over 130 examples of isoflavonoid compounds extracted from non-legume angiosperms, actinomyces, and fungi (Dixon & Steele, 1999, Lapčík, 2007, Veitch, 2007, Wang *et al.*, 2020). The isoflavone aglycones, genistein and daidzein, constitute the two major scaffolds from which thousands of isoflavonoids are derived (Dixon & Paiva, 1995, Dixon & Steele, 1999, Dixon *et al.*, 2002). Isoflavonoids play two predominant roles in plant-environment interactions; first, in their communicative function to induce symbiosis with nitrogen-fixing bacteria, and second, as defence molecules in response to pathogens and abiotic stress.

In the first role, isoflavonoids act as signalling molecules. Plants require nitrogen to maintain central metabolic pathways, including amino acid and nucleic acid biosynthesis. Atmospheric nitrogen remains inaccessible to plants, typically relying on soil-bound nitrogenous compounds. Legumes have developed symbioses with nitrogen-fixing rhizobia to colonize nitrogen-depleted soils (Van de Sande & Bisseling, 1997, Wais *et al.*, 2002). Isoflavonoids are key factors that induce the transcription of bacterial nodulation genes and the subsequent mutualistic rhizobia infection into the developing root nodule (Subramanian *et al.*, 2006). The communicative role of genistein in initiating the formation of plant-rhizobia interactions has been previously described (**2.1.1**). An isoflavonoid-mediated symbiosis between legumes and rhizobia indirectly allows plants to acquire nitrogen from the atmosphere.

Isoflavonoid aglycones and their more complex derivatives, such as pterocarpans, serve as defence compounds following pathogen attacks. Pterocarpans are the second largest group of isoflavonoids synthesized from daidzein and include compounds such as glyceollin, medicarpin, and maackiain which are rapidly synthesized in surrounding healthy tissues following microbial infection (Jiménez-González *et al.*, 2008). In soybean, silencing *IFS* and *CHS* gene expression led to an increased susceptibility to the fungal pathogen *Fusarium virguliforme* because of reduced production of daidzein, the isoflavone aglycone and a precursor to the pterocarpan glyceollin (Lygin *et al.*, 2010). Additionally, in *Cicer arietinum* L. (chickpea), medicarpin and maackiain inhibit the germination of *Fusarium oxysporum* (Stevenson *et al.*, 1997). Thus, the synthesis and release of pterocarpans can prevent fungal infections. The potential utility of (+)-medicarpin as a "green pesticide" has been identified. (+)-Medicarpin isolated from *Colletotrichum phomoides*-infected *Medicago sativa* plants showed elevated antifungal activity

towards *Cladosporium cladosporoides,* superior to a preferred fungicide, Benlate® (Bandara *et al.*, 1989).

Isoflavonoids are renowned for their bioactivity and benefits for human health. Long-term clinical studies have shown that diets containing high isoflavonoid content can reduce the risk of chronic heart disease, diabetes, menopause syndrome, and hormone-related cancers (Beck *et al*., 2005; Sohn *et al*., 2021). Flavonoids have also demonstrated antioxidant and anti-inflammatory health benefits; however, the chemical structures of isoflavonoids have additional utility due to the structural similarities with estrogen hormones, specifically 17-*β*-estradiol (Dixon & Steele, 1999, García-Lafuente *et al.*, 2009, Sarian *et al.*, 2017, Brickman *et al.*, 2023). The human estrogen receptors (hER*α*/*β*) have a substantial affinity for the isoflavone aglycones, which can act as weak estrogen agonists or antagonists, depending on the levels of endogenous estrogens (Kuiper *et al.*, 1998). Isoflavonoids have been used to treat hormone-related cancers, potentially by mitigating receptor disruption away from extremes of function. Specifically, phenoxodiol, a genistein derivative, is an anti-cancer drug used to treat breast cancers, working through interaction with hER*β* (Yang *et al*., 2019).

2.2.2 Overview of flavanone and isoflavonoid biosynthesis

Flavanone synthesis begins with the precursor molecules *p*-coumaroyl-CoA and malonyl-CoA (**Figure 1**). Malonyl-CoA provides a 2-carbon acetate unit for growing polyketide and fatty acid chains. Malonyl-CoA is produced from acetyl-CoA and bicarbonate in the plastid by acetyl-CoA carboxylase (ACCase; EC 6.4.1.2) (Harwood, 1996, Schmid & Ohlrogge, 2002). CHS catalyzes a four-step condensation reaction of three units of malonyl-CoA onto *p*-coumaroyl-CoA,

followed by an intramolecular cyclization to form the chalcone scaffold (Ferrer *et al*., 1999). There are two chalcone starting points, naringenin chalcone, which is synthesized in all land plants, and the legume-specific isoliquiritigenin (2′,4,4′-trihydroxychalcone) when CHS activity is coupled to CHR. Chalcones undergo a stereospecific ring closure either spontaneously or by CHI to yield the flavanones; (2*S*)-naringenin and (2*S*)-liquiritigenin (4′,7-dihydroxyflavanone) from naringenin chalcone and isoliquiritigenin, respectively.

The key branch point towards isoflavonoid biosynthesis is catalyzed by IFS as discussed briefly in section **2.1.1**. IFS produces the intermediates 2,5,7,4′-tetrahydroxyisoflavanone from naringenin or 2,7,4′-trihydroxyisoflavanone from liquiritigenin (Steele *et al*., 1999, Jung *et al.*, 2000). Subsequently, either spontaneously or with the action of 2-HID the intermediates are dehydrated to produce a double bond between C-2 and C-3 (Akashi *et al*., 2005). The isoflavone aglycones, genistein (5,7,4′-trihydroxyisoflavone) and daidzein (7,4′-dihydroxyisoflavone) are produced through this two-step reaction initiated with naringenin and liquiritigenin, respectively (**Figure 1**).

Isoflavonoid and flavonoid metabolism are in constant competition for flavanone intermediates. In plants, P450 enzymes are often membrane-bound proteins that are anchored to the ER, with an N-terminal hydrophobic stretch of amino acids forming an α-helix (Neve & Ingelman-Sundberg, 2008). The ER-anchored P450 enzymes, IFS and C4H, can act as metabolic hubs by directing metabolic flux towards isoflavonoid biosynthesis by aggregating the necessary soluble enzymes. It has been demonstrated that soybean IFS and C4H interact with CHS, CHR, and CHI, in addition to each other, to form a metabolon at the ER (Dastmalchi *et al*., 2016).

Furthermore, the upstream components, PAL and 4CL also associate with part of the complex. Together, the metabolon is proposed to enhance metabolic flux through the pathway. Daidzein is the exclusive precursor to the pterocarpans, which are produced immediately following fungal infection (Jiménez-González *et al*., 2008). Thus, it is reasonable that rapid biosynthesis of isoflavonoids should be enhanced based on environmental requirements.

Synthetic biologists utilize the understanding of isoflavonoid biosynthesis, protein-protein interactions, and distribution of metabolic flux to reconstitute the production of aglycones in heterologous hosts. Effective reconstitution of the isoflavonoid pathway in heterologous hosts such as, *E. coli*, requires developing truncated but functional IFS enzymes without membrane-bound domains, or directly integrating the complete chassis of enzymes into hosts, such as yeast, with endosomal membranes. Specifically, two recent publications in *Nature* and *ACS biochemistry* showcased the possibility of producing daidzein and genistein titres of 85.4 mg/l and 60.8 mg/l using glucose as a precursor (Liu *et al*., 2021, Liu *et al*., 2022). These studies have focused on identifying the optimal genes for producing the isoflavone aglycones, specifically *CHS, CHR, CHI,* and *IFS*. The following section is written with a focus on each enzyme involved in flavanone biosynthesis.

2.2.3 Chalcone synthase

Ubiquitously present in land plants, CHS is part of the largest PKS gene family, precisely type III PKSs. PKS IIIs encompass a broad class of synthases with products across many metabolic classes. For example, stilbene synthase produces small phenylpropanoid phytoalexins; acridone synthase produces acridone alkaloids in *Citrus microcarpa*, and bibenzyl synthase produces

bibenzyls in *Dendrobium christyanum* (orchid) (Jeandet *et al.*, 2010, Choi *et al.*, 2020, Chen *et al.*, 2022). PKS IIIs share similar strategies for generating polyketide structures by loading a starter CoA molecule onto an active cysteine residue site, followed by successive decarboxylation condensation reactions to extend the polyketide chain (Schröder *et al.*, 2002).

Chalcone scaffolds are produced by CHS in a four-step reaction. They first utilize its acetyltransferase activity to load *p*-coumaroyl-CoA onto the active cysteine residue (Cys164), followed by a decarboxylase activity that activates malonyl-CoA. The *β*-keto synthase domains tether to growing polyketide chains via a thioester linkage and catalyze the sequential incorporation of two carbon-acetate units into a growing polyketide (Chen *et al.*, 2006). Thirdly, a condensation reaction couples three acetyl anion units to the growing polyketide. Finally, a cyclase activity generates the final product, naringenin chalcone (Hinderer, 1985). CHS is a homodimer composed of 42 kilodaltons (kDa) monomers with two distinct binding cavities (Austin & Noel, 2003).

The core catalytic machinery within CHS is a trio of amino acid residues, Cys164, His303, and Asn336, that are positioned within the CHS active site and aid in multiple decarboxylation and condensation reactions (Ferrer *et al*., 1999). The CHS active site contains three binding cavities: a coumaroyl-binding pocket, a CoA-binding tunnel, and a cyclization pocket. Cys164 is the acting nucleophile within the active site. Additionally, His303 acts as a general base or anchor for the polyketide through hydrogen bonding. Phe215 and Phe265 are two critical gate-keeping residues that limit water within the active site when accommodating substrates (Austin

& Noel, 2003, Ferrer *et al*., 2008). CHS is non-competitively inhibited by flavonoid pathway products, naringenin and naringenin chalcone (Hinderer, 1985).

Functional CHS isoforms are present in all land plants, but the expression of the downstream enzymes varies significantly between plant taxa; thus, most plants maintain multiple copies of CHS genes that can be expressed across tissues, in different organelles, or based on environmental cues (Austin & Noel, 2003). Specifically, two WGD events in soybean formed at least 14 CHS and CHS-like isoforms across eight different chromosomes (Kumaran *et al.*, 2018). In contrast, Arabidopsis only contains a single *CHS* gene (Burbulis *et al.*, 1996). Soybean CHS isoforms are differentially expressed and display functional divergence. Most *GmCHS* genes are highly expressed in leaf tissues; however, *GmCHS6, GmCHS7, GmCHS8, GmCHS10,* and *GmCHS11* were abundant in root tissues. Furthermore, *GmCHS7* and *GmCHS8* are responsible for isoflavonoid accumulation in developing seed pods (Chennupati *et al.*, 2012). The subfunctionalization by multiple CHS gene copies may allow fine-tuning of soybean chalcone flux.

As in many cases in specialized metabolism, PKSs have some degree of catalytic promiscuity (Austin & Noel, 2003). Some CHS and CHS-like proteins can accept alternative starter-CoA molecules, such as cinnamoyl-CoA and caffeoyl-CoA, allowing for a greater diversity of chalcone scaffolding (Schröder, 1999). For example, CHS can accept a reduced *p*-coumaroyl-CoA structure (*p*-dihydrocoumaroyl-CoA), forming the scaffold to a variety of dihydroxychalcone derivatives, such as phloretin (Ibdah, 2017). Additionally, the target size of polyketide chains can vary between CHS isoforms, ranging from triketide to decaketide

structures (Jez *et al.*, 2000a). *In vitro,* the catalytic promiscuity of CHS can be seen as the production of lactone byproducts BNY and CTAL at successive additions of malonyl-CoA molecules (**Figure 1**).

Interestingly, BNY is classified as a styrylpyrone typically synthesized in fungal specialized metabolism. In fungi and some plant species, BNY is synthesized by styrylpyrone synthase (SPS), a PKS that condenses one molecule of hydroxycinnamoyl-CoA with two molecules of malonyl-CoA (Wu, 2022). Styrylpyrones, such as hispidin, are suggested to have a functional role akin to flavonoids in medicinal fungi such as *Phellinus linteus*. Styrylpyrones appear to be used by medicinal fungi as antioxidant pigments and for cellular signalling. Additionally, researchers are investigating if styrylpyrones are partially responsible for the health-promoting effects of medicinal mushrooms (Lee & Yun, 2011). Interestingly, despite the functions of styrylpyrones in fungi, most plant species do not produce BNY in nature, even with functional isoforms of CHS. Therefore, this byproduct must be limited *in planta.*

2.2.4 Chalcone reductase

In legumes, CHR couples with CHS and acts at a biochemical branching point between canonical tetrahydroxy-naringenin chalcone biosynthesis in land plants and the formation of isoliquiritigenin in legumes (**Figure 1**) (Bomati *et al*., 2005). Isoliquiritigenin is the exclusive chalcone precursor to certain isoflavonoids, such as the aglycone daidzein, glyceollins, and the downstream pterocarpans (Schopfer *et al.*, 1998, Steele *et al*., 1999). RNAi silencing of *CHR* transcripts in soybean led to decreased isoflavonoid production and increased susceptibility to infection by *Phytophthora sojae* (Graham *et al.*, 2007).

Belonging to the aldo/keto reductase (AKR) superfamily, specifically the AKR4 family, CHR is amongst several members involved in specialized metabolism, including the alkaloid biosynthetic COR (Bomati *et al*., 2005, Dastmalchi *et al*., 2018). The AKR superfamily of enzymes is monomeric with an (α/β) 8-barrel and NADPH oxidoreductase function (Mindnich & Penning, 2009). All CHR isozymes contain a catalytic tetrad of Asp53, Tyr58, Lys87, and His120 and predominantly have hydrophobic and aromatic residues that line the unoccupied entrance to the active site (Welle & Grisebach, 1988, Bomati *et al*., 2005). A soybean genome-wide search revealed 14 CHR isoforms; however, three isoforms are missing the catalytic tetrad and are suggested to be inactive (Sepiol *et al.*, 2017). The genes encoding these isoforms are differentially transcribed in specific tissues following elicitation, including *GmCHR14A*, which is highly expressed in root tissues (Sepiol *et al*., 2017).

Soybean cell suspension cultures were first used to purify CHR, where it was demonstrated that CHR works in combination with CHS using the cofactor NADPH to form isoliquiritigenin and naringenin chalcone (Welle & Grisebach, 1987). CHR does not reduce naringenin chalcone but instead likely acts on an intermediate of the CHS reaction, likely *p*-coumaroylcyclohexantrione (Bomati *et al*., 2005). CHR is recruited to the endoplasmic reticulum along with CHS and other enzymatic components as part of the (iso)flavonoid metabolon where it interacts with CHS (Dastmalchi *et al*., 2016, Mameda *et al.*, 2018). However, it is unlikely that CHR can interact with a substrate buried between the CHS homodimer active site, but an *in vivo* mechanism of substrate channelling may exist. Interestingly, CHS-CHR coupled reactions *in vitro* require an increased stochiometric ratio of CHR to CHS to produce an equimolar output of either chalcone (Mameda *et al*., 2018).

2.2.5 Chalcone isomerase

Chiral (2*S*)-flavanones are produced from chalcones through a Michael addition reaction catalyzed by CHI, leading to an intramolecular stereospecific cyclization (Jez *et al.*, 2002, Jez & Noel, 2002). Notably, naringenin chalcone can be cyclized spontaneously in all land plants to form naringenin) (Boland & Wong, 1975, Cheng *et al.*, 2018). However, the CHI fold has been dubbed a "perfect enzyme", because it catalyzes a reaction approaching the diffusion limit, meaning the rate-limiting step of this reaction is the length of time the substrate takes to reach the active site. (Jez & Noel, 2002) (**Figure 1**). CHI activity is therefore a significantly rapid catalyst for this reaction, over spontaneous isomerization. Chalcone-to-flavanone isomerization can proceed spontaneously, although very slowly, with a half-life of 3.5 hours at alkaline pH, and results in racemic flavanone products (2*S*/*R*) (Dixon & Whitehead, 1982). CHI-mediated catalysis of either chalcone type is enantiospecific, producing only (2*S*)-flavanones required for downstream synthesis, and can increase the turnover rate by 10^7-fold compared to spontaneous isomerization (Bednar & Hadcock, 1988).

The CHI family of proteins is classified into four subfamilies (type I to IV) based on amino acid sequence and biochemical activity (Ngaki *et al*., 2012, Dastmalchi & Dhaubhadel, 2015). Type III and II appear to be non-catalytic ancestors to types I and II with *bona fide* chalcone-to-flavanone isomerization. All four CHI subfamilies have a conserved 3D structure; a ligand-binding pocket within a helical layer, bound by two α-helical segments (helix-turn-helix α6-α7 and a helical hairpin α4-α5), β-hairpin, and the face of the core β-sheet (Jez et al., 2000b). A probable trajectory for the appearance of *bona fide* CHI activity from the type IV common

ancestor was mapped out using a directed evolution study and ancestral sequence inference (Kaltenbach *et al*., 2018). They explicitly identified three potential founder mutations in the type IV ancestral sequence that established CHI activity: Lys100Thr and Ser111Asn. Enzymes that diverge into scaffolding or regulatory proteins typically lose their catalytic activity (Adrain & Freeman, 2012). In the case of chalcone-to-flavanone isomerization, it appears that the non-catalytic binding capacity of FAPs led to polyketide binding and subsequently into *bona fide* chalcone-to-flavanone isomerization (Ngaki *et al*., 2012, Kaltenbach *et al*., 2018).

The first evolutionary step seems to have been the formation of substrate-binding pockets in FAPs – also known as type III CHIs. Like other effectors of flavanone metabolism, CHIs arose from central fatty acid biosynthesis. Type III CHIs have been studied in Arabidopsis, where both *in planta* and *in vitro* evidence suggests that type III CHIs affect fatty acid biosynthesis in plant cells and storage in the developing embryo (Ngaki *et al*., 2012). FAP knockouts in Arabidopsis showed elevated *α*-linolenic acid levels with significant reproductive seed formation defects. Evidence also suggests that some FAP isoforms can bind to polyketides during chalcone metabolism (Ban *et al*., 2018). Phylogenetic analysis of FAPs has identified three subfamilies of FAPs (FAPs1-3) that are subdivided by phylogenetic relatedness and substrate binding. FAPs1/2 typically bind to longer-chain fatty acids (i.e., 16-18 carbons), such as laurates, whereas FAP3s typically bind to short-chain fatty acids, such as palmitic acid (i.e., 12 carbons) (Ngaki *et al*., 2012). FAPs localize to the chloroplast and affect fatty acid biosynthesis in Arabidopsis, but the putative chloroplast localization sequence is absent in the later stages of CHI evolution (Dastmalchi & Dhaubhadel, 2015). FAPs are present in plants in addition to some algae, bacteria, and fungi (Gensheimer, 2004).

The next step in developing the CHI fold was the formation of a rigid chalcone binding cleft, found in type IV CHIs – CHI-like (CHIL) (Ngaki *et al*., 2012, Kaltenbach *et al*., 2018). Structural comparisons of FAPs and CHILs reveal multiple substitutions at residues that line the ligand binding pocket. However, the exact mutations that cause CHILs to gain polyketide binding affinity are yet to be confirmed. It is evident that the substrate binding pocket of CHILs is more rigid than FAPs and does not bind to fatty acids of any size.

A unique FAP ortholog from *Humulus lupulus* (hops) (HlCHIL1) can bind to prenylated chalcones (Ban *et al*., 2018). HlCHIL1 belongs to the FAP1 classification; however, researchers identified two substitutions, Phe139Val and Phe143Ala, that differ from the Arabidopsis FAP1 orthologs. Double V139F and A143F mutants in HlCHIL1 completely lost the ability to bind to polyketides, naringenin and prenylated chalcones while gaining stronger fatty acid binding during differential scanning fluorimetry (DSF) (Ban *et al*., 2018). Therefore, these two residues could play a conserved role in polyketide binding. It remains to be seen if other residues are involved in conferring full CHIL function.

Enzymatically active (*bona fide*) CHI isoforms subsequently evolved from CHILs (Kaltenbach *et al*., 2018). Type I CHIs are common in all land plants and capable of cyclizing naringenin chalcone to naringenin. Type II CHIs can accept naringenin chalcone and isoliquiritigenin as a substrate leading to a stereospecific ring closure (Dastmalchi & Dhaubhadel, 2015). Type II CHIs were initially designated as legume-specific; however, recently, type II CHI activity has been discovered in some “basal land plant” taxa, including bryophytes (*Marchantia* spp.) and lycophytes (*Selaginella* spp.) (Cheng *et al*., 2018, Ni *et al.*,

2020). Interestingly, CHI amino acid sequences from *M. polymorpha* and *S. moellendorffii* do not cluster with either type I or legume-derived type II CHIs (**Figure 3**) but can isomerize both naringenin chalcone and isoliquiritigenin, resembling legume type II CHIs (Cheng *et al*., 2018).

Within the CHI active site, the ligand can interact with the enzyme through a series of van der Waals interactions and two hydrogen bonds. The hydrogen bond network within the binding cleft exposes the ketone oxygen of the ligand to undergo cyclization (Jez & Noel, 2002, Ngaki *et al*., 2012). The structure of the CHI fold has been well characterized, as five residues have been identified to contribute to the catalytic core in type I CHIs: Arg36, Tyr48, Tyr106, Asn113, and The190 (residue numbers are based on *M. sativa* MsCHI) (Jez et al., 2000b). Arg36 aligns the substrate and stabilizes the transition by interacting with the nucleophilic 2′-hydroxyl group. The other four residues comprise the hydrogen bond network with carbonyl oxygen and 6′-hydroxyl group, thus directing the substrate binding and promoting a ring closure. Extant CHIs retain three key catalytic residues, Arg36, Thr48, and Tyr106, essential for catalytic function. Legume-specific Type II CHIs have four of the five residues conserved but possess a mutation at Thr190Ser. Structurally, types I, II, and IV share a close resemblance in terms of fold and substrate-binding pockets; however, CHIL homologs have substitutions at many of these critical catalytic residues compared to CHIs, and orient naringenin in an “opposite-mode” within its active site (Kaltenbach *et al*., 2018). Therefore, establishing CHI activity required substrate repositioning within the active site to allow catalysis.

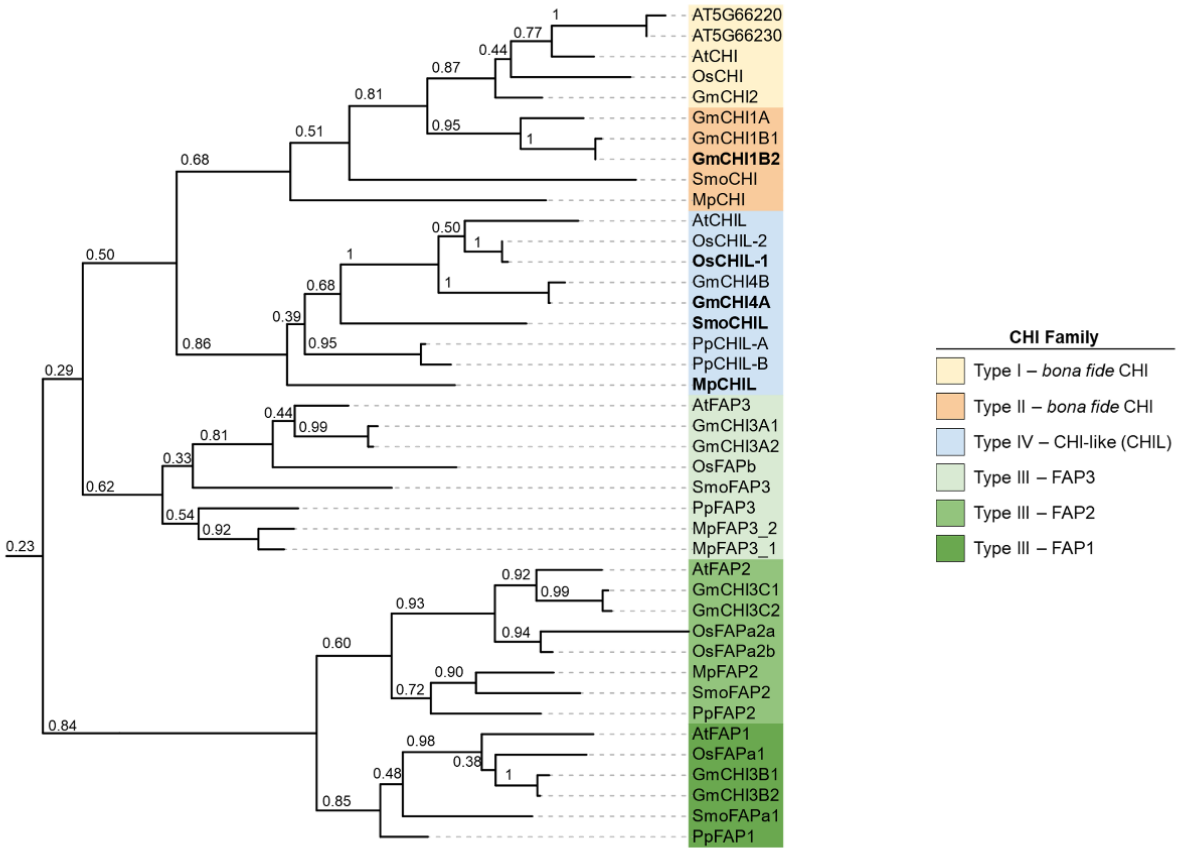


Figure 3. Phylogenetic analysis of the chalcone isomerase (CHI) family. Amino acid sequences from *M. polymorpha*, *P. patens, S. moellendorffii, O. sativa*, *A. thaliana*, and *G. max* were queried in Phytozome v13.0 using BLASTp search algorithm with Arabidopsis CHI (AT2G43570) as a reference sequence (**Appendix 1**). Colour ranges split up each CHI type (I–IV). Bolded names are indicated as proteins used during this study. Sequences were aligned in MEGA X with MUSCLE, and maximum likelihood phylogenetic analysis was completed with 500 bootstrap replicates. The resulting tree was drawn to scale with branch lengths signifying substitutions and visualized using the interactive Tree of Life (iTOL) (https://itol.embl.de/).

2.2.6 The functional characterization of CHIL

In 2014, Morita *et al*. identified an *ENHANCER OF FLAVANOID PRODUCTION* (*EFP*) in *Ipomoea nil* var. *intergriuscula* (Japanese morning glory). Loss of function *efp* mutants demonstrated a pale-coloured flower phenotype, likely due to reduced anthocyanin production. Additionally, *efp* mutants accumulated reduced total colourless flavonoids, including flavanol, relative to the wild-type. Therefore, they reasoned that the protein encoded by *EFP* acted early in flavonoid biosynthesis affecting both flavanol and anthocyanin levels. Moreover, they identified that *EFP* encodes a CHIL protein.

Interestingly, *efp* knockout mutants had impaired flavanone biosynthesis, however, it was unclear how a non-catalytically CHI family member could lead to a significant chemical phenotype shift. Morita *et al*. silenced EFP homologs in *Petunia hybrida* and *Torentia hybrida* using RNAi leading to the same pale flower phenotype. They confirmed the hypothesis that CHIL affected early flavanone biosynthesis by assessing the impact of *efp* mutants on CHS activity. They showed that *p*-coumaroyl-CoA molecules were being diverted towards caffeic acid derivatives in the *efp* mutant, rather than chalcone biosynthesis. Thus, they reasoned that CHIL is an enhancer of early flavonoid biosynthesis by directing metabolic flux towards chalcone biosynthesis through an unknown mechanism.

The initial characterization of CHI activity was completed using *transparent testa 5* (*tt5*) mutants in Arabidopsis by Shirley *et al*. in 1992. The *tt5* mutants are part of an extensive collection of *tt* mutants with impaired flavonoid metabolism, leading to a phenotype of transparent seed coats. In *tt5* mutants, the transparent seed coat arises due to a disruption of

proanthocyanidins (PA) biosynthesis (Shirley *et al.*, 1992). When CHIL was cloned into *tt5* Arabidopsis mutants, CHIL could not restore the wild-type phenotype (Jiang *et al.*, 2015). However, CHIL enhanced the production of PAs and flavonoids in *tt5* mutants. It was reasoned that CHIL had no function within the pathway due to its catalytic inability to complement flavanone biosynthesis. However, CHIL acts as an auxiliary protein that regulates flavonoid biosynthesis (Jiang *et al*., 2015).

Arabidopsis *chil* null mutants accumulated lower levels of flavonols, such as kaempferol, and PAs in seeds, but anthocyanin levels remained unchanged when measured spectrophotometrically. CHIL may impact anthocyanin production because the total metabolites produced were not confirmed via mass spectrometry (Jiang *et al*., 2015, Dastmalchi, 2021). Additionally, CHIL overexpression in Arabidopsis (*AtCHIL*) enhanced the levels of kaempferol biosynthesis. Together the work in morning glory and Arabidopsis demonstrated that CHIL is an enhancer of flavonoid metabolism, likely acting alongside CHS and CHI to promote the production of downstream metabolites.

The next step in CHIL characterization came from female cones of hops. Xanthohumol (XN) is a prenylated chalcone coveted because of its anti-cancer, anti-inflammatory, and antioxidant properties (Stevens & Page, 2004, Zanoli & Zavatti, 2008). Hops are unique in that the characteristic ring closure by CHI is omitted in favour of forming more decorated chalconoids. In hop trichomes, two non-catalytic CHIL isoforms, HlCHIL1 and HlCHIL2, antagonistically regulate this pathway. HlCHIL1 is most closely related to type III CHIs, as previously mentioned. HlCHIL2 is a type IV CHI (**Figure 3**). *In vitro* HlCHIL2 enhanced CHS

activity by increasing the production and prenylation of naringenin chalcone (Ban *et al*., 2018). *In vitro* conversion rates of the starting substrates *p*-coumaroyl-CoA and malonyl-CoA increased 5.6-fold and 17.5-fold, respectively, by introducing HlCHIL2.

When CHILs were added to a reconstituted desmethylxanthohumol (DMX) pathway in yeast, the overall output of DMX products was enhanced. The ubiquity of CHS and CHILs across plant phyla can be associated with the role of flavonoids in land colonization. CHIL isoforms from bryophytes to soybean were determined to physically interact with their native CHSs and other partners in XN synthesis using protein-protein interaction studies (Burbulis & Winkel-Shirley, 1999, Dastmalchi *et al*., 2016, Ban *et al*., 2018). Interestingly, HlCHIL1 does not interact with CHS or CHI, suggesting the required interaction is limited to higher-order CHIs (types IV, I, and II). Later, it was determined that CHIL is a component of flavonoid metabolons that binds to soluble CHS and is recruited to the ER (Waki *et al*., 2020).

Interestingly, Waki *et al*. reasoned that the ability of CHILs to bind to CHS may not be conserved in all cases and is thought to be species-specific (Waki *et al*., 2020). However, there is evidence that suggests that CHILs may be able to interact and affect non-native CHS isoforms. HlCHIL2 can interact with CHS isoforms from rice, *Selaginella,* and Arabidopsis using split-luciferase assays (Ban *et al*., 2018). Additionally, related type III PKSs (CHS-like) in *Hydrangea macrophylla* and *Vitis vinifera* do not interact with their endogenous CHIL isoforms (Ferrer *et al*., 1999, Austin & Noel, 2003). Thus, it is possible that a physical interaction between CHILs and CHS may not be required for the formers corrective function.

The CHIL mechanism of action has yet to be elucidated, but the current theories suggest that either CHIL works with CHS in concert to mediate product selectivity or CHIL binds to lactone derivatives to control metabolic flux. The first theory reasons that CHILs enhance CHS-catalyzed chalcone production by binding to CHS and an intermediate, thus, maintaining the ring-open structure of the chalcone, diminishing lactone production without changing the total polyketide production (Waki *et al*., 2020). Interestingly, CHIL does not affect the reactivity of CHS-like homologs, specifically stilbene synthase (STS; EC 2.3.1.95) or *p*-coumaroyltriacetic acid synthase (CTAS) from *H. macrophylla* var. *thunbergia.*

Recently it has been reasoned that CHIL acts as a regulator of metabolic flux through CHS through interactions with naringenin chalcone and its effects on CHS binding. The CHIL isoform form *V. vinifera* led improved thermostability when associated with naringenin chalcone and reduced protein-protein interaction with VvCHS (Wolf-Saxon *et al.*, 2023). It is possible that CHIL acts as a sensor for the CHS product and subsequently dissociates with CHS to modulate the product specificities of the reaction. It remains to be determined if CHIL isoforms can bind to lactone byproducts or isoliquiritigenin *in vitro*.

By measuring the ability of CHILs from across plant phylogeny to interact with a CHS from soybean, we can potentially identify candidates for synthetic biology and potential steps in the evolutionary development of CHILs. Thus far, synthetic biology efforts have been made to produce the isoflavone genistein *de novo* in *E. coli*. Researchers have included CHIL isoforms from multiple branches of plant phylogeny and found that a combinatorial approach using CHIL from *Malus domestica* (MsCHIL) coupled with CHS from *Sorghum bicolor* (SbCHIL) led to the

highest titre of naringenin (150 mg/l) (Liu *et al*., 2022). These results dissuade the previous notion that CHILs must act in a species-specific manner with a corresponding CHS. Additionally, this paper highlights the impact CHIL can have on a synthetic biology platform.

Although it has not yet been confirmed, CHIL activity may influence isoflavonoid production through the CHIL-CHS protein-protein interactions. Specifically, the downstream effects of CHIL on isoflavonoid metabolites produced are yet to be examined. Furthermore, a three-way partnership between CHS, CHR, and CHIL has not been established in a "legume-specific" reaction. The effect of CHIL on enhancing isoliquiritigenin, or the resulting flavanone liquiritigenin, remains unknown. Due to the complex mechanism by which CHR acts during the CHS reaction, it is important to see how CHIL contacts these enzymes.

This research focuses on the involvement of CHILs from across evolutionary branches of plant phylogeny – lycophytes, bryophytes, angiosperm monocots, and legumes – to retain CHIL activity when working with a soybean CHS *in vitro.* Additionally, this research aims to identify the role of CHILs when coupled with CHS and CHR in a legume-specific reaction to form isoliquiritigenin and liquiritigenin.

Auxiliary proteins have previously been overlooked for their non-catalytic activity but are integral for specialized metabolism; now CHILs are part of this growing list. Other examples include pathogenesis-related proteins 10 (PR10s), such as Fra a (flavonoid pathway) and NISO (morphine pathway) (Muñoz *et al.*, 2010, Dastmalchi *et al*., 2019, Dastmalchi, 2021). Both CHILs and PR10s are members of larger gene families with ligand-binding roles in plants. Like CHILs, PR10s are involved in guiding metabolic flux towards specific products. Therefore, there

is a recent precedent for the role of auxiliary proteins to improve efficiency in engineered pathways.

Understanding how CHILs impact isoflavonoid metabolism and the evolution of the type IV gene family will enable researchers to better understand the components involved in synthetic biology. Lastly, this research aimed to understand a missing link in the evolutionary jump from FAPs to CHILs. Using site-directed mutagenesis experiments and rational design, we aimed to establish CHIL activity, substrate binding, and interactions with CHS in soybean FAP3 (GmCHI3A2), the most recent common ancestor to higher-order CHIs in soybean. Overall, this research offers valuable insight into the role of CHIL during the initial stages of isoflavonoid production and will determine the key residues required for CHIL activity to improve ongoing synthetic biology efforts.

CHAPTER III: MATERIALS & METHODS

3.1 Plasmids and *E. coli* strains

The plasmids used in this study were obtained from the following sources. Plasmids used for recombinant protein purification in *E. coli: pUC-GW-Amp-GmCHIL* (Azenta Life Sciences; Burlington, MA, USA), *pET28α* (+)-*MpCHIL*, *pET28α* (+)-*SmoCHIL*, and *pET28α* (+)-*OsCHIL* were codon-optimized for *E. coli* expression (Soon Goo Lee, University of North Carolina Wilmington; UNCW). Additionally, *pET28α* (+) (Novagen; Darmstadt, Germany), *pET160-DEST* (Invitrogen; Waltham MA, USA), *pDONR™/Zeo* (Invitrogen), *pET160-DEST-GmCHR14A* (Invitrogen), *pET32α* (+)-*GmCHS8* (EMD Biosciences; San Diego, CA, USA), and *pDONR™/221-GmCHI1B2* (Invitrogen) were also used. The *E. coli* strain NEB10*β* was used for plasmid DNA production (New England Biolabs (NEB); Ipswich, MA, USA) and BL21 (DE3) was used for recombinant protein expression.

3.2 Chemicals

Analytical-grade standards were purchased for *in vitro* assays, including *p*-coumaroyl-CoA (MicroCombiChem; Wiesbaden, Germany), malonyl-CoA (Sigma Aldrich; St. Louis, MO, USA), naringenin chalcone (Phytolab; Lenexa, KS, USA), naringenin (Sigma Aldrich), isoliquiritigenin (Sigma Aldrich), liquiritigenin (Sigma Aldrich), NADPH (BioShop; Burlington, ON, CA), and fluorescein (Sigma Aldrich). Metabolites were stored at -20 °C in powder as lyophilized form until use.

3.2 Phylogenetic analysis of the CHI protein family

Deduced amino acid sequences for the CHI protein family were obtained from *M. polymorpha* (bryophyte), *P. patens* (bryophyte), *S. moellendorffii* (lycophyte), rice (monocot), Arabidopsis (angiosperm dicot), and soybean (Fabaceae) using the Phytozome v13.0 (http://www.phytozome.net) protein BLAST algorithm with Arabidopsis CHI (AT2G43570) as a reference sequence and an additional keyword search for "chalcone isomerase" (**Appendix 1**). Sequences were aligned by multiple sequence comparison by log-expectation (MUSCLE) alignment algorithm using the MEGA X platform (Kumar Sudhir, 2018), and alignments were visualized using Boxshade alignment tools (https://bioinformatics.org/SMS/multi_align.html). The phylogenetic relatedness of CHI isoforms was inferred using the Jones-Taylor-Thornton (JTT) matrix-based model (Jones *et al*., 1992) in MEGA X with 500 bootstrap replicates and visualized using the interactive tree of life (iTOL) (https://itol.embl.de/). CHI amino acid sequences were grouped into the appropriate subfamilies based on previous literature (Ngaki *et al.*, 2012, Dastmalchi & Dhaubhadel, 2015). Percent amino acid identity was calculated using pairwise sequence analysis via Clustal ω (https://www.ebi.ac.uk/Tools/msa/clustalo/).

3.3 Plant material

Soybean (*Glycine max* L. Merr. var DH410) were grown for six weeks with controlled conditions in growth chambers (16 hrs light, 8 hrs dark; 22 °C day, 20 °C night; 50-40 % humidity during day-night). They were fertilized once every 2 weeks (4 g/l Green Flag® all-purpose water-soluble fertilizer). Leaf tissue was harvested and flash-frozen in liquid nitrogen before being stored at -80 °C.

3.4 RNA extraction and cDNA synthesis

Total RNA was isolated from approximately 70-100 mg of soybean leaf tissue was homogenized using a Qiagen Tissue Lyser II® (Hilden, Germany) with five 2.8 mm ceramic beads (OMNI international; Kennesaw, GA, USA) per tube. Tissue homogenization was conducted in RLC buffer, followed by subsequent steps outlined in the Rneasy® Plant Mini Kit protocol (Qiagen), yielding 30 µl of the RNA sample. Total RNA was quantified using a NanoPhotometer® NP80 (Implen; Westlake Village, CA, USA). Concentration and integrity were confirmed through gel electrophoresis of 0.2 µg of RNA on a 1 % (w/v) agarose gel with 1 % (v/v) bleach in 1x TAE buffer and stained with GelRed® (Sigma Aldrich) nucleic acid staining solution and Trackit™ CyanOrange loading buffer (Invitrogen). cDNA was synthesized from 1 µg of RNA using the "All-In-One 5X RT MasterMix" (ABM; Richmond, BC, Canada) with a final concentration of 50 ng/µl. Genomic DNA contamination was removed using the master mix.

3.5 Construct design and assembly

Full-length *GmCHIL* (*GmCHI4A*; Glyma.06G143000) was codon-optimized for *E. coli* recombinant protein expression. The restriction cut sites NdeI and XhoI (NEB) were added at the N and C-terminus of the coding sequence, respectively. Plasmid DNA (1 µg of *pET28α* (+) vector and 750 ng of *pUC-GW-AMP-GmCHIL*) was linearized at 37 °C for 1 hr by two units (1 µl) of NdeI and XhoI. The success of the digestion reaction was confirmed using gel electrophoresis by running the total reaction volume with 3 µl of Gel Loading Dye, Purple (6x) (NEB) on a 1 % (w/v) agarose gel. The linearized *pET28α* (+) vector (5369 base pairs (bp)) and the *GmCHIL* insert (639 bp) were extracted using the Monarch® Gel extraction kit (NEB). A 3:1

ratio of insert (*GmCHIL*) to vector (*pET28α* (+)) was ligated for 1 hr at 25 °C with 1 µl of T4 ligase (NEB). The ligation product was transformed (5 µl) into a 50 µl aliquot of chemically-competent NEB10*β E. coli* cells and grown on 50 µg/ml kanamycin lysogenic broth (LB)-agar plates overnight at 37 °C. Single colonies were selected for colony PCR (2x QuickLoad® Taq polymerase (NEB)) using T7 sequencing primers, with empty *pET28α* (+) as a negative control. A complete list of oligonucleotides used during this study is reported in **Appendix 3**. Verified colonies were selected, and plasmids (and all plasmids hereafter) were purified using the EZ-10 Spin Column Plasmid DNA Miniprep Kit protocol by Bio Basics Inc. (Markham, ON, Canada).

Gateway® cloning technology (Invitrogen) was employed to clone *GmCHI1B2* (Glyma.10G292200) from *pDONR221™-GmCHI1B2*, using a two-step process of DNA recombination. Gateway® attachment sites (*att*-b) were added by PCR amplification (**Appendix 3**) using Q5® high-fidelity DNA polymerase (NEB). The PCR product was mixed with Gel Loading Dye, Purple (6x) (NEB) and purified using a Monarch® gel extraction kit (NEB). The *pDONR™/Zeo -GmCHI1B2* entry vector was assembled at 25 °C for 1 hr, using 100 fmoles of the purified PCR fragments (insert) with 100 fmoles of *pDONR™/Zeo* (vector; Invitrogen) using 1 µL of BP clonase™ II enzyme (Invitrogen), followed by termination with 1 µl of proteinase K at 37 °C for 10 min. The BP clonase™ product was transformed with 2 µl into NEB10*β E. coli* (50 µl). Single colonies grown on LB-agar plates with 100 µg/ml of Zeocin® (Invitrogen) were screened using colony PCR with M13 (vector-specific) and gene-specific primers. The destination construct was generated by an LR reaction (cloning into a desired destination vector tailored for specific functions, (e.g., His-tag fusion recombinant proteins). LR clonase™ II (Invitrogen) recombined 39.45 fmoles of *pDONR™/Zeo-GmCHI1B2* (entry vector) and 100

fmoles of *pET160-DEST* (destination vector) using the same assembly protocol as the BP reaction. *pET160-DEST* is a protein expression vector with a 6-residue histidine (His)-tag region at the N terminus of the multiple cloning site, driven by a T7 promoter. Following enzyme inactivation, 5 µl of the LR reaction was transformed into NEB10*β* and LR transformants were screened on appropriate antibiotic selection (100 µg/ml ampicillin) and verified using PCR with T7 (vector-specific) and gene-specific primers. Sanger sequencing at the Centre d'Expertise et de Services Génome Québec (Montréal, QC, Canada) validated that the gene of interest was in frame with the His-tag in the purified plasmid. Finally, constructs were transformed into BL21 (DE3) for protein purification.

3.6 Recombinant protein purification

Recombinant proteins were translationally fused at the N-terminus to a His-tag for protein purification (**Appendix 4**), and 1 µg of each construct was transformed into *E. coli* BL21 (DE3) chemically competent cells. Single colonies were screened and inoculated into 50 ml LB starter cultures containing 50 µg/ml chloramphenicol with the appropriate vector-specific antibiotic (100 µg/ml ampicillin or 50 µg/ml kanamycin) shaking at 37 °C at 225 rotations per minute (rpm) overnight (16 hrs). Starter cultures were diluted 20-fold into 1 L of LB and subcultured until optical density (OD) reached absorbance 0.5 at $UV_{600\ nm}$. Protein production was induced with 1 mM isopropyl-*β*-D-1-thiogalactopryranoside (IPTG) (VWR; Radnor, PA, USA) and incubated for 4 hrs or at 25 °C. Cell paste was harvested at 5000 rpm for 30 minutes at 4 °C using the J2-21 refrigerated floor model centrifuge installed with the Ja-14 rotor (250 ml) (Beckman; Indianapolis, ID, USA).

Individual cell pastes were resuspended in 10 ml/g of pellet in Tris-HCl lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 % glycerol (v/v)) and lysed for 5 minutes at 80 % amplitude (60 Htz) pulsing for 5s on and 5s off via a microprobe ultrasonic liquid processor (Misonix; Farmingdale, NY, USA). Lysed cells were centrifuged at 14, 000 x *g* for 30 minutes at 4 °C to remove cellular debris (J2-21 refrigerated flood model centrifuge; Ja-21 rotor). The supernatant (soluble fraction) was collected and incubated with 1 ml of buffer-equilibrated TALON© resin (Clonetech; TAKARA Bio; Kusatsu, Shiga, Japan) on ice for 1 hr shaking at 65 rpm. The resin-supernatant solution was added to a gravity flow column (Thermofischer Scientific). Bound protein was washed five times with 1 ml of 5 mM imidazole supplemented lysis buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol (v/v), 5 mM imidazole). Proteins were then eluted by at least six washes with 1 ml of 150 mM imidazole supplemented lysis buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol (v/v), 150 mM imidazole) and each eluted fraction collected separately. A Vivaspin® protein filter with a 10 kDa threshold (VWR) was used to concentrate each protein, and a complete buffer exchange to remove imidazole.

Purified recombinant proteins were quantified using two methods: first, using a Bradford assay relative to a Quick Start™ Bovine serum albumin (BSA) standard curve (0.0625 to 2.0 mg/ml) (Bio-Rad). Concentrated protein elution (4 µl) was added to a cuvette (VWR) containing 1 ml of 1x Bradford reagent (R-G250) (Bio-Rad; Hercules, CA, USA), mixed, and measured using an Eppendorf BioSpectometer® basic (Hamburg, Germany) at absorbance $UV_{590\ nm}$. In addition, 1 µl of purified protein was quantified using a Nanophotometer NP80® (Implen) at absorbance $UV_{280\ nm}$ relative to the same BSA standard curve. Pure proteins were then diluted to

1 µg/µl aliquots and flash-frozen in liquid nitrogen before storage at -80 °C. CHIL proteins were purified within one week of each other to ensure structural integrity is preserved for assays.

3.7 SDS-PAGE and Western blots

Fractions from each step of the protein purification protocol were collected and analyzed using SDS-PAGE within a Mini-PROTEAN® tetra cell gel electrophoresis rig (Bio-Rad). The purity of each fraction was measured by running 700 ng of concentrated protein through a 15 % (v/v) polyacrylamide gel with a 29:1 ratio of acrylamide: bis-acrylamide at 100 V (approximately 30 min) through the stacking gel and then through the resolving gel at 200 V (approximately 1 hr). Following electrophoresis, gels were either stained for total protein using Coomassie Brilliant Blue R250 (Bio-Rad) for 1 hr and de-stained using an 80 % methanol and 10 % acetic acid (v/v) solution until the bands were clear (approximately 2 hrs) or electroblotted via a semi-dry transfer cell (Bio-Rad) onto poly(vinylidene) difluoride (PVDF) (Bio-Rad) membranes for Western immunoblotting. Electroblotting transfer efficiency and the relative molecular weight of each protein band (kDa) were qualitatively measured using a pre-stained protein molecular weight marker (SMOBIO; Roswell, GA, USA). Proteins transferred onto PVDF membranes were blocked for 1 hr with 5 % (w/v) skim milk powder (BioShop). Proteins were probed with a 1:1000 dilution of mouse anti-His primary antibody (Abclonal; Woburn MA, USA) resuspended in Tris-buffered saline (20 mM Tris pH 7.5, 150 mM NaCl) with 0.1 % (v/v) Tween-20 (TBST) containing 1 % (w/v) BSA. The membrane was subsequentially washed, and incubated with a 1:5000 dilution of goat anti-mouse secondary antibody, conjugated with horseradish peroxidase

(HRP) (Abclonal). Immunoreactive peptides were visualized using Clarity© Western enhanced chemiluminescence (ECL) substrate (Bio-Rad) within a ChemiDoc™ imaging system (Bio-Rad).

3.8 *In vitro* CHIL coupled assays

The standard *in vitro* coupled assay consists of a 50 µl reaction containing 50 mM HEPES-NaOH buffer pH 7.5, 100 µM *p*-coumaroyl-CoA, 200 µM malonyl-CoA, 500 µM of fluorescein as an internal standard, and 3 µg of CHS in all cases. Each CHIL ortholog (4 µg) was independently incorporated into the assays. Denatured (boiled) CHIL proteins through boiling were used as negative controls for each ortholog. To determine the effect of CHILs on isoliquiritigenin production, 4 µg of CHR and 10 mM NADPH were included in the reaction mix. The impact of CHILs on flavanone production was confirmed by introducing 1 µg of CHI. Reaction constituents were pre-heated at 30 °C for 10 minutes before the reaction was initiated with the addition of CHS and incubated at 30 °C for 40 minutes. The reaction was quenched with 20 µl of 1 N hydrochloric acid (HCl), followed by the addition of 180 µl of ethyl acetate and centrifuged at 16, 000 x *g* for 30 minutes. The upper phase (100 µl) was dried down using the Savant SpeedVac™ DNA130 vacuum concentrator system (Thermofischer Scientific) and stored at -20 °C. Prior to liquid chromatography-mass spectrometry (LC-MS) analysis, the samples were resuspended in 80 µl of absolute methanol and transferred into glass vials (Agilent Technologies; Santa Clara, CA, USA).

3.9 LC-MS analysis

Products of *in vitro* coupled assays were measured using an LC-MS 6120 (single quad rupole) (Agilent Technologies). The following system was used to separate compounds: column (Poroshell 120 EC-C18 3.0 x 50 mm, 2.7 μM) (Agilent Technologies); guard (Poroshell 120, UHPLC Guard.EC-C18, 3.0 mm) (Agilent Technologies); flow rate of 0.5 ml/min; solvent A, 0.1 % formic acid (v/v) in H_2O; solvent B, 0.1 % formic acid (v/v) in acetonitrile. The column was equilibrated with 100 % solvent A from 0 – 0.5 min. Subsequently, 5 μl of the sample was injected and a gradient was developed by washing the column with 50 % solvent B (v/v) from 6-17 min; 90 % (v/v) solvent B from 17-17.5 min; followed by 100 % solvent B from 17.5-19 min. The column was rewashed and equilibrated with 100 % solvent A. Standards for all substrates and products were tested using this method, and their retention times were noted. Metabolite detection was completed as absorbance $UV_{254\ nm}$ for fluorescein and liquiritigenin, $UV_{280\ nm}$ for CTAL, naringenin, and isoliquiritigenin, and $UV_{360\ nm}$ for naringenin chalcone.

Quantification of metabolites was completed using LC-MS and the OpenLab CDS software (Agilent Technologies) by solving the area under the curve (integration) of a given peak. A standard curve for each metabolite was generated to quantify the samples (0.5 – 160 μM). Using the mass/charge ratio (*m*/*z*) and the daughter ion chromatogram, the byproduct, CTAL, was identified. Responses were normalized to the fluorescein internal standard, followed by statistical analysis.

3.10 Statistical analysis

All experiments were conducted with a minimum of three replicates. Data is represented as sample mean ± standard deviation. Two sets of independent variables were compared using a two-tailed Student's *t*-test with an assumption of equal variance, and $p < 0.05$ was considered significant. When comparing the effect on each product between treatment groups, a two-way Analysis of Variance (ANOVA) was completed using JMP® statistical discovery from SAS (SAS Institute Inc, 2021). The least squared means were determined by a Tukey honestly significant difference (HSD) test in JMP®. In these cases, the two independent variables are condition and compound, while the concentration is continuous. Letters denote statistically similar groups between and within treatments. Levels not connected by the same letter are significantly different.

3.11 *In silico* FAP identification

Two Soybean FAP3 isoforms (GmCHI3A; Glyma.13G262500, GmCHI3A2; Glyma.15G242900) were identified following the query of CHI amino acid sequences (**Appendix 1** and **Figure 3**). Pairwise percent identity analysis between the amino acid sequences of GmFAP3s and MpCHIL, SmoCHIL, OsCHIL, AtCHIL, and GmCHIL was completed using Clustal ω (https://www.ebi.ac.uk/Tools/msa/clustalo/). Additionally, CHIL and FAP amino acid sequences were aligned in MEGA X using the MUSCLE alignment algorithm and visualized using Boxshade, where residue positions are black if 60 % of sequences are identical and grey if 60 % of sequences are similar. TargetP (https://services.healthtech.dtu.dk/services/TargetP-2.0/) predicts the presence and the cleavage site location of N-terminal pre-

sequences, including chloroplast transit peptides. This tool was used to asses the amino acid sequences of GmFAP3s, and the truncated amino acid, with 60 amino acids removed from the N-termius based on the predicted cleavage site, was selected for further analysis.

3.12 FAP protein modelling

The resolved crystal structures of AtFAP3 (PDB ID, 4DOL), and AtCHIL (PDB ID, 4DOK) were obtained from the Protein Data Bank (https://www.rcsb.org/structure/) as .PBD files. These crystal structures served as a template for receptor molecules during homology modelling. The predicted tertiary structure of GmFAP3s and GmCHIL was modelled using the SWISS Model (Expasy; https://swissmodel.expasy.org/). The interaction between palmitic acid and AtFAP3 was previously resolved and retained in the .PBD file. The predicted structures were exported and visualized using PyMOL (Schrödinger), where polar contacts between ligands and active-site residues can be queried.

CHAPTER IV: RESULTS

4.1 Identifying CHIL orthologs

The soybean CHI, GmCHI1B2 (Glyma.10G292200), a *bona fide* catalyst of flavanone formation, was used to query all related CHI family sequences (types I-IV) from *M. polymorpha, P. patens, S. moellendorffii, O. sativa, A. thaliana*, and *G. max* databases in Phytozome v13.0 (https://phytozome-next.jgi.doe.gov/). An additional keyword search was performed for "chalcone isomerase" within the same datasets. A total of 59 sequences were identified and aligned in MEGA X using the MUSCLE algorithm (Kumar Sudhir, 2018). Phylogenetic relationships between the compiled amino acid sequences were inferred through a maximum likelihood analysis using a JTT matrix-based model with 500 bootstrap replicates (Jones *et al.*, 1992) (**Figure 3**, **Appendix 4**). The resulting phylogenetic tree was annotated with predicted functions (types I-IV) based on grouping with previously characterized soybean and Arabidopsis CHI proteins (Ngaki *et al*., 2012, Dastmalchi & Dhaubhadel, 2015, Ni *et al*., 2020).

Two proteins within the CHI family have previously been designated as CHILs (Dastmalchi & Dhaubhadel, 2015). GmCHI4A (Glyma.06G143000) and GmCHI4B (Glyma.04G222400); in addition to these, the phylogenetic tree (**Figure 3**) includes *M. polymorpha* MpCHIL (Mapoly0175s0004), one in *S. moellendorffii*, SmoCHIL (227414), and two members in rice, OsCHIL-1 (LOC_Os11g02440), OsCHIL-2 (LOC_Os12g02370), listed here with their Phytozome identifier tags. These curated CHIL isoforms share between 37.2 % to 60.39 % amino acid identity (**Appendix 2**). From the available orthologs in rice and soybean,

OsCHIL-1 (hereafter, OsCHIL) and GmCHI4A (hereafter, GmCHIL) were selected for further analysis.

4.2 Purification of recombinant proteins from *E. coli*

The purity of protein preparations, free from non-desired, contaminating proteins, is essential for downstream characterization processes, including structure and function analyses. The unique ability of CHIL to rectify CHS promiscuity and lactone formation has only been observed *ex planta* (Ban *et al.*, 2018, Waki *et al.*, 2020), therefore, necessitating the comparative analysis of CHIL orthologs in coupled assays *in vitro*. All proteins were expressed in and purified from *E. coli*. Coding sequences were cloned or received in the pET (Thermofischer Scientific) vector series with N- or C-terminal 6-His-tags, and induced with the lactose analog, IPTG. The constructs were introduced into the *E. coli* strain BL21 (DE3).

The constructs were either received or cloned, depending on availability. Briefly, *pET32α* (+)-*GmCHS8* (Glyma.11G011500) and *pET160-DEST-GmCHR14a* (Glyma.14G005700) were obtained from Dr. Sangeeta Dhaubhadel at Agriculture and Agri-Food Canada (London, ON, Canada) (**Appendix 4**). *pET32α* (+) adds an N-terminal 6-His-tag; the CHS construct is also in-frame with an S-tag at the N-terminus, leading to a total addition of 12 kDa. The *pET160-DEST* vector is a Gateway®-compatible expression vector with an N-terminal 6-His-tag. *GmCHI1B2* (Glyma.10G292200) was also cloned into *pET160-DEST*. All four CHILs were codon-optimized for *E. coli* expression and cloned into *pET28α* (+) via restriction digest (NdeI and XhoI), yielding an N-terminal 6-His-tag. Sanger sequencing before transformation into BL21 (DE3) *E. coli* for recombinant protein expression and purification.

Protein purification using affinity chromatography proved to be challenging in this project. Many troubleshooting steps were undertaken to optimize recombinant protein expression, solubility, purity, and yield (**Appendix 5**). Initially, protein induction was optimized with a range IPTG concentrations (0.1 – 1 mM) in 50 ml cultures of LB. Secondly, the duration of incubation following IPTG addition was optimized for *pET32α* (+)-*GmCHS8*. Finally, a 1 mM IPTG concentration was selected for a 4-hour induction at room temperature. These settings led to recombinant protein expression and its appearance as a new band, compared to the uninduced control, visualized using Coomassie-stained SDS-PAGE gels.

Following the successful induction of protein expression, the steps leading to protein purification were also optimized. The N-terminal 6-His-tag, translationally fused to each protein, allowed purification using a cobalt resin. The six negatively charged His residues ionically bind to positively-charged cobalt molecules, when incubated for an hour at 4 °C, with agitation. Cobalt resin has the capacity to bind 10 mg of soluble protein with higher stringency than classical nickel resin (Jiang *et al.*, 2004). Initial purification attempts were thwarted by the loss of recombinant proteins in preliminary wash steps. Recombinant proteins were not binding to the resin. A potential lack of complete cell lysis was ruled out by supplementing the buffer with additives (including the detergents Triton X-100 (Bio-Rad) and Tween-80 (Bio-Rad), the reducing agent *β*-mercaptoethanol (Bio-Rad), the protease inhibitor phenylmethylsulphonyl fluoride (PMSF) (BioBasic), and osmolytes such as glucose and trehalose (Sigma Aldrich) that would reduce aggregation and promote the stability and solubility of proteins (**Appendix 5a**) There was no change in soluble protein yield under any of the conditions. Interestingly, recombinant proteins did become soluble when the initial cell paste was resuspended in an

excess volume of lysis buffer (10 ml per g of cell paste) (**Appendix 5b**). All seven genes of interest were purified using the same purification protocol over a gravity flow column and concentrated with a centrifugal filter. Typical protein yields were between 3 and 7 mg, and proteins were aliquoted in 1 µg/µl aliquots and stored at -80 °C.

The purity and size of the recombinant proteins were confirmed by Coomassie-stained SDS-PAGE and Western blot. **Figure 4** shows a unique band for GmCHS8 (60 kDa), GmCHR14A (48.8 kDa), GmCHI1B2 (28.8 kDa), MpCHIL (25.2 kDa), SmoCHIL (24.7 kDa), GmCHIL (25.7 kDa), and OsCHIL (25.7 kDa) (**Appendix 4**). These protein preparations were subsequently used for all *in vitro* assays.

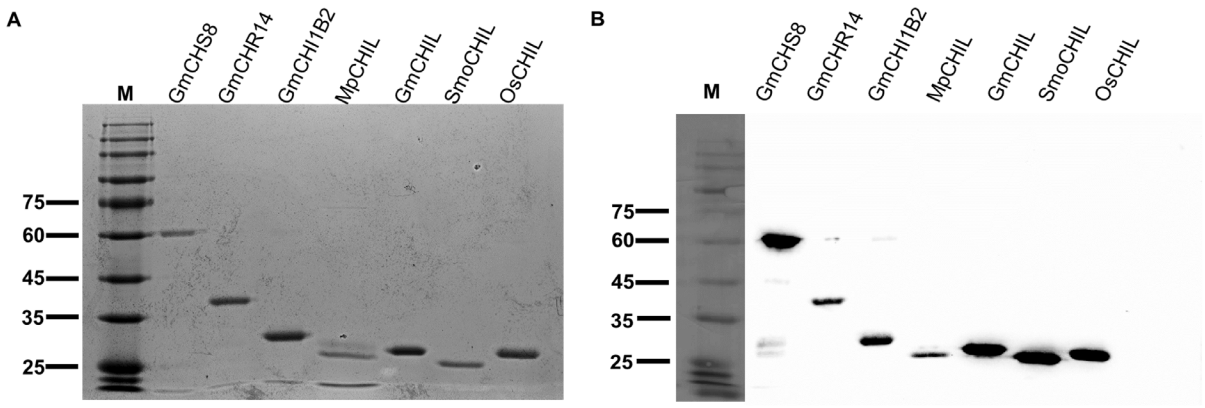


Figure 4. Purified recombinant proteins for *in vitro* assays. A) Confirmation via Coomassie Brilliant Blue R250 and **B**) Western blot analysis. Purified proteins (700 ng) were observed with Coomassie staining. Recombinant His-tagged proteins (500 ng) were confirmed by Western blot, with mouse anti-His primary antibody and goat anti-mouse secondary (Abclonal). Expected sizes of proteins are GmCHS8, 60 kDa; GmCHR14A, 38.8 kDa; GmCHI1B2, 28.8 kDa; MpCHIL, 25.2 kDa; SmoCHI, 24.7 kDa; OsCHIL, 25.7 kDa; GmCHIL, 25.7 kDa.

4.3 CHIL isoforms reduce the accumulation of CTAL in an isoform-specific manner

CHIL orthologs were assayed in conjunction with soybean GmCHS8 (hereafter, CHS) in a standard assay measuring the reaction products naringenin chalcone, naringenin, and CTAL. CHILs were added in a molar surplus to CHS (4:1) to promote conclusive determination of their function. The standard assay *in vitro* consisted of 100 µM of *p*-coumaroyl-CoA and 200 µM malonyl-CoA as substrates and 3 µg of CHS in a HEPES-NaOH buffer, pH 7.5.

In the uncoupled reaction, CHS activity yielded the expected products, naringenin chalcone, naringenin, and CTAL. All three have identical computed mass of 272.25 g/mol. Naringenin chalcone and naringenin were identified by comparison to authentic standards, under $UV_{360\ nm}$ and $UV_{280\ nm}$, separating at 7.2 and 7.3 min, respectively (**Figure 5a**). Conversely, the appearance of CTAL was putatively identified according to its mass, under $UV_{280\ nm}$ absorbance, at a retention time of 6.1 min and *m*/*z* ratio of 271.1 in negative ion mode (**Figure 5b**). We did not record the presence of BNY (230.22 g/mol) in any of our reaction assays, which is supported by previous reports with similar constituents. Due to their instability, no authentic standard could be acquired for either of the lactone byproducts. The approximate quantification of CTAL was completed using the standard curve generated with naringenin chalcone.

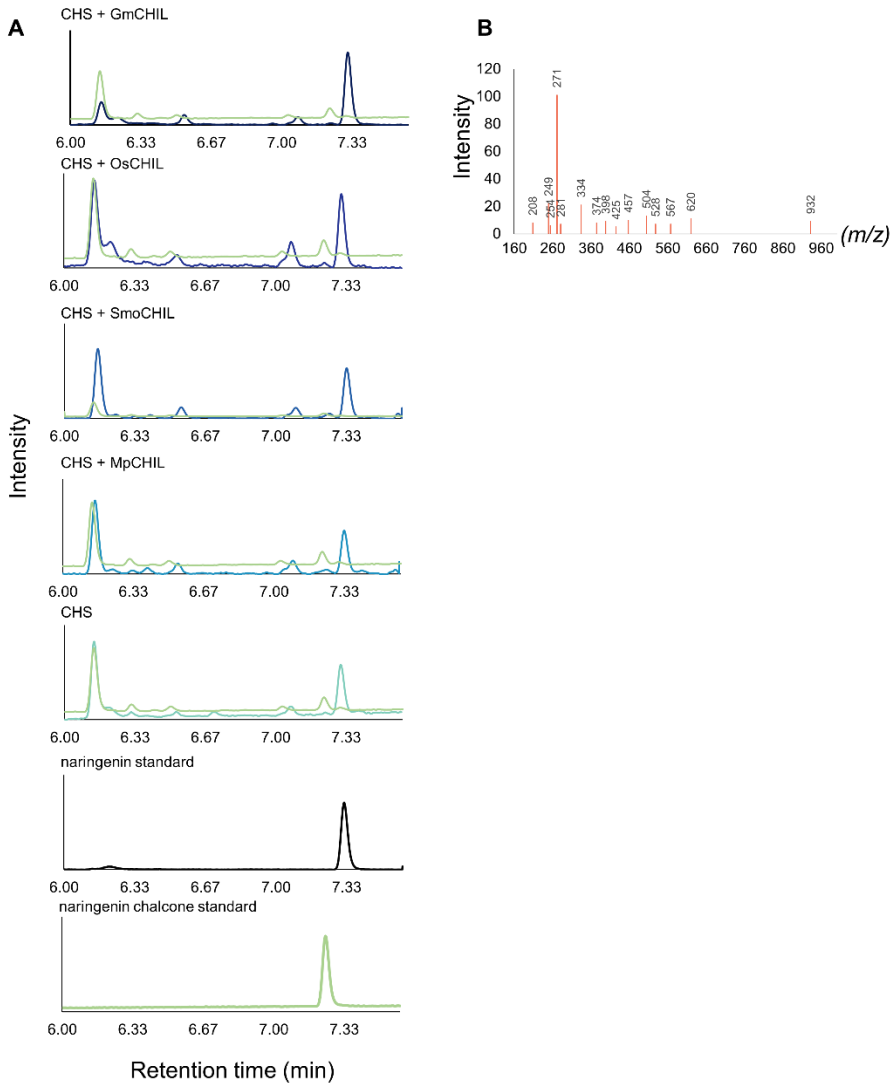


Figure 5. Representative chromatograms demonstrating the effect of CHIL isoforms on CHS product specificities. The x-axis indicates each chromatogram peak's retention time (minutes). **A**) Chromatograms of each assay were compared to a 25 µM naringenin (MicroCombiChem) at absorbance $UV_{280\ nm}$ and naringenin chalcone standard (Phytolab) at absorbance $UV_{360\ nm}$. Two identifiable peaks were seen in all assays, naringenin (retention time of 7.3 min) and a second at retention time 6.1 min. **B**) using *m*/*z* data analysis, the significant peak at the retention time of 6.1 was identified to have a *m*/*z* ratio of 271.1 identifying the peak as *p*-coumaroyltriacetic acid lactone (CTAL).

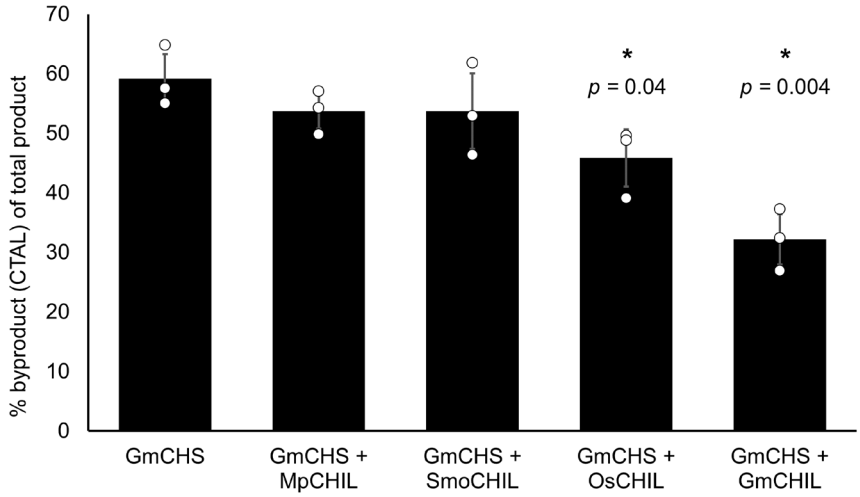


Figure 6. Reduction of byproduct (%) formation by CHIL-CHS coupling. The y-axis indicates percentage of byproduct (CTAL) to the total reaction products (naringenin chalcone, naringenin, and CTAL). The data represents the mean ± standard deviation of three independent assays (data points shown). The change in byproduct percentage was measured between CHS alone against reactions supplemented with each CHIL ortholog. The asterisk and *p* values indicate statistical significance as determined using a Student's two-tailed *t*-test with an equal assumption of variance ($p < 0.05$).

Byproduct levels were calculated as a percentage of total products to assess the diversion of carbon flux. CHS alone results in the formation of 60.55 % (**Figure 6**). Coupling CHS with the bryophyte MpCHIL and lycophyte SmoCHIL did not significantly reduce byproduct formation. Conversely, coupling CHS with either OsCHIL or GmCHIL significantly reduced byproduct formation to 45.83 % and 32.1 %, respectively.

4.4 GmCHIL and OsCHIL appear to increase product formation

Including GmCHIL and OsCHIL appears to increase overall product formation in the coupled assays (**Figure 7**). We are not able to detect the substrates (*p*-coumaroyl-CoA and malonyl-CoA) in our method (**3.9**); therefore, quantification of substrate conversion, which is preferable, was not possible. Specifically, the inclusion of GmCHIL increased the production of naringenin chalcone and naringenin by 1.9 and 2.4-fold, respectively; this increase cannot be explained by the reduction of CTAL levels (1.7-fold decrease). Interestingly, the addition of OsCHIL leads to a different product profile. In these samples, the overall production of naringenin is significantly greater than in samples without CHIL, while the production of CTAL remains unchanged (**Figure 7**). Conversely, the overall metabolic flux in both SmoCHIL and MpCHIL assays is reduced.

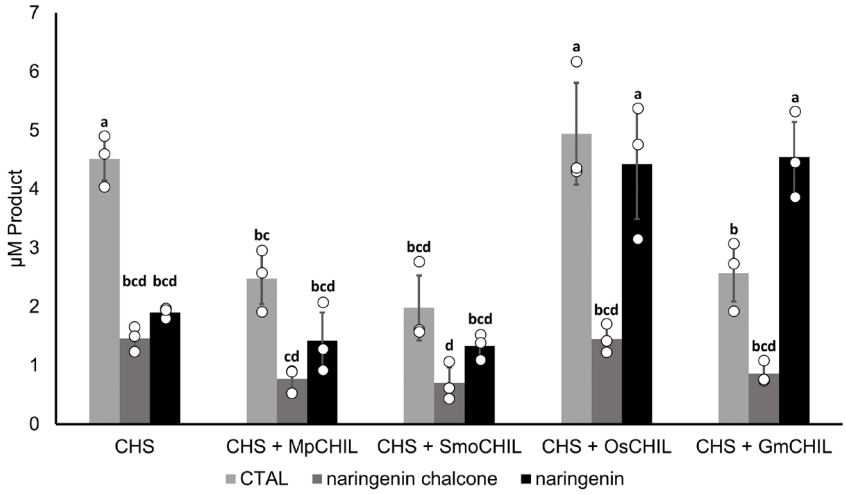


Figure 7. Product distribution of CHS-CHIL coupled assays. The production (µM) of CTAL (light grey), naringenin chalcone (dark grey), and naringenin (black) was quantified in couple assays using LC-MS. Statistically significant differences between products within and between treatment types were determined using a two-way Analysis of Variance (ANOVA) with an HDS Tukey test in JMP® (SAS Institute Inc, 2021). Bars that share a letter are not statistically different.

4.5 CHIL activity is conserved in isoliquiritigenin formation

While the activity of CHIL in rectifying CHS lactone diversion has been established in the context of naringenin chalcone formation, its utility in the legume-characteristic branch of isoliquiritigenin has not been confirmed. The latter is missing a hydroxyl group at the 6′ position and is produced by CHS coupled with CHR. Therefore, the standard assay described in section **4.3** was supplemented with 4 µg of GmCHR14A (hereafter referred to as CHR) and 10 mM of NADPH (**3.8**). CHS and CHR coupled reactions produced three distinct peaks: CTAL, naringenin, and isoliquiritigenin (computed mass of 256.1 g/mol; retention time of 7.7 min and *m*/*z* of 255.1) (**Figure 8**). Interestingly, in all reactions with CHR, naringenin chalcone, the intermediate to naringenin, was absent. Conversely, and expectedly, liquiritigenin (4′,7-dihydroxyflavanone), the flavanone product of isoliquiritigenin isomerization, was absent as this reaction does not occur spontaneously.

The coupled reaction of CHS and CHR produces 59.94 % byproduct, showing that the reductase does not rectify the catalytic promiscuity of the former (**Figure 9**). The inclusion of CHILs, irrespective of origin and isoform, significantly reduced byproduct formation between 13.3 % and 9.08 %. Notably, levels of isoliquiritigenin and naringenin are similar in all reaction assays, indicating near molar equity in the carbon flux through the diverging pathways (**Figure 10**).

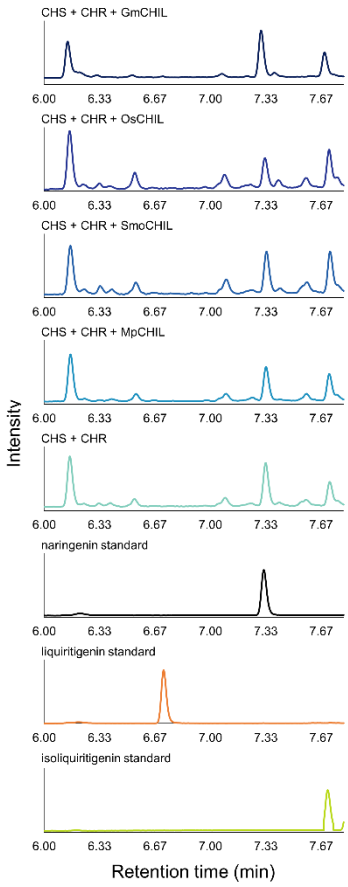


Figure 8. Representative chromatograms of the CHS, CHR, and CHIL coupled assays. The x-axis indicates the retention time of each chromatogram peak. Chromatograms of each assay were compared to 25 µM isoliquiritigenin (Sigma Aldrich), naringenin (MicroCombiChem) ($UV_{280\ nm}$), and liquiritigenin (Sigma Aldrich) ($UV_{254\ nm}$) standards. Three identifiable peaks were seen in all assays, isoliquiritigenin (retention time of 7.7 min), naringenin (retention time of 7.3 min) and CTAL (retention time of 6.1 min).

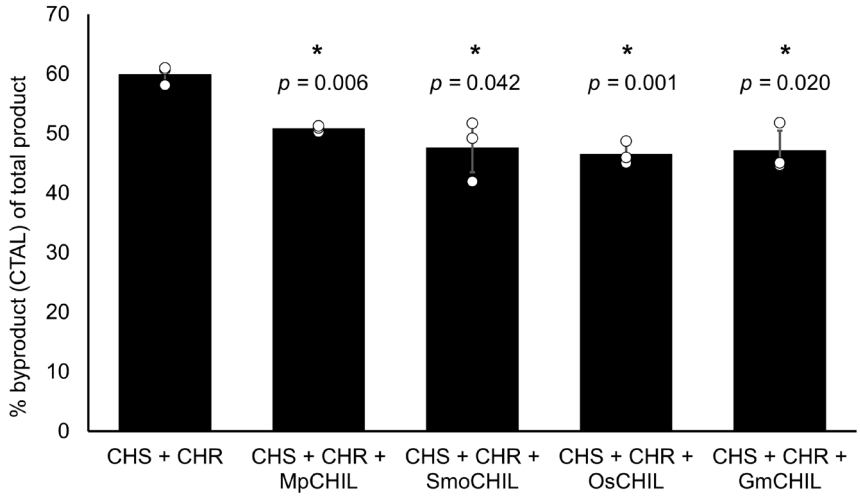


Figure 9. Reduction of percent byproduct formation in coupled assays with CHS, CHR, and CHIL. The y-axis indicates byproduct (CTAL) as a percentage of total reaction products (naringenin chalcone, naringenin, isoliquiritigenin, and CTAL). The data represents the mean ± standard deviation of three independent assays (data points shown). *P* values indicate statistical significance as determined using a Student's two-tailed *t*-test with an equal assumption of variance ($p < 0.05$).

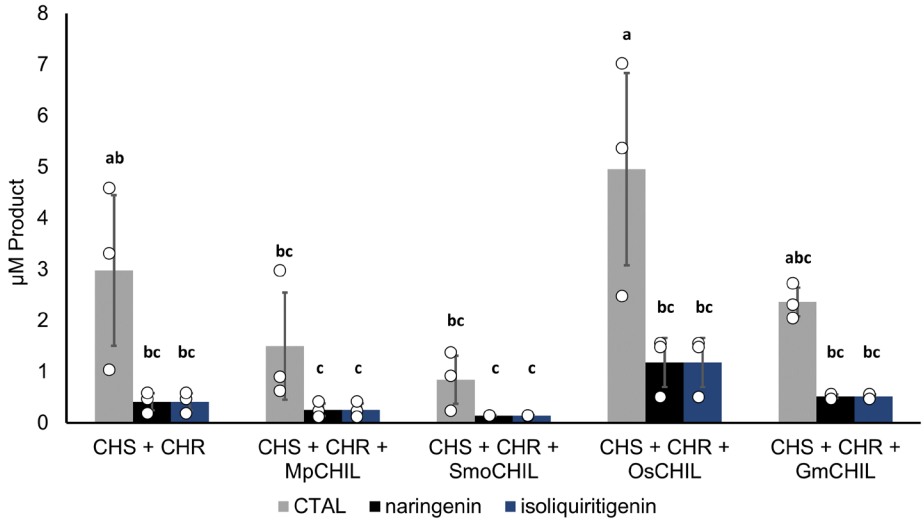


Figure 10. Product distribution of coupled assays with CHS, CHR, and CHIL. The production of each product, isoliquiritigenin (blue), naringenin (black), and CTAL (light grey), were quantified in coupled assays. Statistically significant differences between products within and between treatment types were determined using a two-way ANOVA with an HDS Tukey test.

4.6 CHIL inclusion increases flavanone production

The direct impact of CHIL orthologs in the context of flavanone production was identified using four-protein reactions, including CHS, CHR, GmCHI1B2 (hereafter CHI), and each CHIL ortholog (**Figure 11**). The inclusion of 1 μg of CHI allowed for the direct assessment of liquiritigenin production as GmCHI1B2 belongs to the type II group (**Figure 3**). As expected, the production of naringenin chalcone was not detected during these assays; however, trace amounts (0.3 μM – 1.2 μM) of isoliquiritigenin were produced in all cases (**Figure 11a**). Additionally, metabolic flux continued to be equally directed toward either branch of the pathway as the titres of liquiritigenin and isoliquiritigenin are between ± 0.08 μM to ± 0.4 μM of naringenin.

Consistent with previous assays (**4.3** and **4.5**), coupled reactions with CHS, CHR, and CHI led to a significant excess of CTAL at 59.05 % of total products, again showing that CHR and CHI do not rectify the catalytic promiscuity of CHS (**Figure 11b**). Notably, MpCHIL could not significantly reduce byproduct formation at 41.4 %. However, the inclusion of SmoCHIL, OsCHIL, and GmCHIL significantly reduced byproduct synthesis by 17.3 %, 25.2 %, and 17.25 %, respectively.

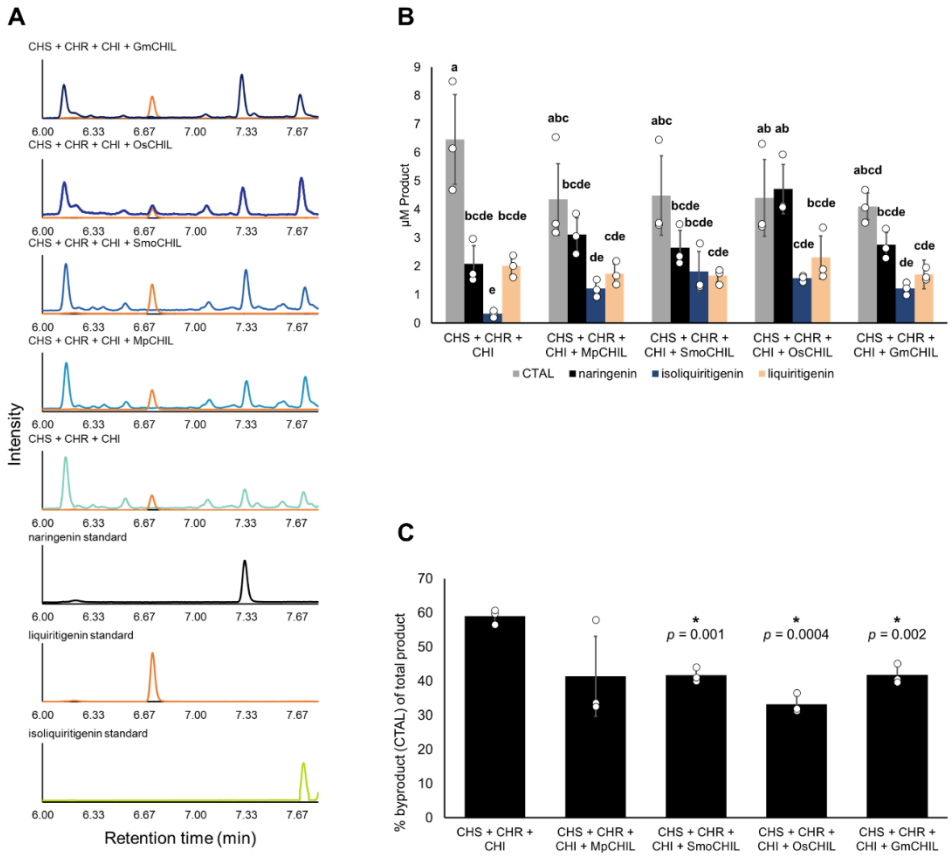


Figure 11. CHIL enhances flavanone production *in vitro.* **A**) Representative chromatogram of products when CHS is coupled with CHR, CHI, and CHIL orthologs, yielding CTAL, naringenin, isoliquiritigenin ($UV_{280\ nm}$) and liquiritigenin ($UV_{254\ nm}$). **B**) Product distribution of coupled assays. **C**) Percent byproduct of each assay when CHS, CHR, and CHI are coupled with CHIL orthologs. A statistically significant reduction in byproduct is indicated by an asterisk and *p* value and was calculated using a two-tailed Student's *t*-test with an equal assumption of variance.

4.7 FAP3s are the most recent ancestor of CHILs in soybean

We aimed to establish the residues sufficient for the evolution of CHIL activity from the FAP sister clade, which has been speculated as being an evolutionary progenitor. FAPs have been grouped into the larger CHI protein family and are also known as type III CHIs. Based on the phylogenetic analysis here (**Figure 3**) and previous reports, FAP3s appear to be the most recent common ancestor of CHILs and *bona fide* CHIs (Ngaki *et al*., 2012, Kaltenbach *et al*., 2018). Two soybean FAP3 isoforms (GmCHI3A1; Glyma.13G262500 and GmCHI3A2; Glyma.15G242900) have been identified; we decided to proceed with GmCHI3A2, hereafter referred to as GmFAP (**Appendix 1**). The two isoforms share 95 % amino acid identity and 24.87 % and 25.27 %, respectively, with GmCHIL (**Appendix 2**).

The FAP subfamily is also plastidic (Ngaki *et al*., 2012, Dastmalchi & Dhaubhadel, 2015), with an N-terminal chloroplast transit peptide, which is absent in the rest of the CHI family (**Appendix 6**). TargetP (https://services.healthtech.dtu.dk/services/TargetP-2.0/) analysis revealed the presence of a 68-amino acid N-terminal transit peptide. The amino acid percentage identity of truncated GmFAP, without the transit peptide, to GmCHIL was 26 % (**Appendix 2**). An examination of previously resolved crystal structures of AtFAP3 (PDB ID, 4DOL) and AtCHIL (PDB ID, 4DOK) in PyMOL revealed significant structural overlap (**Figure 12a**).

Homology modelling was completed using SWISS Model (https://swissmodel.expasy.org/) to predict the three-dimensional structure of GmFAP, and GmCHIL (**Figure 12a-b**) based on the Arabidopsis proteins (AtFAP3, and AtCHIL). GmFAP retained the interaction with the fatty acid

substrate, palmitic acid, seen in the Arabidopsis model. Similarly, GmCHIL had significant overlap with AtCHIL.

The functional utility of FAPs as binding proteins, similar to the speculated activity of CHILs in tetraketide or chalcone-binding led us to the hypothesis that a minimum number of sites (residues) could be mutagenized to swap functions. FAPs have a slightly larger active site to bind bulky and hydrophobic C-12 hydrocarbon structures (Ngaki *et al*., 2012). As previously mentioned, GmCHI3A2 was selected as the candidate GmFAP for pairwise comparison to GmCHIL due to the higher conservation of amino acid residues (**Appendix 2**).

4.8 Six active-site residues may be responsible for CHIL function

Structure and sequence comparison of GmFAP and GmCHIL revealed several significant mutations (in the direction of CHIL) at active site residues (**Figure 12c-d**; **Appendix 7**). Some of these mutations were previously characterized as key residues for *bona fide* CHI catalytic activity (Kaltenbach *et al*., 2018); including mutations in Tyr106, Asn113, and Thr190. Specifically, we have identified 6 "tier-1" targets as potentially sufficient for CHIL function. Using the amino acid number for GmFAP: Glu109Asn, Val119Phe, Ala122Ile, Phe177Tyr, Val257Trp, and Phe258Tyr are changes that line the active site, with proximity (< 7 Å) to the substrate. Single or combinatorial changes at these residues should be sufficient and necessary to accommodate a polyketide substrate and introduce additional polar contacts with naringenin chalcone. Therefore, future work can be aimed at site-directed mutagenesis of these six residues in GmFAP to establish CHIL function.

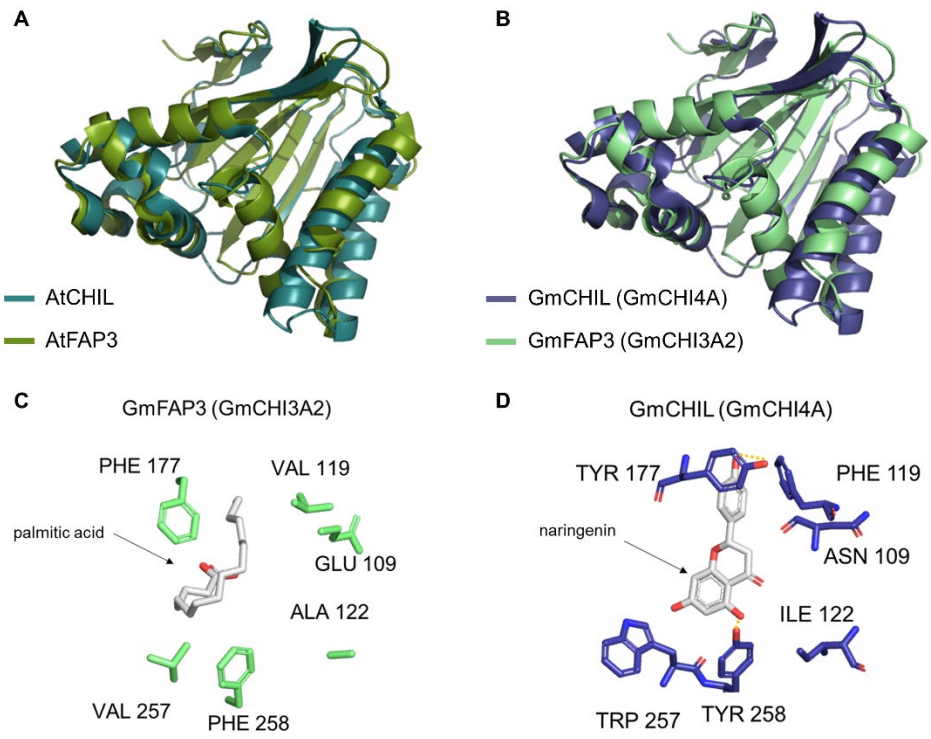


Figure 12. Three-dimensional structures of related CHI proteins from Arabidopsis and soybean. A) PyMOL (Schrödinger) modelling was completed using the PDB files for AtCHIL (4DOK) and AtFAP3 (4DOL). **B**) SWISS model (https://swissmodel.expasy.org/) was used to generate predicted 3D structures for both GmCHIL and GmFAP based on previously resolved structures. **C**) Six amino acid residues are displayed for GmFAP (full length): Glu109, Val119, Ala122, Phe177, V257, and Phe58. These residues show the ability to accommodate palmitic acid. **D**) In GmCHIL the same residues have mutated to Asn109, Phe119, Ile122, Tyr177, Trp257, and Tyr258, and the protein interacts with naringenin. Polar regions are indicated by blue or red coloration and contacts with a yellow line.

CHAPTER V: DISCUSSION

This research has focused on the ability of CHIL isoforms from across the plant kingdom to direct the activity of CHS, enhancing flavanone biosynthesis by reducing byproduct accumulation. We have shown a differential capacity in our selection of CHILs, with GmCHIL and OsCHIL outperforming the others in CHS-CHIL coupled assays, increasing naringenin production. In legume-specific three-protein assays, with CHS, CHR, and CHIL, all four orthologs significantly reduced byproduct levels and channelled flux into both the 6-deoxychalcone (isoliquiritigenin) and trihydroxy (naringenin) flavanone branches. The inclusion of CHIL, when reconstituting the entire pathway for flavanone biosynthesis *in vitro* (CHS, CHR, and CHI), continued the trend of repressing carbon loss.

The function of CHIL in rectifying CHS activity, belies the question of why this problem arises. Catalytic promiscuity is believed to serve as a basis for the functional evolution of novel metabolic routes in metabolism. Protein scaffolds and catalytic machinery implicated in primary metabolism can develop novel functions (neofunctionalization) in specialized metabolism through progressive neutral or advantageous mutations (**Figure 2a**) (Weng, 2012, Moghe, 2015). For example, the CHS protein family likely evolved from a *β*-ketoacyl-acyl carrier protein synthase (EC 2.3.1.180) responsible for fatty acid biosynthesis (Austin & Noel, 2003). Furthermore, it is conjectured that the promiscuity of CHS activity allowed for the development of stilbene and *p*-coumaroyltriacetic acid metabolism via subfunctionalization into STS and CTAS (Tropf *et al.*, 1994, Akiyama, 1999). Enzymes show three levels of specificity or selectivity: substrate(s) and cofactor(s), reaction mechanisms, and product(s). CHS and CHS-like

proteins, without the addition of new protein domains, have demonstrated catalytic promiscuity for substrates and products, while maintaining the reaction mechanism (Ferrer et al., 1999, Jez et al., 2000a, Abe & Morita, 2010). The typical result of CHS catalytic promiscuity, which forms the focus of the current research, is the production of lactone byproducts (BNY and CTAL). This side-branch does not appear to serve a physiological role; nevertheless it can represent an evolutionary access point to novel chemical structures (Waki *et al.*, 2021). However, from a synthetic biology perspective, such promiscuity or lack of functional acuity can lead to "loss of carbon" and metabolic flux to undesired products.

Auxiliary proteins are non-catalytic players involved in biochemical reactions, typically by binding a *bona fide* enzyme, substrate, product, or reaction intermediates. In the most positive sense, seen through an anthropogenic lens, we can view auxiliary proteins as significant corrective tools to contravene enzymatic promiscuity and to direct metabolic flux towards physiologically-relevant or value-add products. In the examples of NISO (morphine pathway) and Fra a (flavonoid pathway), auxiliary proteins can shift isomeric chemical mixtures in one direction or render a novel chemical profile; however, the exact mechanisms remain obscure (Muñoz *et al.*, 2010, Dastmalchi *et al.*, 2019). CHIL has demonstrated the ability to rectify the catalytic promiscuity of CHS, without significantly changing the total reaction output. However, the *in vitro* characterization of CHIL-CHS coupled assays has been limited to homo-specific protein pairs (from the same species) (Ban *et al*., 2018, Ni *et al*., 2020, Waki *et al*., 2020, Liu *et al*., 2021). Here we go beyond that framework, selecting CHILs from across the plant kingdom and using them as "parts" in an *in vitro* coupled assay approach.

First, CHS activity was confirmed using the standard assay (**3.8**). Expectedly, the production of naringenin significantly exceeded the production of naringenin chalcone under most conditions (**Figure 5**). Naringenin chalcone can spontaneously convert into a racemic mixture of (2*S*/2*R*)-naringenin at an alkaline pH (Bednar & Hadcock, 1988). Therefore, at pH 7.5, the spontaneous conversion of chalcone to flavanone, without the addition of CHI, explains the accumulation of naringenin in all samples. Additionally, BNY was expected to be produced in all cases. However, its production seemed to be unchanging at low concentrations *in vitro* (Waki *et al*., 2020), and the expected *m*/*z* ratio (229.1) was not detected in any of the sampled assays. Furthermore, there are no authentic standards for either CTAL or BNY and we calculated an approximation of CTAL concentrations using a naringenin chalcone standard curve. Naringenin chalcone has the same computed molar mass (271.22 g/mol) as CTAL and can be expected to ionize similarly.

In the standard assay, CHS produced an 11 % excess of CTAL compared to naringenin and naringenin chalcone (**Figure 7**). Previous *in vitro* characterization using the isoform GmCHS7 exhibited a 5 % molar excess of CTAL without CHIL (Waki *et al*., 2020). In soybean, GmCHS7 and GmCHS8 isoforms likely arose following a gene duplication event and share 98 % amino acid identity, and they are both implicated in isoflavonoid biosynthesis (Kumaran et al., 2018). Therefore, while we have opted for GmCHS8 in our assays, we do not expect the results to vary significantly compared to GmCHS7.

An investigation of CHIL function in tandem with CHS and CHR has previously been overlooked, likely due to a focus by researchers on non-legume model systems, such as

Arabidopsis, and their corresponding pathways. Furthermore, the literature has not suggested direct interaction between CHIL and CHR in soybean, perhaps leading researchers to cast off the idea of the former enhancing isoliquiritigenin production (**Appendix 8**). However, as the concept of a supramolecular complex involving CHS, CHI, and others has been indicated for (iso)flavonoid biosynthesis, the inclusion of CHIL in legume-specific pathways can be expected (Dastmalchi *et al*., 2016, Waki *et al.*, 2016, Mameda *et al*., 2018). Furthermore, CHR and CHIL may act on the same spontaneous and unstable intermediate of the CHS-catalyzed reaction. While the exact substrate has not been confirmed, CHR is theoretically predicted to bind *p*-coumaroylcyclohexantrione (Bomati *et al*., 2005). Similarly, modelling studies with AtCHIL predict that the most likely ligand is *p*-coumaroyl-tetraketide-CoA, but this has not been confirmed experimentally (Waki *et al*., 2020).

Reactions with CHS and CHR produced isoliquiritigenin, naringenin, and CTAL (**Figure 8**). Following previous work, naringenin chalcone is converted into naringenin, while, in the parallel pathway, isoliquiritigenin does not spontaneously rearrange into liquiritigenin (Mameda *et al*., 2018). Liquiritigenin was only present in reactions supplemented with CHI (**Figure 11a**). Notably, the CHI isoform selected for this study belongs to the type II legume classification (**Figure 3**), accepting either chalcone scaffold for stereospecific intramolecular isomerization into (2*S*)-flavanone products (**Figure 1**) (Bednar & Hadcock, 1988).

As expected, CHR alone did not rectify the catalytic promiscuity of CHS, with the reaction accumulating a 59.8 % majority of byproducts (**Figure 9**). Interestingly, in CHR-catalyzed reactions, we saw an equimolar division into isoliquiritigenin and naringenin (**Figure 10**).

Previous *in vitro* characterization of soybean CHR isoforms revealed that the production of isoliquiritigenin never exceeds 50 % of total products while utilizing a molar excess of CHR isoforms to CHS (Mameda *et al*., 2018). It is proposed that during the course of the CHS and CHR coupled reaction, the linear tetraketide of which CHR acts upon diffuses out of its active-site prior to its reduction (Nakayama *et al.*, 2019). Similarly, in these findings, the total production of isoliquiritigenin never exceeded that of naringenin. Notably, previous characterizations using 0.05 μM CHR have noted marginal isoliquiritigenin production (> 25 % of total products). Further, the production of CTAL exceeded the production of either canonical product.

Liquiritigenin is the exclusive precursor to certain isoflavonoid phytoalexins, including many pterocarpans (Jiménez-González et al., 2008). Furthermore, RNAi silencing of soybean CHR14A led to increased susceptibility to infection from *P. sojae* (Graham *et al.*, 2007). Following elicitation, legume species may increase the production of liquiritigenin-derived products. Interestingly, CHIL does not appear to enhance the production of liquiritigenin at the expense of naringenin when coupled with CHS and CHR *in vitro* (**Figure 11**). Therefore, additional characterization of the regulatory function of CHIL in legume plant models following elicitation may uncover additional roles within this pathway.

Quantifying byproduct (% of total product) levels in our assays was used as a rapid means to interpret the impact of CHIL inclusion, in two, three, or four-protein mixtures. However, we noticed that the overall output did change from one “treatment” to another. For example, in the CHS-CHIL reactions, MpCHIL and SmoCHIL seem to have an inhibitory effect on the overall

metabolic output. Specifically, the average total products are 1.2 to 1.8-fold lower than the reactions containing just CHS or CHS-CHR, respectively (**Figure 7**, **Figure 9**). Technical error in substrate availability may explain the observed lower output of MpCHIL and SmoCHIL coupled assays.

Direct protein-protein interactions between CHS and CHIL are proposed to be critical for the corrective function (Ban *et al*., 2018, Waki *et al*., 2020). However, two CHIL isoforms from *P. patens* (PpCHIL-A and PpCHIL-B) and MpCHIL did not interact with Arabidopsis CHS (**Appendix 8**). Additionally, bryophyte CHIL isoforms were unable to establish a wild-type phenotype in *Atchil3* loss-of-function mutants with a transparent seed coat phenotype and a reduced production of proanthocyanidins (Waki *et al*., 2020). Deciphering the differences between CHIL isoforms that render GmCHIL and OsCHIL superior to their bryophyte/lycophyte counterparts will be essential for future structure-function analyses.

Remarkably, coupling MpCHIL or SmoCHIL to CHS and CHR significantly reduces the percentage byproduct by 9 % and 12.3 %, respectively (**Figure 8**), a trend that persists with CHI inclusion (**Figure 11c**). Introducing CHR can potentially alter the conformation of the CHS active site, exposing the reaction intermediates to Mp/SmoCHIL function. Such potential has been evidenced by PpCHIL-A (bryophyte) binding isoliquiritigenin, when incubated in a thermal shift assay (Wolf-Saxon *et al.*, 2023). Such preference for isoliquiritigenin, might explain why these CHILs function in the CHS-CHR mixture, but not with CHS alone.

We have demonstrated that the effect of OsCHIL, a CHIL ortholog with 60 % amino acid sequence identity with GmCHIL (**Appendix 2**), can reduce the percent byproduct accumulation

by 13.3 % when coupled to CHS. Previously, researchers have noted a greater impact for OsCHIL (40 % reduction in CTAL); however, these assays utilized the homospecific OsCHS, which produces greater levels of CTAL (25 % excess compared to 5 % excess for GmCHS) (Waki *et al*., 2020). Therefore, the impact of the rice isoform cannot be judged similarly across both experiments.

While a physical interaction between OsCHIL and GmCHS8 has not been established (**Appendix 8**), there is a statistically significant effect on the overall byproduct output (p = 0.004) in our *in vitro* assays (**Figure 8**). The only reported interactions between CHS and CHIL have involved homospecific isoforms, or weak associations between members from closely related taxa. Specifically, VvCHIL and ImCHIL have been shown to interact with AtCHS (all angiosperm dicot species) (Waki *et al.*, 2020). It is unclear whether CHIL from rice (monocot) could physically interact with CHS isoforms from distant species (e.g., Fabaceae); however, we do see the positive corrective influence of OsCHIL on GmCHS8.

Finally, the inclusion of GmCHIL reduced byproduct levels in all coupled assays: with CHS (30 %), CHS-CHR (12.8 %), or CHS-CHR-CHI (21.1 %) (**Figure 6**, **Figure 8**, **Figure 11b-c**). Consistent with previous literature, the overall product composition does not change significantly, but the increase of naringenin (1.6-fold) appears to be gained from a reduction in CTAL (1.7-fold) (**Figure 7**) (Waki *et al.*, 2020). Therefore, it appears that carbon inherently lost to byproducts in an *ex planta* system can be reallocated to the canonical pathway with the inclusion of CHIL.

Additional experiments will be required to assess the impact of CHIL on downstream isoflavonoid biosynthesis, building on the utility evidenced here in producing both precursors (liquiritigenin and naringenin). Notably, GmCHIL has been shown to interact with GmIFS1 and GmIFS2 (**Appendix 8**). Due to the limitation of working with cytochrome P450s *in vitro* assays with IFS were not attempted. IFS would have to be prepared from yeast endosomes, without accurate molar quantification. Furthermore, an *in vitro* reassembly of the isoflavonoid pathway (including > 5 proteins), lacking metabolon formation, might not be an accurate facsimile of the plant context. Future work could be attempted in an engineered microbial or algal strain to recreate the membrane-tethered nature of the complex.

The inclusion of CHIL homologs within synthetic biology modules has been reconstituted in yeast and *E. coli* hosts. Specifically, *E. coli* cultures demonstrated a 14 % increase in naringenin titres when incorporating the heterospecific isoforms *Petunia hybrid* CHS (dicot) and *M. domestica* CHIL (dicot) compared to strains lacking CHIL (Liu et al., 2022). Furthermore, the overall naringenin output increased disproportional to CTAL reduction, indicating enhanced metabolic flux with PhCHS and MdCHIL strains. Similarly, in a yeast model, heterospecific coupling of *Erigeron breviscapus* CHS (dicot) and *S. bicolor* CHIL (monocot) enhanced naringenin titres while reducing CTAL production (Meng *et al.*, 2022). These studies demonstrate that CHIL can function *in vivo*, guiding product biosynthesis toward downstream isoflavonoid derivatives, including genistein.

In lieu of microbial reconstruction or *in vitro* assays, we can return to investigations of *in planta* CHIL function for greater detail of physiological impact. Studies of loss-of-function *chil*

mutants in Japanese morning glory and Arabidopsis revealed reduced flavonoid and anthocyanin accumulation (Morita *et al.*, 2014, Jiang *et al*., 2015). Similarly, model legume organisms, such as *Medicago truncatula* could be used to study the effect of CHIL knockout mutants on isoflavonoid biosynthesis. *M. truncatula* is a useful legume model organism due to its relatively small and fully-annotated genome, amenability to functional genomics tools, and short generation time. The division of Agricultural Sciences and Natural Resources at Oklahoma State University has curated thousands of T-DNA insertion loss of function mutants in *M. truncatula*, including two mutant lines with disruptions to the CHIL coding region (NF7721 and NF17281) (Tadege *et al.*, 2008). Knockout lines can be used to produce a detailed chemical profile of isoflavones (daidzein and genistein) and downstream derivatives (e.g., biochanin A and formononetin).

An evolutionary connection between fatty acid metabolism and flavonoid biosynthesis has been well-established based on the structural, mechanistic, and phylogenetic relationships between key enzymes in both pathways, including CHS and CHR (Austin & Noel, 2003, Bomati *et al*., 2005). We can add the CHI family to this list, with their speculated evolution from FAPs, as illustrated in **Figure 3**. The most recent common ancestor are the FAP3s. The proposed CHIL substrate likely consists of 12 carbon molecules and resembles the preferred ligand of GmFAP, palmitic acid (C-12) (Ngaki *et al*., 2012, Kaltenbach *et al*., 2018, Waki *et al*., 2020).

CHILs likely evolved their function through minimal changes to the active site of FAP3s, whereby additional polar contacts and cavity rigidity enabled them to accept chalcone intermediates. Reported here is a curated set of six amino acid residues in GmFAP that are

substituted in CHILs. Each residue is generally conserved in CHIL sequences but is substituted in the GmFAP sequence (**Appendix 7**). Specifically, using the numbering for GmFAP, we suggested that mutations at the following residues would introduce new polar contacts with the polyketide scaffold: Glu109Asn, Phe177Tyr, and Phe258Tyr (**Figure 12c-d**). Additionally, Val119Phe, Ala122Ile, and Val257Trp might provide rigidity to the chalcone-binding cleft, enabling the transition from FAP to CHIL activity.

During the preparation of this study, another group reported analysis of the Trp257 residue by site-directed mutagenesis (Wolf-Saxon *et al*., 2023). Trp257 is a largely conserved site amongst CHIL orthologs but replaced with other amino acids in PpCHIL-A/B and GmFAPs. Site-directed mutagenesis of PpCHIL-A with Leu257Trp, significantly enhanced protein thermostability, when bound to naringenin chalcone, suggesting an increased positive affinity. Conversely, in VvCHIL, a Trp257Leu mutation did not alter the binding affinity. These seemingly contradictory results obscure the probable role of the residue in stabilizing a chalcone intermediate. However, exploration of this residue in conjunction with the other five, outlined above, in combinatorial mutants could render a FAP-to-CHIL conversion with a parsimonious mutagenesis regime.

CHAPTER VI: CONCLUSION

6.1 General conclusions

We hypothesized that CHIL orthologs from *M. polymorpha, S. moellendorffii, O. sativa*, and *G. max* could regulate metabolic flux in flavanone biosynthesis by repressing byproduct formation. The findings of this study suggest that CHIL orthologs from diverse origins can improve flavanone biosynthesis and reduce byproduct formation in four-protein reactions with CHS, CHR, and CHI. Thus, we validated that CHIL does not require a homospecific isoform of CHS to regulate production. Furthermore, we confirmed that CHR and CHI alone do not reduce byproduct accumulation *in vitro.* Notably, we did see differential capacities when CHILs were complementing CHS activity directly, with GmCHIL and OsCHIL rendering the best outcomes. Finally, the inclusion of CHIL can direct metabolic flux equally toward either flavanone product (liquiritigenin and naringenin), suggesting putative utility in downstream isoflavonoid applications.

We also hypothesized that CHIL function was established through amino acid substitutions within the ligand-binding pocket FAP3 ancestors. Using *in silico* predictive tools, we have identified six amino acid residues that may be critical for establishing soybean CHIL function in the most recent common ancestor, GmFAP3. Through homology modelling and ligand interactions in SWISS model and PyMOL, these residues were proposed to establish new polar contacts and increase the structural rigidity of the active site to accommodate polyketides. These curated residues should serve as “tier 1” targets for site-directed mutagenesis to uncover the molecular mechanism and evolution of the CHIL protein family.

6.2 Contributions to Science

By completing **Objectives 1** and **2**, this study demonstrated that CHIL function from *M. polymorpha*, *S. moellendorffii*, *O. sativa*, and *G. max* extends beyond the conserved pathway to naringenin chalcone (2′,4,4′,6′-tetrahydroxychalcone) and highlights their utility in isoflavonoid precursor production, specifically the flavanones naringenin (4′,5,7-trihydroxyflavanone) and liquiritigenin (4′,7-dihydroxyflavanone). This study also suggests that CHIL function may not be restricted to species-specific CHS, CHR, and CHI isoforms. Finally, the completion of **Objective 3** has provided six amino acid residue targets for future investigations into understanding CHIL function, mechanism, and evolutionary transformation from FAPs. These targets include previously identified residues critical for *bona fide* CHI function (Tyr177), thermostability when binding to naringenin chalcone (Trp257), and others which have not been investigated.

6.3 Future directions

We can further expand our *in vitro* characterization of CHIL isoforms by resolving their crystal structure, helping us to determine the physical basis for functional differences. Furthermore, we would like to substantiate physical interactions between CHIL and its catalytic partners (**Appendix 8**). While MpCHIL reportedly does not interact with AtCHS, there is a need to understand why the activity changes when CHR is included (Waki *et al.*, 2020). Therefore, an important next step would be to utilize assays, such as co-immunoprecipitation, for multi-component protein interaction analysis.

Two major limitations of the present study are that *in vitro* reaction conditions vary from the physiological context, and the direct impact of CHIL on isoflavonoid biosynthesis was not measured. Future work using *in vivo* models, such as *S. cerevisiae* or *M. truncatula,* could solve both concerns. Specifically, integrating the complete isoflavonoid biosynthetic chassis into yeast with each CHIL ortholog should further support the evidence provided here. Additionally, the aforementioned T-DNA insertion *chil* mutants (NF7721 and NF17281) in *M. truncatula* can be used as tools for an unbiased isoflavonoid metabolomic analysis and *in planta* complementation assays.

Lastly, **Objective 3** highlighted amino acid targets to establish CHIL function in FAPs. Site-directed mutagenesis of each residue, shown in **Figure 12**, can be used to screen GmFAP3 mutants for CHIL-associated ligand binding, protein-protein interactions with CHS and CHR, and the residue(s) required to rectify CHS catalytic promiscuity.

Explicitly, DSF can be used to determine if any GmFAP mutants can bind to naringenin chalcone. Swapping FAP affinity for fatty acids, such as palmitic acid, with a chalcone scaffold would provide proof for this concept. Further, the novel mutants can be deployed in place of CHIL for *in vitro* coupled assays, to quantify reduction of byproducts (CTAL). Finally, proof of FAP mutant interaction with CHS can be assayed by bimolecular fluorescence complementation (BiFC) assays. Briefly, the gene encoding a fluorophore, *yellow fluorescent protein* (YFP), will be split and translationally fused to the target proteins. Subsequently the cloned genes will be transiently expressed in *Nicotiana benthamiana* (*Agrobacterium tumefaciens*-mediated infiltration) and assayed for interaction with confocal laser scanning microscopy (Mehlhorn *et al.*, 2018). TargetP analysis and subcellular localization indicated that GmFAP and GmCHS isoforms localize to distinct subcellular compartments (**Appendix 6**) (Ngaki *et al*., 2012, Dastmalchi & Dhaubhadel, 2015, Dastmalchi *et al*., 2016). However, it remains to be shown whether a truncated GmFAP, with the transit peptide removed, can interact with CHS. Therefore, future work must first establish if this interaction can occur before proceeding through site-directed mutagenesis.

Overall, by understanding how the biological function of CHIL was established, we can resolve a missing link in the evolution of the CHI protein family and potentially provide the context required to engineer a more efficient auxiliary protein.

CHAPTER VII: APPENDICES

Appendix 1. Gene identifiers used for phylogenetic analysis.

Name	Gene ID (Phytozome)	Species
AtCHI	AT3G55120	*Arabidopsis thaliana*
AtCHI2	AT5G66220	*Arabidopsis thaliana*
AtCHI3	AT5G66230	*Arabidopsis thaliana*
AtCHIL	AT5G05270	*Arabidopsis thaliana*
AtFAP1	AT3G63170	*Arabidopsis thaliana*
AtFAP2	AT2G26310	*Arabidopsis thaliana*
AtFAP3	AT1G53520	*Arabidopsis thaliana*
GmCHI1A	Glyma.20G241500	*Glycine max*
GmCHI1B1	Glyma.20G241600	*Glycine max*
GmCHI1B2	Glyma.10G292200	*Glycine max*
GmCHI2	Glyma.20G241700	*Glycine max*
GmCHI3A1	Glyma.13G262500	*Glycine max*
GmCHI3A2	Glyma.15G242900	*Glycine max*
GmCHI3B1	Glyma.03G154600	*Glycine max*
GmCHI3B2	Glyma.19G156900	*Glycine max*

GmCHI3C1	Glyma.14G098100	*Glycine max*
GmCHI3C2	Glyma.17G226600	*Glycine max*
GmCHI4A	Glyma.06G143000	*Glycine max*
GmCHI4B	Glyma.04G222400	*Glycine max*
MpCHI	Mapoly0167s0012	*Marchantia polymorpha*
MpCHIL	Mapoly0175s0004	*Marchantia polymorpha*
MpFAP2	Mapoly0058s0114	*Marchantia polymorpha*
MpFAP3_1	Mapoly0044s0117	*Marchantia polymorpha*
MpFAP3_2	Mapoly0077s0003	*Marchantia polymorpha*
OsCHI	LOC_Os03g60509	*Oryza sativa*
OsCHIL-1	LOC_Os11g02440	*Oryza sativa*
OsCHIL-2	LOC_Os12g02370	*Oryza sativa*
OsFAPa1	LOC_Os07g38390	*Oryza sativa*
OsFAPa2a	LOC_Os06g10210	*Oryza sativa*
OsFAPa2b	LOC_Os02g53810	*Oryza sativa*
OsFAPb	LOC_Os02g21520	*Oryza sativa*
PpCHIL-A	Pp3c4_25770	*Physcomitrella patens*
PpCHIL-B	Pp3c26_4040	*Physcomitrella patens*

PpFAP1	Pp3c11_25580	*Physcomitrella patens*
PpFAP2	Pp3c21_9340	*Physcomitrella patens*
PpFAP3	Pp3c10_20100	*Physcomitrella patens*
SmoCHI	151616	*Selaginella moellendorffii*
SmoCHIL	227414	*Selaginella moellendorffii*
SmoFAP2	112032	*Selaginella moellendorffii*
SmoFAP3	16313	*Selaginella moellendorffii*
SmoFAPa_1	428339	*Selaginella moellendorffii*

Gene Name																																			
HlCHIL1	100																																		
SmoFAPa1	31.22	100																																	
OsFAPa1	31.16	47.69	100																																
AtFAP1	33.17	45.02	53.49	100																															
GmCHI3B2	33.81	47.83	62	59.26	100																														
GmCHI3B1	31.43	44.39	59.5	56.87	92.88	100																													
AtFAP2	26.07	32.07	28.02	26.91	27.68	27.17	100																												
GmCHI3C2	25.12	31.78	30.12	29.35	32.35	31.2	50.26	100																											
GmCHI3C1	26.07	31.78	30.5	29.35	33.09	31.95	50.39	92.72	100																										
SmoFAP2	26.97	36	39.34	35.33	37.36	37.91	41.62	43.24	44.32	100																									
MpFAP2	28.91	34.19	29.46	32.12	31.25	32.21	30.46	32.4	32.91	61.08	100																								
OsFAPa2b	24.17	32.07	28.74	27.8	32.23	30.71	43.47	43.62	43.51	43.78	34.79	100																							
OsFAPa2a	19.43	27.2	24.71	25.09	26.91	26.59	36.53	35.2	35.37	34.59	31.31	59.72	100																						
OsFAPb	13.5	21.52	19.46	17.23	16.17	15.42	14.5	15.65	16.79	18.97	16.79	15.53	13.91	100																					
AtFAP3	15.35	22.97	18.02	16.88	18.72	18.94	18.98	18.66	18.73	24.29	17.41	16.54	12.41	39.41	100																				
GmCHI3A2	19.49	20.75	19.72	18.26	18.34	18.55	21.64	20.6	19.55	24.86	18.53	19.85	14.87	39.44	55.64	100																			
GmCHI3A1	20.1	20.18	20.09	17.84	19.58	18.53	19.44	18.51	17.86	24.29	17.99	19.23	14.93	38.89	53.74	92.17	100																		
SmoFAP3	16.24	22.41	18.1	17.24	16.38	15.52	17.24	18.97	19.83	25.86	25.86	19.83	13.79	36.89	42.62	43.44	43.44	100																	
MpFAP3_2	17.73	22.67	24.55	23.75	23.33	21.98	21.63	22.18	22.99	25.42	21.77	21.66	17.56	33.33	33.81	36.23	34.28	42.62	100																
MpFAP3_1	19.61	20.35	20.18	22.54	21.31	21.61	21.18	21.35	22.42	24.72	22.18	19.43	16.14	32.6	34.89	38.75	37.02	46.72	70	100															
MpCHIL	14.95	17.28	20.74	16.33	19.29	19.29	18.32	18.81	18.81	19.55	16.83	17.33	14.36	23.27	22.93	22.44	22.44	21.31	24.15	23.08	100														
SmoCHIL	14.81	14.05	21.98	18.95	20.31	19.27	16.33	16.33	16.33	23.6	18.88	18.88	16.33	17.35	21.11	20.6	20.6	18.03	22.77	21.67	38.54	100													
OsCHIL1	14.66	16.49	20.21	20.92	21.13	20.62	19	19.5	19.5	18.78	19	18.5	14.5	20.71	26.5	21	21	20.66	22.77	23.65	41.46	39.3	100												
AtCHIL	12.44	16.84	18.18	18.97	18.88	17.86	17.91	16.92	16.92	20	19.9	15.92	14.93	20.6	23.27	24.26	24.26	20.66	25.49	25.85	36.71	37.44	61.84	100											
HlCHIL2	11.46	17.46	19.25	17.95	16.92	15.9	17.91	18.91	18.41	20	17.91	19.4	14.43	24	25.25	25.25	24.26	22.31	27.45	26.83	43	41.38	66.18	69.23	100										
GmCHI4A	14.58	15.34	17.65	13.85	17.95	17.95	16.92	17.91	17.91	23.33	18.41	14.93	11.94	19	25.25	25.25	24.26	18.18	25.98	26.83	37.2	43.35	59.9	61.06	67.94	100									
GmCHI4B	15.1	14.81	17.65	15.38	17.95	17.95	16.42	17.91	17.91	23.33	18.41	16.42	12.44	20	25.25	25.74	24.75	18.18	27.45	27.8	37.2	44.83	60.39	62.5	68.9	95.22	100								
MpCHI	14.98	15.53	16.82	17.95	17.83	18.14	17.9	17.34	16.94	19.44	19.05	15.87	15.81	16.67	19.35	19.48	18.73	20.66	22.66	25.29	25.96	23.27	27.32	26.21	24.27	25.73	24.27	100							
SmoCHI	17	19.29	19.37	18.5	22.28	21.29	19.71	18.75	18.75	21.67	21.15	21.63	13.94	21.18	21.26	23.79	24.15	30.83	25.24	25.59	25.73	20.9	24.26	24.51	23.65	21.18	21.18	33.8	100						
GmCHI1A	16.33	18.56	18.95	16.08	19.7	18.18	15.46	17.39	17.39	17.13	18.1	16.83	14.9	19.12	21.84	21.84	22.33	22.95	24.04	25.84	29.67	21.67	31.07	28.02	28.99	29.47	29.47	28.24	37.14	100					
GmCHI1B2	16.58	17.26	20.83	19.31	22.39	20.9	17.62	20.48	20.48	20.44	17.84	18.01	15.17	20.29	23.92	24.52	24.4	24.59	25.59	26.42	27.27	24.14	27.67	25.6	26.57	28.5	29.47	29.02	39.44	66.97	100				
GmCHI1B1	16.58	16.75	21.35	19.31	22.89	21.89	17.62	19.52	19.52	19.34	16.9	18.96	15.17	21.26	24.88	25	25.36	25.41	25.59	26.42	26.79	23.15	28.64	25.6	26.57	28.5	29.47	28.57	38.97	66.06	96.46	100			
OsCHI	18.78	22.05	22.51	23	24.12	22.61	21.15	23.08	23.56	21.98	21.8	20.1	16.27	19.02	25.12	25.12	24.64	23.77	26.32	27.62	27.14	17.65	28.99	27.88	28.37	28.85	28.85	30.53	34.43	52.29	45.29	45.74	100		
GmCHI2	16	18.18	22.28	17.73	21.29	19.8	17.54	17.54	17.54	20.44	16.82	18.4	16.04	17.79	23.33	25	24.29	20.49	25.47	24.88	25.84	20.2	27.67	27.05	27.54	28.99	28.02	29.65	36.62	48.17	50	50.44	53.57	100	
AtCHI	14.71	18.72	17.68	18.27	18.36	16.91	15.07	17.05	17.05	17.13	17.35	16.97	13.76	16.74	24.3	24.76	24.77	22.95	22.94	24.66	27.27	23.15	27.18	27.05	28.02	28.99	28.99	28.81	36.15	50	46.02	46.46	55.7	60.96	100

Appendix 2. Percent identity matrix of amino acid residues in the CHI protein family. Pairwise analysis of amino acids for each CHI type in *Marchantia polymorpha, Selaginella moellendorffii, Oryza sativa, Arabidopsis thaliana, Humulus lupus,* and *Glycine max*. Computed using Clustal ω (https://www.ebi.ac.uk/Tools/msa/clustalo/).

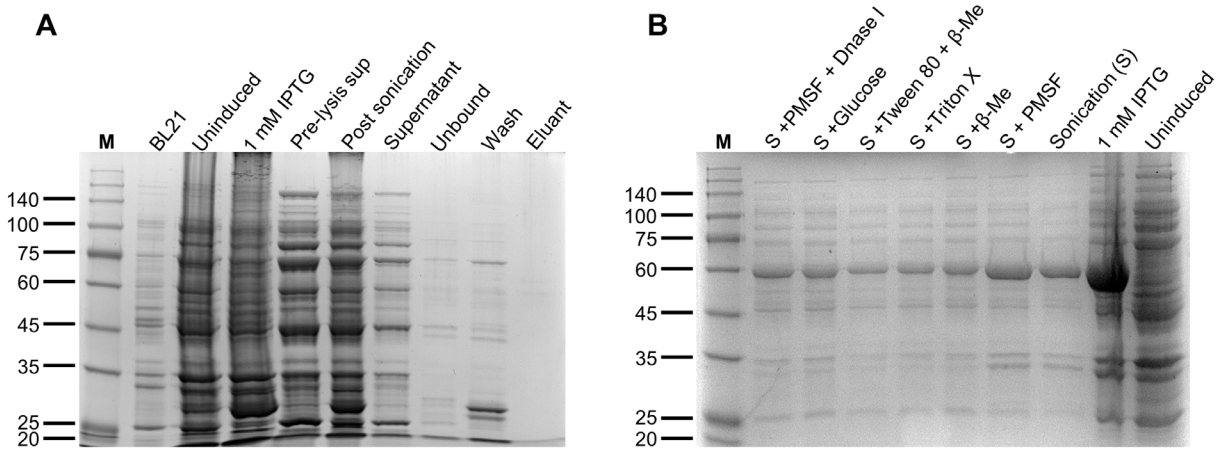
Appendix 3. Oligonucleotide list.

Gene Target	Sequence (5´-3´)	Amplicon size (bp)	Specifications
GmCHI1B2_F_GW	GGGG ACA AGT TTG TAC AAA AAA GCA GGC TTT GCC ACA CCA GCA TC	808	Gateway® cloning
GmCHI1B2_R_GW	GGGG ACC ACT TTG TAC AAG AAA GCT GGG T TCA GTT TTC ATT GTT GGG ATT GGC		Gateway® cloning
CHI1B2_F	ATGGCCACACCAGCATC	791	Insert-specific sequencing primer
CHI1B2_R	TCA GTT TTC ATT GTT GGG ATT GGC		Insert-specific sequencing primer
M13_F	TGT AAA ACG GCC AGT	-	*pDONR™/Zeo* specific sequencing primer
M13_R	CAG GAA ACA GCT ATG AC	-	*pDONR™/Zeo* specific sequencing primer
T7_F	TAATACGACTCACTATAGGG	-	pET vector-specific Sequencing primer
T7_R	GCTAGTTATTGCTCAGCGG	-	pET vector-specific Sequencing primer

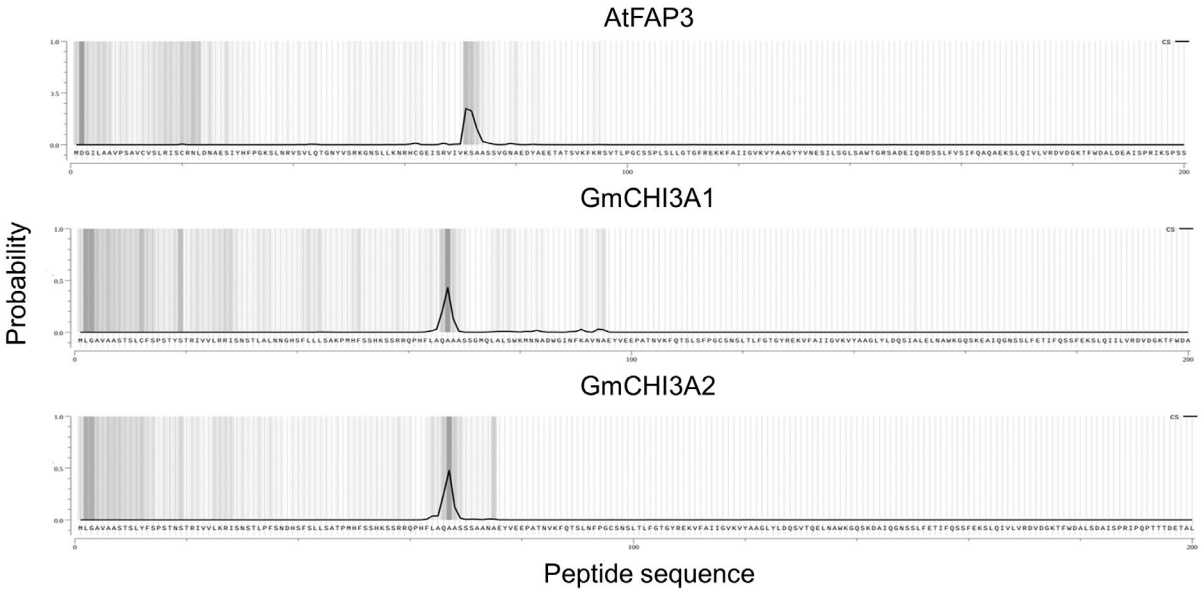
Appendix 4. Genes encoding proteins used in this study.

Gene name	Gene identifier*	Species	Base pairs	Expected kDa†	Vector
GmCHS8	Glyma.11G011500	*Glycine max*	1170	60	*pET32α* (+)
GmCHR14A	Glyma.14G005700	*Glycine max*	906	38.8	*pET160-DEST*
GmCHI1B2	Glyma.10G292200	*Glycine max*	791	28.9	*pET160-DEST*
MpCHIL	Mapoly0175s0004	*Marchantia polymorpha*	636	25.15	*pET28α* (+)
SmoCHIL	227414	*Selaginella moellendorffii*	618	24.73	*pET28α* (+)
OsCHIL-1	LOC_Os11g02440	*Oryza sativa*	639	25.7	*pET28α* (+)
GmCHIL (*GmCHI4A*)	Glyma.06G143000	*Glycine max*	630	25.66	*pET28α* (+)

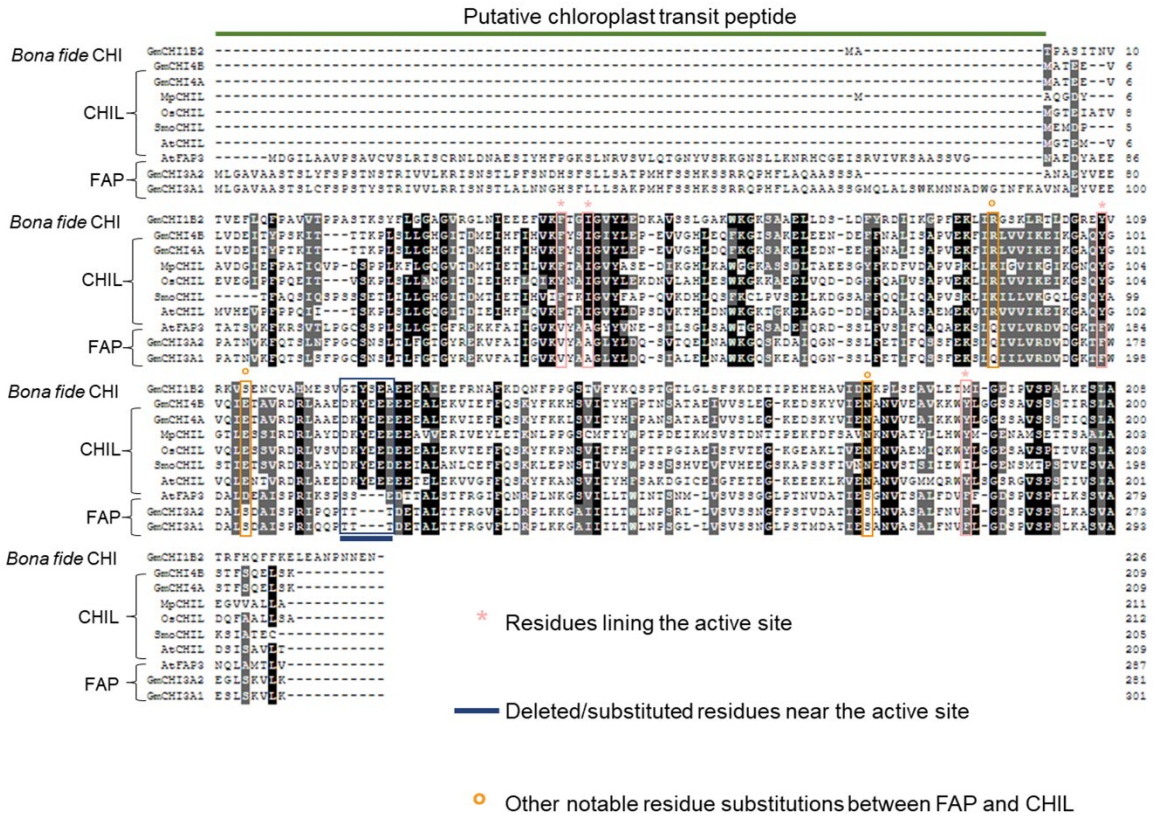
* Phytozome v13.0 (https://phytozome-next.jgi.doe.gov/) gene identifiers. † Expected molecular weight (kDa), including N-terminal tags.



Appendix 5. Troubleshooting protein purification. . Coomassie total protein staining depicting common issues during protein purification and troubleshooting steps**. A**) GmCHI1B2 (28.8 kDa) expression is induced; however, the expected band is not found in the supernatant and thus cannot be eluted from the Talon© resin. **B**) Solubility test of GmCHS8 (60 kDa) protein with additional lysis additives (protease inhibitors, detergents, reducing agents, and osmolytes). Proteins were resuspended in 10 ml of lysis buffer (Tris-HCl pH 7.5, 300 mM NaCl, 10 % glycerol) per g of cell paste.



Appendix 6. GmFAP TargetP sequence characterization. TargetP prediction of N-terminal chloroplast transit peptides in GmFAPs (GmCHI3A1, GmCHI3A2), and AtFAP3 (https://services.healthtech.dtu.dk/services/TargetP-2.0/).



Appendix 7. GmFAP amino acid sequence characterization. Multiple sequence alignment generated with MUSCLE (MEGA X) and visualized using a Boxshade plot (bioinformatics.org). Using a 60 % threshold, the plot is black and grey at identical and similar amino acid residues, respectively. Tier 1 residues for CHIL activity are highlighted in red boxes with red asterisks. Blue highlights a region absent in FAP sequences but present in other CHI sequences. Orange boxes highlight other differences between FAPs and CHILs.

Appendix 8. Literature review of reported CHIL interaction partners to date.

CHIL isoform	Species	Interaction partner	Method(s)	Source*
AmCHIL	*Antirrhinum majus*	AmFNSII	BifC	(Waki *et al*., 2020)
AmCHIL	*Antirrhinum majus*	AmCHS	BifC	(Waki *et al*., 2020)
AtCHIL	*Arabidopsis thaliana*	AtCHS	yeast two-hybrid, Pulldown, Split luciferase, Bifc	(Waki *et al*., 2020, Ban *et al*., 2018)
GbCHIL	*Ginka biloba*	GbCHS	yeast two-hybrid, BifC	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmIFS1	BifC, yeast two-hyrbrid	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmIFS2	BifC, yeast two-hybrid	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmCHS-1	BifC, yeast two-hybrid	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmCHS7	BifC, yeast two-hybrid	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmCHR1	yeast two-hybrid (no interaction)	(Waki *et al*., 2020)
GmCHIL	*Glycine max*	GmCHR5	yeast two-hybrid (no interaction)	(Waki *et al*., 2020)
HlCHIL-2	*Humulus lupulus*	HlCHS_H1	yeast two-hybrid, Split luciferase	(Ban *et al*., 2018)
HlCHIL-2	*Humulus lupulus*	HlPT1	yeast two-hybrid, Split luciferase	(Ban *et al*., 2018)
HmCHIL	*Hydrangea macrophylla*	HmCHS	yeast two-hybrid, BifC	(Waki *et al*., 2020)
InCHIL	*Ipomoea nil*	LnCHS-D	yeast two-hybrid, BifC	(Waki *et al*., 2020)
InCHIL	*Ipomoea nil*	LnCHS-E	yeast two-hybrid	(Waki *et al*., 2020)
InCHIL	*Ipomoea nil*	AtCHS	yeast two-hybrid	(Waki *et al*., 2020)
LoCHIL	*Lindsaea orbiculata*	LoCHS1	yeast two-hybrid, Split luciferase	(Ni *et al*., 2020)
LoCHIL	*Lindsaea orbiculata*	LoCHS2	yeast two-hybrid, Split luciferase	(Ni *et al*., 2020)
LoCHIL	*Lindsaea orbiculata*	LoCHI1	yeast two-hybrid, Split luciferase	(Ni *et al*., 2020)
OsCHIL-1	*Oryza sativa*	OsCHS-1	yeast two-hybrid, BifC, Split luciferase	(Waki *et al*., 2020, Ban *et al*., 2018)

OsCHIL-2	*Oryza sativa*	OsCHS-1	yeast two-hybrid, Split luciferase	(Waki *et al*., 2020, Ban *et al*., 2018)
PpCHIL-A	*Physcomitrella patens*	PpCHS	Pulldown, yeast two-hybrid, Split luciferase	(Wolf-Saxon *et al*., 2023, Waki *et al*., 2020, Ban *et al*., 2018);
PpCHIL-B	*Physcomitrella patens*	PpCHS	yeast two-hybrid, Split luciferase	(Waki *et al*., 2020, Ban *et al*., 2018)
MpCHIL	*Marchantia polymorpha*	AtCHS	No interaction using yeast two-hybrid	(Waki *et al*., 2020)
PpCHIL-A	*Physcomitrella patens*	AtCHS	No interaction using yeast two-hybrid	(Waki *et al*., 2020)
SmoCHIL	*Selaginella moellendorffii*	SmoCHS	yeast two-hybrid; Split luciferase	(Waki *et al*., 2020, Ni *et al*., 2020)
ThCHIL-B	*Torenia fournieri*	ThCHS	yeast two-hybrid, BifC	(Waki *et al*., 2020)
VvCHIL	*Vitis vinifera*	AtCHS	Pulldown	(Wolf-Saxon *et al*., 2023)
VvCHIL	*Vitis vinifera*	VvCHS-1	yeast two-hybrid	(Waki *et al*., 2020)

* References appear in order of methods completed.

CHAPTER VIII: REFERENCES

Abe, I. and Morita, H. (2010) Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Natural Product Reports,* **27,** 809-838.

Adrain, C. and Freeman, M. (2012) New lives for old: evolution of pseudoenzyme function. *Nature Reviews Molecular Cell Biology,* **13,** 489-498.

Aharoni, A. and Galili, G. (2011) Metabolic engineering of the plant primary-secondary metabolism interface. *Current Opinion in Biotechnology,* **22,** 239-244.

Akashi, T., Aoki, T. and Ayabe, S. I. (2005) Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. *Plant Physiology,* **137,** 882-891.

Akiyama, T. (1999) *p*-Coumaroyltriacetic acid synthase, a new homologue of chalcone synthase, from *Hydrangea macrophylla* var. *thunbergii*. *European Journal of Biochemistry,* **263,** 834.

Austin, M. B. and Noel, J. P. (2003) The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports,* **20,** 79-110.

Ban, Z., Qin, H., Mitchell, A. J., Liu, B., Zhang, F., Weng, J. K.*, et al.* (2018) Noncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America,* **115,** E5223-E5232.

Bandara, B. R., Kumar, N. S. and Samaranayake, K. S. (1989) An antifungal constituent from the stem bark of *Butea monosperma*. *Journal of ethnopharmacology,* **25,** 73-75.

Bednar, R. A. and Hadcock, J. R. (1988) Purification and characterization of chalcone isomerase from soybeans. *The Journal of Biological Chemistry,* **263,** 9582-9588.

Blount, J. W., Korth, K. L., Masoud, S. A., Rasmussen, S., Lamb, C. and Dixon, R. A. (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. *Plant Physiology,* **122,** 107-116.

Board, I. (2023) Narcotic Drugs: Estimated World Requirements for 2022-Statistics for 2021. *United Nations Publication*.

Boland, M. J. and Wong, E. (1975) Purification and kinetic properties of chalcone-flavanone isomerase from Soya Bean. *European Journal of Biochemistry,* **50,** 383-389.

Bomati, E. K., Austin, M. B., Bowman, M. E., Dixon, R. A. and Noel, J. P. (2005) Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. *Journal of Biological Chemistry,* **280,** 30496-30503.

Booth, M. (2013) *Opium: A history*. St. Martin's Griffin.

Bradsher, K. (2014) The shake-up on opium island; Tasmania needs to adapt to remain the world's top source of a key painkiller. *International New York Times*.

Brickman, A. M., Yeung, L.-K., Alschuler, D. M., Ottaviani, J. I., Kuhnle, G. G., Sloan, R. P.*, et al.* (2023) Dietary flavanols restore hippocampal-dependent memory in older adults with lower diet quality and lower habitual flavanol consumption. *Proceedings of the National Academy of Sciences,* **120,** e2216932120.

Burbulis, I. E., Iacobucci, M. and Shirley, B. W. (1996) A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in *Arabidopsis*. *The Plant Cell,* **8,** 1013-1025.

Burbulis, I. E. and Winkel-Shirley, B. (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proceedings of the National Academy of Sciences,* **96,** 12929-12934.

Carrington, Y., Guo, J., Le, C. H., Fillo, A., Kwon, J., Tran, L. T.*, et al.* (2018) Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. *The Plant Journal,* **95,** 823-833.

Chemler, J. A. and Koffas, M. A. G. (2008) Metabolic engineering for plant natural product biosynthesis in microbes. *Current Opinion in Biotechnology,* **19,** 597-605.

Chemler, J. A., Lim, C. G., Daiss, J. L. and Koffas, M. A. G. (2010) A versatile microbial system for biosynthesis of novel polyphenols with altered estrogen receptor binding activity. *Chemistry and Biology,* **17,** 392-401.

Chen, A. Y., Schnarr, N. A., Kim, C.-Y., Cane, D. E. and Khosla, C. (2006) Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society,* **128,** 3067-3074.

Chen, H., Seguin, P., Archambault, A., Constan, L. and Jabaji, S. (2009) Gene expression and isoflavone concentrations in soybean sprouts treated with chitosan. *Crop Science,* **49,** 224-236.

Chen, Y., Wang, Y., Liang, C., Liu, L., Song, X., Zhao, Y.*, et al.* (2022) Characterization of the key bibenzyl synthase in *Dendrobium sinense*. *International Journal of Molecular Sciences,* **23,** 6780.

Cheng, A. X., Zhang, X., Han, X. J., Zhang, Y. Y., Gao, S., Liu, C. J.*, et al.* (2018) Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. *New Phytologist,* **217,** 909-924.

Chennupati, P., Seguin, P., Chamoun, R. and Jabaji, S. (2012) Effects of high-temperature stress on soybean isoflavone concentration and expression of key genes involved in isoflavone synthesis. *Journal of Agricultural and Food Chemistry,* **60,** 12421-12427.

Choi, G.-S., Choo, H. J., Kim, B.-G. and Ahn, J.-H. (2020) Synthesis of acridone derivatives via heterologous expression of a plant type III polyketide synthase in *Escherichia coli*. *Microbial Cell Factories,* **19**.

Dastmalchi, M. (2021) Elusive partners: a review of the auxiliary proteins guiding metabolic flux in flavonoid biosynthesis. *The Plant Journal,* **108,** 314-329.

Dastmalchi, M., Bernards, M. A. and Dhaubhadel, S. (2016) Twin anchors of the soybean isoflavonoid metabolon: Evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. *The Plant Journal,* **85,** 689-706.

Dastmalchi, M., Chang, L., Torres, M. A., Ng, K. K. S. and Facchini, P. J. (2018) Codeinone reductase isoforms with differential stability, efficiency and product selectivity in opium poppy. *The Plant Journal,* **95,** 631-647.

Dastmalchi, M., Chen, X., Hagel, J. M., Chang, L., Chen, R., Ramasamy, S.*, et al.* (2019) Neopinone isomerase is involved in codeine and morphine biosynthesis in opium poppy. *Nature Chemical Biology,* **15,** 384-390.

Dastmalchi, M. and Dhaubhadel, S. (2015) Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta,* **241,** 507-523.

Davies, K. M., Jibran, R., Zhou, Y., Albert, N. W., Brummell, D. A., Jordan, B. R.*, et al.* (2020) The evolution of flavonoid biosynthesis: a bryophyte perspective. *Frontiers in Plant Science,* **11,** 7.

Dixon, R. A. (1999a) Isoflavonoids: biochemistry, molecular biology and biological functions. *Comprehensive natural products chemistry,* **1,** 773-823.

Dixon, R. A. (1999b) Plant natural products: the molecular genetic basis of biosynthetic diversity. *Current Opinion in Biotechnology,* **10,** 192-197.

Dixon, R. A., Achnine, L., Kota, P., Liu, C. J., Reddy, M. S. S. and Wang, L. (2002) The phenylpropanoid pathway and plant defence - A genomics perspective. *Molecular Plant Pathology,* **3,** 371-390.

Dixon, R. A. and Paiva, N. L. (1995) Stress-induced phenylpropanoid metabolism. *Plant Cell,* **7,** 1085-1097.

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

Dixon, R. A. and Whitehead, I. M. (1982) Purification and properties of chalcone isomerase from cell suspension cultures of *Phaseolus vulgaris*. *Biochemica et Biophysica Acta,* **715,** 25-33.

Dostálek, T., Rokaya, M. B., Maršík, P., Rezek, J., Skuhrovec, J., Pavela, R.*, et al.* (2016) Trade-off among different anti-herbivore defence strategies along an altitudinal gradient. *AoB Plants,* **8,** 026.

Ehlting, J., Shin, J. J. K. and Douglas, C. J. (2001) Identification of 4-coumarate:coenzyme-A ligase (4CL) substrate recognition domains. *The Plant Journal,* **27,** 455-465.

Fani, R. (2009) Origin and evolution of metabolic pathways. *Physics of Life Reviews,* **6,** 23.

Felton, G. W., Korth, K. L., Bi, J. L., Wesley, S. V., Huhman, D. V., Mathews, M. C.*, et al.* (1999) Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. *Current Biology,* **9,** 317-320.

Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R. and Thomson, J. D. (2004) Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics,* **35,** 375-403.

Ferrer, J. L., Austin, M. B., Stewart, C. and Noel, J. P. (2008) Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiology and Biochemistry,* **46,** 356-370.

Ferrer, J. L., Jez, J. M., Bowman, M. E., Dixon, R. A. and Noel, J. P. (1999) Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology,* **6,** 775-784.

Fliegmann, J., Furtwängler, K., Malterer, G., Cantarello, C., Schüler, G., Ebel, J.*, et al.* (2010) Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). *Phytochemistry,* **71,** 508-514.

Flores-Sanchez, I. J. and Verpoorte, R. (2008) PKS activities and biosynthesis of cannabinoids and flavonoids in *Cannabis sativa* L. Plants. *Plant and Cell Physiology,* **49,** 1767.

Fontaine, T. D. and Irving Jr, G. (1948) Isolation and partial characterization of crystalline tomatine, an antibiotic agent from the tomato plant. *Archives of Biochemistry,* **18,** 467-475.

Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. and Smolke, C. D. (2015) Complete biosynthesis of opioids in yeast. *Science,* **349,** 1095-1100.

Gao, D., Liu, T., Gao, J., Xu, J., Gou, Y., Pan, Y.*, et al.* (2022) *De novo* biosynthesis of vindoline and catharanthine in *Saccharomyces cerevisiae*. *BioDesign Research,* **2022,** 0002.

Garagounis, C., Delkis, N. and Papadopoulou, K. K. (2021) Unraveling the roles of plant specialized metabolites: using synthetic biology to design molecular biosensors. *New Phytologist,* **231,** 1338-1352.

García-Lafuente, A., Guillamón, E., Villares, A., Rostagno, M. A. and Martínez, J. A. (2009) Flavonoids as anti-inflammatory agents: implications in cancer and cardiovascular disease. *Inflammation research,* **58,** 537-552.

Gensheimer, M. (2004) Chalcone isomerase family and fold: No longer unique to plants. *Protein Science,* **13,** 540-544.

Georgaki, S., Skopeliti, M., Tsiatas, M., Nicolaou, K. A., Ioannou, K., Husband, A.*, et al.* (2009) Phenoxodiol, an anticancer isoflavene, induces immunomodulatory effects *in vitro* and *in vivo*. *Journal of Cellular and Molecular Medicine,* **13,** 3929-3938.

Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., E Baart, G. J.*, et al.* (2014) Detection of flavonoids in microalgae from different evolutionary lineages. *Journal of phycology,* **50,** 483-492.

Graham, T. L., Graham, M. Y., Subramanian, S. and Yu, O. (2007) RNAi silencing of genes for elicitation or biosynthesis of 5-deoxyisoflavonoids suppresses race-specific resistance and hypersensitive cell death in *Phytophthora sojae* infected tissues. *Plant Physiology,* **144,** 728-740.

Harwood, J. L. (1996) Recent advances in the biosynthesis of plant fatty acids. *Biochemica et Biophysica Acta* **1301,** 7-56.

Hinderer, W. (1985) Chalcone Synthase from cell suspension cultures of *Daucus carota* L. In: *Archives of Biochemistry and Biophysics.* pp. 265-272.

Houseman, J. G., Campos, F., Thie, N. M. R., Philogene, B. J. R., Atkinson, J., Morand, P.*, et al.* (1992) Effect of the maize-derived compounds DIMBOA and MBOA on growth and digestive processes of European Corn-Borer (*Lepidoptera*: *Pyralidae*). *Journal of Economic Entomology,* **85,** 669-674.

Hrazdina, G. and Wagner, G. J. (1985) Metabolic pathways as enzyme complexes: Evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Archives of Biochemistry and Biophysics,* **237,** 88-100.

Ibdah, M. (2017) Biosynthetic pathway and metabolic engineering of plant dihydrochalcones. *Journal of Agricultural and Food Chemistry,* **66,** 2273.

Inc, S. I. (2021) Using JMP® 16. Cary, NC: SAS Institute Inc.

Jeandet, P., Delaunois, B., Conreux, A., Donnez, D., Nuzzo, V., Cordelier, S.*, et al.* (2010) Biosynthesis, metabolism, molecular engineering, and biological functions of stilbene phytoalexins in plants. *BioFactors,* **36,** 331-341.

Jez, J. M., Austin, M. B., Ferrer, J. L., Bowman, M. E., Schröder, J. and Noel, J. P. (2000a) Structural control of polyketide formation in plant-specific polyketide synthases. *Chemistry and Biology,* **7,** 919-930.

Jez, J. M., Bowman, M. E., Dixon, R. A. and Noel, J. P. (2000b) Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nature Structural Biology,* **7,** 786-791.

Jez, J. M., Bowman, M. E. and Noel, J. P. (2002) Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. *Biochemistry,* **41,** 5168-5176.

Jez, J. M. and Noel, J. P. (2002) Reaction mechanism of chalcone isomerase: pH dependence, diffusion control, and product binding differences. *Journal of Biological Chemistry,* **277,** 1361-1369.

Jiang, C., Wechuck, J. B., Goins, W. F., Krisky, D. M., Wolfe, D., Ataai, M. M.*, et al.* (2004) Immobilized cobalt affinity chromatography provides a novel, efficient method for Herpes Simplex Virus Type 1 gene vector purification. *Journal of Virology,* **78,** 8994-9006.

Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R. A.*, et al.* (2015) Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *Journal of Experimental Botany,* **66,** 7165-7179.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E.*, et al.* (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature,* **473,** 97-100.

Jiménez-González, L., Álvarez-Corral, M., Muñoz-Dorado, M. and Rodríguez-García, I. (2008) Pterocarpans: Interesting natural products with antifungal activity and other biological properties. *Phytochemistry Reviews,* **7,** 125-154.

Jones, D. T., Taylor, W. R. and Thornton, J. M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics,* **8,** 275.

Jung, W., Yu, O., Lau, S.-M. C., O'Keefe, D. P., Odell, J., Fader, G.*, et al.* (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. *Nature Biotechnology,* **18,** 208-212.

Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A.*, et al.* (2018) Evolution of chalcone isomerase from a noncatalytic ancestor article. *Nature Chemical Biology,* **14,** 548-555.

Keasling, J. D. (2010) Manufacturing molecules through metabolic engineering. *Science,* **330,** 1355-1358.

Kodama, O., Miyakawa, J., Akatsuka, T. and Kiyosawa, S. (1992) Sakuranetin, a flavanone phytoalexin from ultraviolet-irradiated rice leaves. *Phytochemistry,* **31,** 3807-3809.

Koirala, N., Thuan, N. H., Ghimire, G. P., Van Thang, D. and Sohng, J. K. (2016) Methylation of flavonoids: Chemical structures, bioactivities, progress and perspectives for biotechnological production. *Enzyme and microbial technology,* **86,** 103-116.

Koukol, J. and Conn, E. E. (1961) The metabolism of aromatic compounds in higher plants. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare*. *Journal of Biological Chemistry,* **236,** 2692-2698.

Kuiper, G. G., Lemmen, J. G., Carlsson, B., Corton, J. C., Safe, S. H., Van Der Saag, P. T.*, et al.* (1998) Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor *β*. *Endocrinology,* **139,** 4252-4263.

Kumar Sudhir, S. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution,* **35,** 1547-1549.

Kumaran, A. V. A., Krysiak, K., Tian, G. and Dhaubhadel, S. (2018) Genome-wide identification and localization of chalcone synthase family in soybean (*Glycine max* L. Merr). *BMC Plant Biology,* **18**.

Landmann, C., Hücherig, S., Fink, B., Hoffmann, T., Dittlein, D., Coiner, H. A.*, et al.* (2011) Substrate promiscuity of a rosmarinic acid synthase from lavender (*Lavandula angustifolia* L.). *Planta,* **234,** 305-320.

Lapčík, O. (2007) Isoflavonoids in non-leguminous taxa: A rarity or a rule? *Phytochemistry,* **68,** 2909-2916.

Le Bail, J. C., Varnat, F., Nicolas, J. C. and Habrioux, G. (1998) Estrogenic and antiproliferative activities on MCF-7 human breast cancer cells by flavonoids. *Cancer letters,* **130,** 209-216.

Lee, I. K. and Yun, B. S. (2011) Styrylpyrone-class compounds from medicinal fungi *Phellinus* and *Inonotus spp.*, and their medicinal importance. *J Antibiot,* **64,** 349-359.

Leveson-Gower, R. B., Mayer, C. and Roelfes, G. (2019) The importance of catalytic promiscuity for enzyme design and evolution. *Nature Reviews Chemistry,* **3,** 687-705.

Liu, Q., Liu, Y., Li, G., Savolainen, O., Chen, Y. and Nielsen, J. (2021) *De novo* biosynthesis of bioactive isoflavonoids by engineered yeast cell factories. *Nature Communications,* **12,** 6085.

Liu, Q., Luo, L. and Zheng, L. (2018a) Lignins: biosynthesis and biological functions in plants. *International Journal of Molecular Sciences,* **19,** 335.

Liu, X., Li, L. and Zhao, G.-R. (2022) Systems metabolic engineering of *Escherichia coli* coculture for *de novo* production of genistein. *ACS Synthetic Biology,* **11,** 1746-1757.

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F., Visser, R. G. and Bovy, A. (2018b) Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. *Frontiers in chemistry,* **6,** 52.

Lv, X., Xue, H., Qin, L. and Li, C. (2022) Transporter engineering in microbial cell factory boosts biomanufacturing capacity. *BioDesign Research,* **2022**.

Lygin, A. V., Hill, C. B., Zernova, O. V., Crull, L., Widholm, J. M., Hartman, G. L.*, et al.* (2010) Response of soybean pathogens to glyceollin. *Phytopathology,* **100,** 897-903.

Maeda, H. A. (2019) Harnessing evolutionary diversification of primary metabolism for plant synthetic biology. *Journal of Biological Chemistry,* **294,** 16549-16566.

Malla, A., Shanmugaraj, B., Srinivasan, B., Sharma, A. and Ramalingam, S. (2020) Metabolic engineering of isoflavonoid biosynthesis by expressing *Glycine max* isoflavone synthase in *Allium cepa* L. for genistein production. *Plants,* **10,** 52.

Mameda, R., Waki, T., Kawai, Y., Takahashi, S. and Nakayama, T. (2018) Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. *Plant Journal,* **96,** 56-74.

Mansfield, J. W. (2000) The Role of Phytoalexins and Phytoanticipins. *Mechanisms of Resistance to Plant Diseases,* **Antimicrobial Compounds and Resistance,** 325-370.

Martens, S., Forkmann, G., Matern, U. and Lukačin, R. (2001) Cloning of parsley flavone synthase I. *Phytochemistry,* **58,** 43-46.

Martens, S. and Mithöfer, A. (2005) Flavones and flavone synthases. *Phytochemistry,* **66,** 2399-2407.

Matsuno, M., Compagnon, V., Schoch, G. A., Schmitt, M., Debayle, D., Bassard, J.-E*., et al.* (2009) Evolution of a novel phenolic pathway for pollen development. *Science,* **325,** 1688-1692.

Mehlhorn, D. G., Wallmeroth, N., Berendzen, K. W. and Grefen, C. (2018) 2in1 vectors improve *in planta* BiFC and FRET analyses. *The Plant Endoplasmic Reticulum* **1691,** 139-158.

Meng, Y., Liu, X., Zhang, L. and Zhao, G.-R. (2022) Modular engineering of *Saccharomyces cerevisiae* for *de novo* biosynthesis of genistein. *Microorganisms,* **10,** 1402.

Mindnich, R. D. and Penning, T. M. (2009) Aldo-keto reductase (AKR) superfamily: Genomics and annotation. *Human Genomics,* **3,** 362-370.

Moghe, G. D. (2015) Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiology,* **169,** 1512.

Mondal, S. K. and Roy, S. (2018) Genome-wide sequential, evolutionary, organizational and expression analyses of phenylpropanoid biosynthesis associated MYB domain transcription factors in *Arabidopsis*. *Journal of Biomolecular Structure and Dynamics,* **36,** 1577-1601.

Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E*., et al.* (2014) A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant Journal,* **78,** 294-304.

Mouradov, A. and Spangenberg, G. (2014) Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Frontiers in Plant Science,* **5,** 620.

Muñoz, C., Hoffmann, T., Escobar, N. M., Ludemann, F., Botella, M. A., Valpuesta, V*., et al.* (2010) The strawberry fruit Fra a allergen functions in flavonoid biosynthesis. *Molecular Plant,* **3,** 113-124.

Nabavi, S. M., Šamec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S*., et al.* (2020) Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. *Biotechnology Advances,* **38,** 107316.

Nakayama, T., Takahashi, S. and Waki, T. (2019) Formation of flavonoid metabolons: functional significance of protein-protein interactions and impact on flavonoid chemodiversity. *Frontiers in Plant Science,* **10,** 821.

Nam, H., Lewis, N. E., Lerman, J. A., Lee, D.-H., Chang, R. L., Kim, D.*, et al.* (2012) Network context and selection in the evolution to enzyme specificity. *Science,* **337,** 1101-1104.

Neve, E. P. A. and Ingelman-Sundberg, M. (2008) Intracellular transport and localization of microsomal cytochrome P450. *Analytical and Bioanalytical Chemistry,* **392,** 1075-1084.

Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E.*, et al.* (2012) Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature,* **485,** 530-533.

Ni, R., Zhu, T. T., Zhang, X. S., Wang, P. Y., Sun, C. J., Qiao, Y. N.*, et al.* (2020) Identification and evolutionary analysis of chalcone isomerase-fold proteins in ferns. *Journal of Experimental Botany,* **71,** 290-304.

O'Connor, S. E. (2012) Strategies for engineering plant natural products. *Methods in Enzymology,* **515,** 189-206.

O'Connor, S. E. and Maresh, J. J. (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Natural product reports,* **23,** 532-547.

Padmakshan, D., Timokhin, V. I., Lu, F., Schatz, P. F., Vanholme, R., Boerjan, W.*, et al.* (2022) Synthesis of hydroxycinnamoyl shikimates and their role in monolignol biosynthesis. *Holzforschung,* **76,** 133-144.

Panchy, N., Lehti-Shiu, M. and Shiu, S. H. (2016) Evolution of gene duplication in plants. *Plant Physiology,* **171,** 2294.

Petersen, M., Abdullah, Y., Benner, J., Eberle, D., Gehlen, K., Hücherig, S.*, et al.* (2009) Evolution of rosmarinic acid biosynthesis. *Phytochemistry,* **70,** 1663-1679.

Pratyusha, D. S. and Sarada, D. V. L. (2022) MYB transcription factors—master regulators of phenylpropanoid biosynthesis and diverse developmental and stress responses. *Plant Cell Reports,* **41,** 2245-2260.

Rakwal, R., Agrawal, G. K., Yonekura, M. and Kodama, O. (2000) Naringenin 7-O-methyltransferase involved in the biosynthesis of the flavanone phytoalexin sakuranetin from rice (*Oryza sativa* L.). *Plant Science,* **155,** 213-221.

Ralph, J. (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids. *Phytochemistry Reviews,* **3,** 29.

Ralston, L. and Yu, O. (2006) Metabolons involving plant cytochrome P450s. *Phytochemistry Reviews,* **5,** 459-472.

Sajid, M., Stone, S. R. and Kaur, P. (2021) Recent advances in heterologous synthesis paving way for future green-modular bioindustries: a review with special reference to isoflavonoids. *Frontiers in Bioengineering and Biotechnology,* **9,** 1-18.

Sánchez-Baracaldo, P. and Cardona, T. (2020) On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytologist,* **225,** 1440-1446.

Sarian, M. N., Ahmed, Q. U., Mat So'Ad, S. Z., Alhassan, A. M., Murugesu, S., Perumal, V.*, et al.* (2017) Antioxidant and antidiabetic effects of flavonoids: a structure-activity relationship based study. *BioMed Research International,* **2017,** 1-14.

Schmid, K. M. and Ohlrogge, J. B. (2002) Lipid metabolism in plants. *New Comprehensive Biochemistry***,** 93.

Schopfer, C. R., Kochs, G., Lottspeich, F. and Ebel, J. (1998) Molecular characterization and functional expression of dihydroxypterocarpan 6a-hydroxylase, an enzyme specific for pterocarpanoid phytoalexin biosynthesis in soybean (*Glycine max* L.). *FEBS Letters,* **432,** 182-186.

Schröder, G., Wehinger, E. and Schröder, J. (2002) Predicting the substrates of cloned plant *O*-methyltransferases. *Phytochemistry,* **59,** 1-8.

Schröder, J. (1999) The chalcone/stilbene synthase-type family of condensing enzymes. *Comprehensive Natural Products Chemistry***,** 749.

Scossa, F. and Fernie, A. R. (2020) The evolution of metabolism: How to test evolutionary hypotheses at the genomic level. *Computational and Structural Biotechnology Journal,* **18,** 482.

Sepiol, C. J., Yu, J. and Dhaubhadel, S. (2017) Genome-Wide Identification of Chalcone Reductase Gene Family in Soybean: Insight into Root-Specific GmCHRs and Phytophthora sojae Resistance. *Frontiers in Plant Science,* **8**.

Sharma, A., Shahzad, B., Rehman, A., Bhardwaj, R., Landi, M. and Zheng, B. (2019) Response of phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules,* **24,** 2452.

Shirley, B. W., Hanley, S. and Goodman, H. M. (1992) Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis *transparent testa* mutations. *Plant Cell,* **4,** 333-347.

Souto, A. L., Sylvestre, M., Tölke, E. D., Tavares, J. F., Barbosa-Filho, J. M. and Cebrián-Torrejón, G. (2021) Plant-derived pesticides as an alternative to pest management and sustainable agricultural production: prospects, applications and challenges. *Molecules,* **26,** 4835.

Srere, P. A. (1985) The metabolon. *Trends in Biochemical Sciences,* **10,** 109.

St-Pierre, B. and De Luca, V. (2000) Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent advances in phytochemistry,* **34,** 285-315.

Steele, C. L., Gijzen, M., Qutob, D. and Dixon, R. A. (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. *Archives of Biochemistry and Biophysics,* **367,** 146-150.

Stevens, J. F. and Page, J. E. (2004) Xanthohumol and related prenylflavonoids from hops and beer: to your good health! *Phytochemistry,* **65,** 1317.

Stevenson, P., Turner, H. and Haware, M. (1997) Phytoalexin accumulation in the roots of chickpea (*Cicer arietinum* L.) seedlings associated with resistance to fusarium wilt (*Fusarium oxysporumf sp. ciceri*). *Physiological and Molecular Plant Pathology,* **50,** 167-178.

Stovicek, V., Holkenbrink, C. and Borodina, I. (2017) CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Research,* **17,** 30-30.

Stracke, R., Werber, M. and Weisshaar, B. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Current opinion in plant biology,* **4,** 447-456.

Subramanian, S., Stacey, G. and Yu, O. (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *The Plant Journal,* **48,** 261-273.

Symons, G., Hudson, C. and Livermore, M. (2009) *Papaver somniferum* with high concentration of codeine. WO Patent 2009143574A1.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., *et al.* (2008) Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal,* **54,** 335-347.

Tanaka, Y. (2008) Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. *Current Opinion in Biotechnology,* **19,** 190.

Thodey, K., Galanie, S. and Smolke, C. D. (2014) A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nature Chemical Biology,* **10,** 837-844.

Tohge, T. (2013) On the regulation and function of secondary metabolism during fruit development and ripening. *Journal of Experimental Botany,* **65,** 4599.

Tropf, S., Lanz, T., Rensing, S. A., Schroder, J. and Schroder, G. (1994) Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *Journal of Molecular Evolution,* **38**.

Utomo, J. C., Chaves, F. C., Bauchart, P., Martin, V. J. J. and Ro, D. K. (2021a) Developing a Yeast Platform Strain for an Enhanced Taxadiene Biosynthesis by CRISPR/Cas9. *Metabolites,* **11,** 1-15.

Utomo, J. C., Hodgins, C. L. and Ro, D. K. (2021b) Multiplex genome editing in yeast by CRISPR/Cas9 – A potent and agile tool to reconstruct complex metabolic pathways. *Frontiers in Plant Science,* **12,** 1639-1639.

Van de Sande, K. and Bisseling, T. (1997) Signalling in symbiotic root nodule formation. *Essays in biochemistry,* **32,** 127-142.

Veitch, N. C. (2007) Isoflavonoids of the leguminosae. *Natural Product Reports,* **24,** 417-464.

Ververidis, F., Trantas, E., Douglas, C. J., Vollmer, G., Kretzschmar, G. and Panopoulos, N. (2007) Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part I: chemical diversity, impacts on plant biology and human health. *Biotechnology journal,* **2,** 1214-1234.

Wais, R. J., Keating, D. H. and Long, S. R. (2002) Structure-function analysis of Nod factor-induced root hair calcium spiking in rhizobium-legume symbiosis. *Plant Physiology,* **129,** 211.

Waki, T., Mameda, R., Nakano, T., Yamada, S., Terashita, M., Ito, K., *et al.* (2020) A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nature Communications,* **11,** 1-14.

Waki, T., Takahashi, S. and Nakayama, T. (2021) Managing enzyme promiscuity in plant specialized metabolism: A lesson from flavonoid biosynthesis. *BioEssays,* **43,** 2000164.

Waki, T., Yoo, D. C., Fujino, N., Mameda, R., Denessiouk, K., Yamashita, S., *et al.* (2016) Identification of protein-protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). *Biochemical and Biophysical Research Communications,* **469,** 546-551.

Wang, J.-F., Liu, S.-S., Song, Z.-Q., Xu, T.-C., Liu, C.-S., Hou, Y.-G., *et al.* (2020) Naturally occurring flavonoids and isoflavonoids and their microbial transformation: a review. *Molecules 2020, Vol. 25, Page 5112,* **25,** 5112-5112.

Wang, S., Alseekh, S., Fernie, A. R. and Luo, J. (2019) The structure and function of major plant metabolite modifications. *Molecular Plant,* **12,** 899-919.

Welle, R. and Grisebach, H. (1987) Purification and properties of chalcone synthase from cell suspension cultures of soybean. In: *Z. Naturforsch.* pp. 1200-1206.

Welle, R. and Grisebach, H. (1988) Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6′-deoxychalcone. *FEBS Letters,* **236,** 221-225.

Weng, J. K. (2012) The rise of chemodiversity in plants. *Science,* **336,** 1667.

Weng, J. K. and Noel, J. P. (2012) The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harbor Symposia on Quantitative Biology,* **77,** 309-320.

Wolf-Saxon, E. R., Moorman, C. C., Castro, A., Ruiz, A., Mallari, J. P. and Burke, J. R. (2023) Regulatory ligand binding in plant chalcone isomerase-like (CHIL) proteins. *Journal of Biological Chemistry***,** 104804.

Wouters, F. C., Blanchette, B., Gershenzon, J. and Vassão, D. G. (2016) Plant defense and herbivore counter-defense: benzoxazinoids and insect herbivores. *Phytochemistry Reviews,* **15,** 1127-1151.

Wu, Y. (2022) *De novo* biosynthesis of diverse plant-derived styrylpyrones in *Saccharomyces cerevisiae*. *Metabolic Engineering Communications,* **14,** 00195.

Yan, H., Pei, X., Zhang, H., Li, X., Zhang, X., Zhao, M.*, et al.* (2021) MYB-mediated regulation of anthocyanin biosynthesis. *International Journal of Molecular Sciences,* **22,** 3103.

Zanoli, P. and Zavatti, M. (2008) Pharmacognostic and pharmacological profile of *Humulus lupulus* L. *Journal of Ethnopharmacology,* **116,** 383.

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrübbers, L.*, et al.* (2022) A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature,* **609,** 341-347.