Construction of Microbial Expression Systems for the Investigation of CsCHI-L Function in the Cannabinoid Biosynthetic Pathway

by

Lindsay Young

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfilment of the requirements

of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

Copyright © 2023 by Lindsay Elizabeth Young

Abstract

Phytocannabinoids, a group of specialized metabolites that accumulate in the glandular trichomes of *Cannabis sativa* L., are compounds of high economic value recognized for both their recreational and medicinal purposes. Cannabinoids exert their effects in the body by binding to endocannabinoid receptors localized to the central and peripheral nervous system, and components of the immune system. Modulation of these systems by cannabinoids has shown efficacy in a myriad of therapies, including treatment of pain management, epilepsy, asthma, inflammation and neurodegenerative disease. Cannabinoids are produced primarily in the flowering trichomes of cannabis plants through the enzymatic convergence of a resorcinol core (featuring an alkyl side chain) and a monoterpene moiety derived from geranylpyrophosphate (GPP). The biosynthetic pathways responsible for the production of both of these compounds have largely been determined in the past few decades, allowing for partial reconstruction in heterologous system. While these systems show promise for the sourcing of cannabinoids, improvements are necessary for economical production on a larger scale. Low yields of cannabinoids in these systems suggest that there may be unidentified accessory proteins that assist in the biosynthesis of these compounds. Our current understanding suggests that cannabinoid biosynthesis is primarily limited by the production of olivetolic acid (OA) from malonyl-CoA and hexanoyl-CoA. This research project aimed to create synthetic biology system(s) that incorporate the enzymatic components for the OA-forming steps of the cannabinoid biosynthetic pathway into microbial hosts, *S. cerevisiae* and *E. coli*. to facilitate the investigation of a potential chalcone isomerase-like (CHI-L) accessory protein in said pathway. Accessory proteins such as CHI-L's are known to interact with synthases and to be necessary for the processing of the polyketide intermediate, and observed function of similar CHI-L proteins in other embryophytes suggest a mechanism of action for this proposed CsCHI-L in cannabis.

While the experimentation was ultimately successful in creating synthetic biology systems capable of incorporating relevant enzymes and producing the intermediate compound OA, there was no observable difference when comparing compound production in systems with the proposed CsCHI-L protein and those without. This data is indicative that the CsCHI-L protein does not function in these early steps of the pathway.

List of Terms

ACN: Acetonitrile

ACP: Acyl-carrier protein

AMP: Adenosine monophosphate

AMP: Ampicillin

ATP: Adenosine triphosphate

BjMCS: Bradyrhizobium japonicum malonyl-CoA synthetase

BL21: *E. coli* B strain

cAMP: cyclic adenosine pyrophosphate

CB: Cannabinoid

CB1: Cannabinoid receptor 1

CB2: Cannabinoid receptor 2

CBC: Cannabichromene

CBD: Cannabidiol

CBDA: Cannabidiolic acid

CBDP: Cannabidiphorol

CBDV: Cannabidivarin

CBG: Cannabigerol

CBGA: Cannabigerolic acid

CBL: Cannabicyclol

CBN: Cannabinol

CBR: Cannabinoid receptor

CHI-L: Chalcone isomerase-like

CHI: Chalcone isomerase

CHS: Chalcone synthase

CloL: cloacin

Col: Colicin

COPD: Chronic obstructive pulmonary disease

CP: Cyclohexylphenols

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

CsAAE1: Cannabis sativa acyl-activating enzyme

CsOAC: Cannabis sativa olivetolic acid cyclase

CsPT1: Cannabis sativa prenyltransferase 1

CsPT4: Cannabis sativa prenyltransferase 4

CsTKS: Cannabis sativa triketide synthase

CTAL: p-coumaroyltriacetic acid lactone

DABB: Dimeric α + β barrel

DMAPP: Dimethylallyl diphosphate

DMX: Demethylxanthohumol

DNA: Deoxyribonucleic acid

ECS: Endocannabinoid system

(delta) EMV: Multiplier voltage

ESI: Electrospray ionization

EST: Expressed sequence tag

EtAc: Ethyl acetate

G418: Geneticin

GAL: Galactose

GOI.: Gene of interest

GPCR: G-protein coupled receptor

GPP: Geranylpyrophosphate

HCl: Hydrochloric acid

HDR: Homology-directed repair

HIS: Histidine

HPLC-UC: High-performance liquid chromatography-Ultraviolet

HTAL: Hexanoyl triacetic acid lactone

HXT: High-affinity hexose transporter

INVSc1: Diploid yeast (S. cerevisiae) strain

IPTG: Isopropyl β-D-1-thiogalactopyranoside

KAN: Kanamycin

kDa: Kilo daltons

KR: B-ketoacyl-ACP-reductase

kV: Kilo volts

LacI: Lac repressor

LC-MS: Liquid chromatography mass spectrometry

LOD: Limit of detection

LOQ: Limit of quantification

MAGE: Multiplex Automated Genome Engineering

MCS: Multiple cloning site

MD: Molecular dynamics

MeOH: Methanol

MEP: Methylerythritol phosphate

mRNA: Messenger RNA

MRM: Multiple reaction monitoring

MS/MS: Tandem mass spectrometry

NHEJ: Non-homologous end joining

NMR: Nuclear magnetic resonance

OA: Olivetolic acid

OD600: Optical density at 600 nm wavelength

OL: Olivetol

Opt: Optimized

ORF: Open reading frame

Ori: origin of replication

PCDF-Duet1:

pCEV-G4-Km:

pCOLA-Duet1:

PCR: Polymerase chain reaction

PDAL: pentyl diacetic acid lactone

pESC-HIS:

pJET1.2:

PKS: Polyketide synthase

pLATE52:

PNP: Plant natural products

PP1: Pyrophosphate

Ppb: parts per billion

Ppm: parts per million

RLR: retron library recombineering

RNA: Ribonucleic acid

ROP: Repressor of primer

SBS: Synthetic biology systems

SD: Selective dropout

SDS-PAGE: Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SEER: serial enrichment for efficient recombineering

S/N: Signal to noise ratio

SPEC: Spectinomyciomycin

ssDNA: Single-stranded DNA

TALEN: Transcription activator-like effector nucleases

TEF1: Translational elongation factor 1α

THC: Δ9-Tetrahydrocannabinol

THCA: Δ9-Tetrahydrocannabinolic acid

THCV: tetrahydrocannabivarin

TRIS: trishydroxymethylaminomethane

tRNA: transfer RNA

UV-VIS: Ultraviolet visible

v/v: volume/volume

w/v: weight/volume

XN: xanthohumol

YPD: Yeast extract peptone dextrose

YPG: Yeast extract peptone-galactose

ZFN: Zinc-finger nuclease

List of Figures

Figure 1. Reaction schemes of different types of polyketide synthases. 5

Figure 2. Compound production through chalcone synthase and chalcone isomerase enzymes.... 7

Figure 3. Proposed reaction mechanism for the cyclization of chalcone to flavanone through chalcone isomerase (CHI) activity. 9

Figure 4. Cannabinoid biosynthetic pathway 17

Figure 5. Selective dropout media and antibiotic selection 47

Figure 6. Chromatograms and UV-VIS spectra from *in vitro* assays 50

Figure 7. Compound production in in vitro assays 53

Figure 8. Calibration curves of authentic analytical standards 54

Figure 9. E. coli SBS plasmid expression cassettes. 56

Figure 10. S. cerevisiae SBS plasmid expression cassettes 57

Figure 11. Chromatograms and UV-VIS to confirm presence of HTAL and PDAL from *E. coli* SBS feeding experiments 80

Figure 12. HPLC chromatograms from in vivo E. coli SBS experiments 81

Figure 13. Compound production from in vivo E. coli SBS experiments. 84

Figure 14. Presence of Olivetolic Acid in spiked experimental samples. 85

Figure 15. Response of LC-MS/MS to in vivo S. cerevisiae SBS culture extractions 87

Figure 16. Compound production from in vivo S. cerevisiae SBS experiments. 89

Figure 17. Comparison of Olivetolic acid production in SBS systems 104

Figure 18. Phylogenetic analysis of CHI and CHI-L proteins. 108

Figure 19. Protein modelling of CsCHI-L 110

List of Tables

Table 1. Primers used in construction of synthetic biology systems 31

Table 2. Plasmids used in synthetic biology systems 41

Table 3. Limit of detection and quantification for LC-MS/MS. 90

Table 4. LC-MS/MS analysis of HTAL, PDAL, OA and Cinnamic Acid standards 91

Table of Contents

Abstract ..... **iii**

List of Tables ..... **ix**

Chapter 1: Introduction ***1***

Thesis Literature Review **1**

Plant specialized metabolites 1

Polyketide synthases and chalcone synthases 3

CHI and CHI-L accessory proteins 8

Cannabis and specialized metabolites phytocannabinoids 10

In planta production of phytocannabinoinds 15

Alternative production of cannabinoids 20

Synthetic biology systems to produce cannabinoids 22

Hypothesis 24

Chapter 2: Building Synthetic Biology Expression Systems ***25***

Introduction **25**

Materials and Methods **28**

Experimental overview (designing and constructing the synthetic biology systems) 28

Cannabis candidate sequence identification 28

Determination of vector cassettes and in silico construction 29

PCR amplification of acquired gene sequences 30

Initial cloning of sequences into an entry vector 32

Plasmid insert and verification 33
Attempted cloning of CsPT1 and CsPT4 35
Insertion of sequences of interest into HIS-tag protein expression plasmids 35
Protein expression 36
In vitro assays 37
HPLC-UV analysis of in vitro assay samples 37
Statistical analysis of data 38
Cloning into specialized expression vector for E. coli expression 39
S. cerevisiae expression construct construction 42
Gibson assembly for plasmid cassette creation 43
Transformation of E. coli bacterial cells 43
Confirmation of cellular transformations with expression cassettes 44
Transformation of S. cerevisiae yeast cells 45
Selectivity of media and antibiotics with INVSc1 cellular growth 46

Results 46

Selective media is capable of restricting colony growth 46
CsCHI-L inclusion is not correlated with higher OA production in in vitro assays 48
Building of SBS (E. coli and S. cerevisiae) 54

Discussion 57

Building of the synthetic biology system(s) 57
Codon optimization of sequences 59
Cloning techniques used in construction of SBS 60
Protein expression and the lac operon 60

GAL-induced protein expression 61
Multiple-vector transformations of hosts 62
Use of selective media and antibiotics in SBS construction 63
OA is produced in in vitro enzymatic assays 64
Chapter 3: Olivetolic Acid production in E. coli and S. cerevisiae SBS 67
Introduction 67
Materials and Methods 68
Experimental overview 68
E. coli in vivo assays/ culture growth and feeding experiments 70
OA-Spike feeding experiments 72
S. cerevisiae feeding experiments 73
E. coli culture extraction 73
S. cerevisiae culture extraction 74
HPLC sample preparation 75
LC-MS sample preparation 75
Compound analysis using HPLC-UV 75
Compound analysis using LC-MS/MS 76
Statistical analysis of data 78
Results 78
OA production in E. coli SBS in vivo experiments 78
OA is not degraded by the culturing process and can be detected 84
S. cerevisiae SBS are able to produce high levels of OA in vivo experiments 86
LC-MS/MS Standard Analysis and MRM 90

Discussion **91**

Initial E. coli synthetic biology system in vivo experiments do not produce OA 91

Plasmid incompatibility in the E. coli SBS (Duet1 system) 92

E. coli synthetic biology system in vivo experiments are able to produce OA. HTAL and PDAL 93

Yeast cell degradation for analysis of compounds in S. cerevisiae SBS 97

S. cerevisiae synthetic biology system was not able to be analyzed by HPLC-UV 98

S. cerevisiae synthetic biology system produces OA in vivo 99

Comparison of E. coli and S. cerevisiae SBS in vivo production of OA 103

Chapter Four: Conclusions and future directions ***105***

Introduction **105**

Materials and Methods **106**

Phylogenetic tree construction 106

Homology modeling of CsCHI-L 106

Results **107**

Phylogenetic relationships between CsCHI-L and other CHI and CHI-L proteins 107

Homology modeling of CsCHI-L 109

Discussion **111**

Evolutionary backgrounds of flavonoids, CHS, CHI and FAP 111

Types of CHI and CHI-L proteins 113

Phylogenetic relationships of CHI and CHI-L proteins from constructed tree 115

Homology model of CsCHI-L protein 117

CsCHI-L may function in other sections of the cannabinoid pathway 118

Other proteins may function to stabilize OA biosynthesis in cannabis 120

Alternative synthetic biology system construction 121

Conclusion 125

References 126

Chapter 1: Introduction

Thesis Literature Review

Plant specialized metabolites

Plants are everywhere. Comprising an entire kingdom of life and harbouring an estimated 450 gigatons of carbon biomass, plant life is one of the most domineering forces on the planet.[1] Originating over 500 million years ago, land plants have radiated and colonized all terrestrial environmental niches.[2] Alongside this prolific diversification came the evolution and formation of distinctive features. While many remain developmental or morphological, some of the most unique plant characteristics are the compounds found at the molecular level. A number of compounds are produced as part of the plant's natural metabolism and can be split into primary or secondary classes. Primary natural products, or primary metabolites, are those used by the plant for regular growth, development and reproduction, including carbohydrates (sugars), proteins (amino acids), fats and oils. Plant metabolism also produces other useful compounds that traditionally were not seen as vital for plant regular functioning and termed secondary metabolites. Today, these compounds are referred to as plant specialized metabolites, as it has been demonstrated in numerous organisms that they are indeed necessary for life and prosperity. For example, lignin, a complex natural phenolic polymer accounting for 30% of organic carbon present in the biosphere, is essential for cell wall structure and produced as a specialized metabolite through the phenylalanine/tyrosine pathway.[3,4] These organic specialized metabolites are advantageous to the plant, often allowing protection against abiotic and biotic stress. The advantageous behaviour can manifest in many ways, such as conferring increased reproductive capacity as seen in as attractants, pheromones, and colour signaling.[5,6] Plant specialized metabolites can also be important molecules for defense, as many possess anti-herbivory or anti-

pathogenic properties.[7,8] With a suite of diverse functions, these compounds may be responsible in a large part for the successful outspreading of the embryophyte lineage. Millions of specialized metabolites have been identified, and can largely be divided into four major groups: terpenoids, phenolics, alkaloids and sulfur-containing.[9]

Specialized metabolic pathways share many enzymatic components and compounds with those of primary metabolism, indicative of either a branching from primary metabolism or a shared evolutionary origin.[10] An example of this is phenylalanine, which is used for protein biosynthesis in all life forms yet in plants it is also the first molecule in the phenylpropanoid pathway which produces lignin, flavonoids, and other specialized metabolites.[11] The ancestral genes that encode the enzymes in specialized metabolism usually arose via gene duplication events of genes that operate in primary metabolism. The interconnectivity of these primary and secondary pathways is indicative that secondary metabolism arose due to natural selection acting upon the raw metabolic potential of the enzymatic space in past plant lineages. This can be attributed, in part, to the natural enzymatic promiscuity observed where proteins catalyze non-canonical reactions, resulting in the branching diversification of pathways to new intermediates and compounds. Unlike the ubiquitous primary metabolism, secondary metabolism is commonly lineage-specific, oftentimes arising out of environmental variation and stressors.

With numerous different effects and applications, specialized metabolites are commonly utilized in medical research, pharmaceutical production, the food industry and industrial or agricultural sectors. [12–18] Specialized metabolites are often produced in specific plant tissues at low concentrations and as such the purification and extraction of the compounds results in a low yield.[19] Widespread use and consumption of specialized compounds has led to increased demand and subsequent indiscriminate over-harvesting of many species of wild plants.[20]

Conservation of specialized metabolite-producing plants is dependent on increasing yield via plant breeding and improved agricultural practices, or through the production of the metabolites from alternative sources.

Polyketide synthases and chalcone synthases

Specialized metabolites, while being a diverse group of molecules, often originate from the same basic building blocks derived from the plant's primary metabolic pathways. Knowledge of these existing mechanisms lends insight towards how the enzymes and scaffolds in the specialized metabolite pathways function. For example, a subclass of phenolic compounds produced by plants are the polyketides, which are in part produced through the activity of Polyketide Synthases (PKSs). PKS enzymes produce a polyketide chain from simple starter molecules through a series of decarboxylative condensation reactions.[21] These PKS enzymes share structural and catalytic similarity to the fatty acid synthase (FAS) enzymes that act in plant primary metabolism. Fatty acid biosynthesis elongates the hydrocarbon portion of an acyl substrate by the addition of two carbons. The acyl substrate must first be activated by acyltransferase activity and is subsequently acted upon by the ketoacyl synthase (KS), which is the condensing enzyme responsible for the addition of a two-carbon acetate unit to the substrate.[22] The produced B-ketoacyl thioester is reduced by the B-ketoacyl-ACP-reductase (KR) to an alcohol.[23] A B-hydroxyacyl dehydratase removes water from the intermediate, leaving a double carbon bond. Enoyl-ACP-reductase completes reduction of the B-carbon, producing an acyl-ACP much like the initial compound, but having gained two additional methylene groups.[23] It is common for this product to undergo several rounds of elongation before it is released as a free acid. In animals and fungi, FAS systems are amalgamated into one or two multifunctional protein domains – for example, in mammals, FAS activities are harboured in a single

homodimeric polypeptide that carries out fatty acid synthesis on its own.[24] These are known as type 1 FAS systems. In plants and bacteria, each of the FAS activities is carried out by an individual enzyme.[25] All PKS enzymes have the same KS activity as noted in the FAS pathways.[26] However, instead of proceeding into the reduction and dehydration reactions noted in the FAS mechanisms, PKS systems will often modify or eschew these reactions to preserve polar chemical reactivity in the growing polyketide.[27] PKS enzymes will act upon this intermediate and foster intramolecular cyclization or bond rearrangements to produce a vast collection of compounds. In the same vein as FAS systems have been classified, so have the PKS (Figure 1). Type I PKS, resembling the animal and fungi FAS, form large biosynthetic complexes to carry out their activities. Type I PKS can be sub-divided into iterative and modular PKSs, where iterative PKS will catalyze multiple elongation cycles with the same enzymatic domains and modular PKS will move the polyketide chain through multiple modules, with each module responsible for the catalysis of one elongation cycle (Figure 1).[28] Type II PKS consist of discrete and dissociable proteins that can interact *in vivo* to form complexes similar to those in the Type I PKS (Figure 1). While type I and II PKSs are large protein complexes, type III PKS are smaller homodimeric proteins.[29] Type III PKS enzymes have two independent active sites that facilitate the decarboxylation, condensation and cyclization that results in the production of certain natural products (Figure 1).

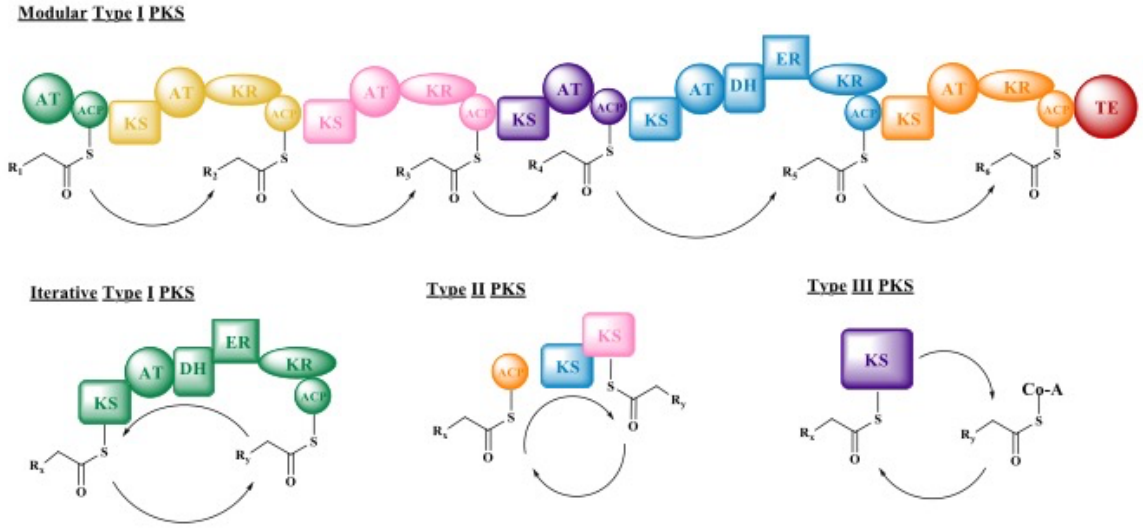


Figure 1. Reaction schemes of different types of polyketide synthases.

Cartoon representation of hypothetical type I, type II and type II PKS enzymes where putative domains are represented by shapes. Type I PKS use an AT loading domain to attach started substrate with ACP. Modular type I PKS passes the substrate/intermediate through different modules (differentiated here with the use of different colour groups), catalyzing subsequent ketone additions until a terminator (TE) domain is reached. Iterative type I PKS catalyzes addition of ketone groups through the same decarboxylative condensation, however the same domain is used repeatedly. Type II PKS are dissociable enzyme complexes, where the domains are used iteratively to produce polyketides but do not require a separate TE, as a portion of the KS domain has TE activity. Type III polyketide synthases are homodimers of KS that act directly on CoA-bound substrates that use iterative decarboxylation and iinternal active site with both KS and TE activity. PKS (polyketide synthase), AT: Acyltransferase; ACP: Acyl carrier protein; KS: Ketosynthase; KR: Ketoreductase; DH: Dehydratase; ER: Enoyl reductase; TE: Thioesterase; CoA: Coenzyme A.

The best-studied PKS is chalcone synthase (CHS), essential for the formation of hydroxychalcone. This enzyme, as well as other CHS-like proteins, acts in the specialized metabolisms of many different plants and some bacteria. CHS catalyzes the addition of three acetate units (from malonyl-CoA) to a *p*-coumaroyl-CoA molecule, sourced from phenylalanine (Figure 2). [30] Chalcones are then formed in the active site via intramolecular cyclization and aromatization. Chalcones produced from the CHS are acted upon in a series of downstream reactions to produce a vast array of important compounds – such as those responsible for

anthocyanin pigmentation, anti-microbial and anti-herbivory defenses, signalling, attractants and more.[31] CHS enzymes have a conserved structure consisting of a dimerized interface and a five-layer αβαβα core, theorized to have arisen from a gene duplication event.[27] The CHS homodimer has two distinct active site cavities, located along the bottom edge of each monomer that are separated by six-residue loops protruding from each interface.[29] The active site of the CHS is buried; accessible only by a CoA-binding tunnel. Reactions facilitated inside the CHS active site include: acyltransferase activity to load the *p*-coumaroyl-CoA starter substrate, a decarboxylase activity to activate the malonyl-CoA substrate, a condensing activity to couple acetyl anions to the polyketide chain, a cyclization activity via a claisen condensation, and finally an aromatase-like activity action to produce the final chalcone product.[21] Conserved among the CHS crystal structure are three residues, referred to as a "catalytic triad" that is responsible for the necessary reactions in the CHS mechanism.[21] Although typically responsible for the formation of 2′,4,4′,6′-tetrahydroxychalcone, CHS can facilitate the catalyzation of other polyketides, such as *p*-coumaroyltriacetic acid lactone (CTAL), a derailment product in the flavonoid pathway.[27] This activity is due to the promiscuity of the enzyme, where the dynamic nature of a protein structure and different chemical mechanisms enable a variety of starter molecules to be incorporated in place of the intended, or result in an alternate end-product.[32]

Through gene and genome duplications, subsequent divergence of the ancestral CHS gene function has resulted in a superfamily of homologous PKS enzymes across the plant kingdom. These PKSs differ in their preference for starter molecules, specific activity, mechanism of addition and termination, number of acetyl units incorporated in the growing polyketide, and alternative patterns of cyclization.[27] The divergence of this PKS family has resulted in a plethora of specialized metabolic pathways in different plant lineages. An example

of this is in the family Cannabaceae, in which the PKS valerophenone synthase (VPS) in *Humulus lupulus* L. (hops) is a component of the pathway that produces the bitter phenolic acids that are used to flavor beer.[33] Similarly, in *Cannabis sativa* L. (cannabis) the PKS tetraketide synthase (TKS) is a component of the cannabinoid biosynthetic pathway.

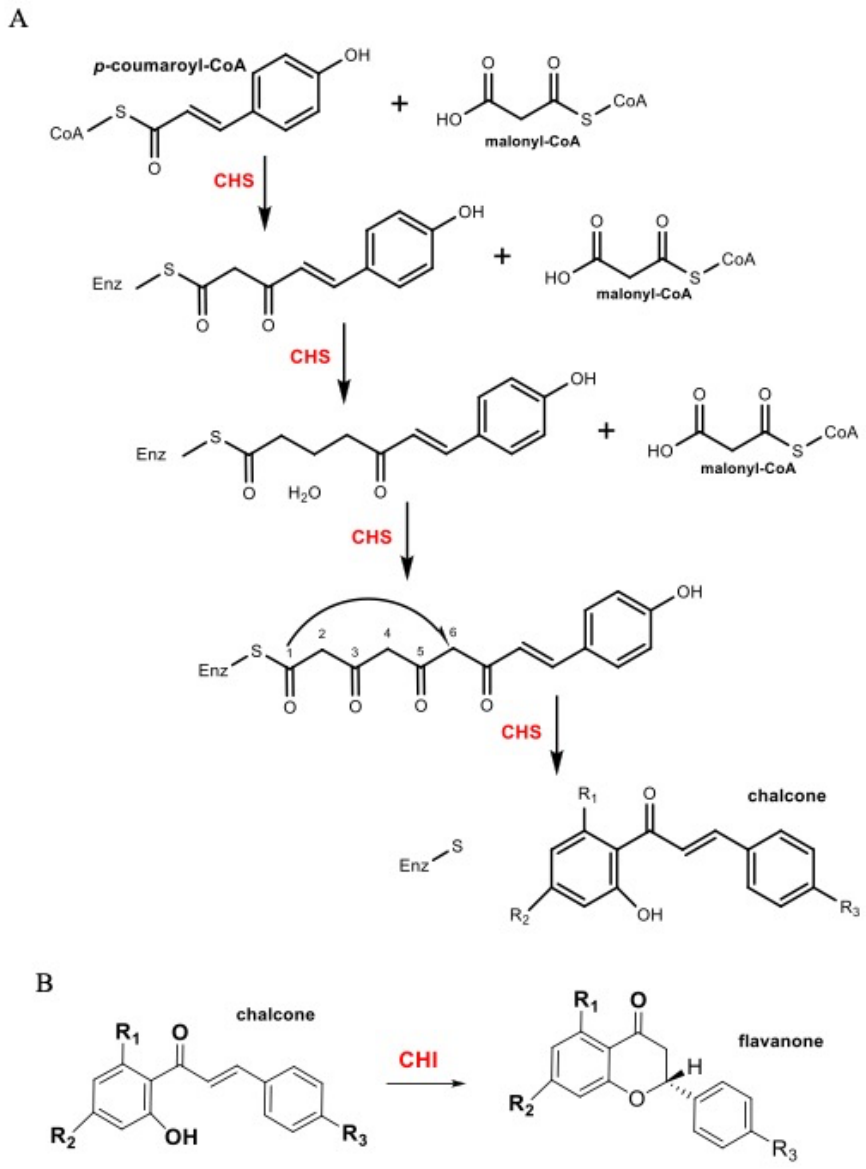


Figure 2. Compound production through chalcone synthase and chalcone isomerase enzymes.
A) Production of chalcone from *p*-coumaroyl-CoA and malonyl-CoA through chalcone synthase (CHS) enzymatic activity. B) Production of flavanone from chalcone through chalcone isomerase (CHI) enzymatic activity.

CHI and CHI-L accessory proteins

An integral component of many biosynthetic pathways are proteins known as accessory proteins. While not able to directly catalyze reactions themselves, accessory proteins accompany and assist the major enzymatic components in some fashion. Examples of this include helping to hold a substrate in place, chaperoning an intermediate from one enzyme to another, holding an enzyme in proper conformation, and many others.

In the flavonoid pathway the chalcone produced by CHS is next converted to flavanone via chalcone isomerase (CHI; Figure 3). This stereospecific isomerization is a committed step in the flavonoid and anthocyanin production pathway in plants. Although this isomerization can occur spontaneously, a 10^7 fold increase in flavanone production is noted with the presence of CHI enzymatic activity.[34] CHI enzymes can be divided into four groups, largely based on phylogenetic analysis.[35] Type I and II are considered bona fide catalysts, with enzymatic activity where type II exhibits broad substrate specificity.[36–38] Type II CHI proteins are regulated by an isoflavaniod-specific gene and as such, appear to be somewhat legume-specific, operating in isoflavanoid production.[35,37] Type III and IV CHIs do not have cyclization activity and are denoted as "Chalcone isomerase-like" (CHI-L) proteins. CHI-L proteins are a type of accessory proteins, where they bind to CHS and enhance product formation and rectify derailment compound production.[39–41] Type III CHIs are predominant in land plants and algae and assist in the binding and storage of fatty acids.[42] Type IV CHI function remains to be somewhat of a mystery, with some evidence pointing to their function in flower coloration or flavonoid production.[43] CHI-L presence has shown to have a direct positive correlation with flavanol and flavone production , as demonstrated in tobacco plants.[44]

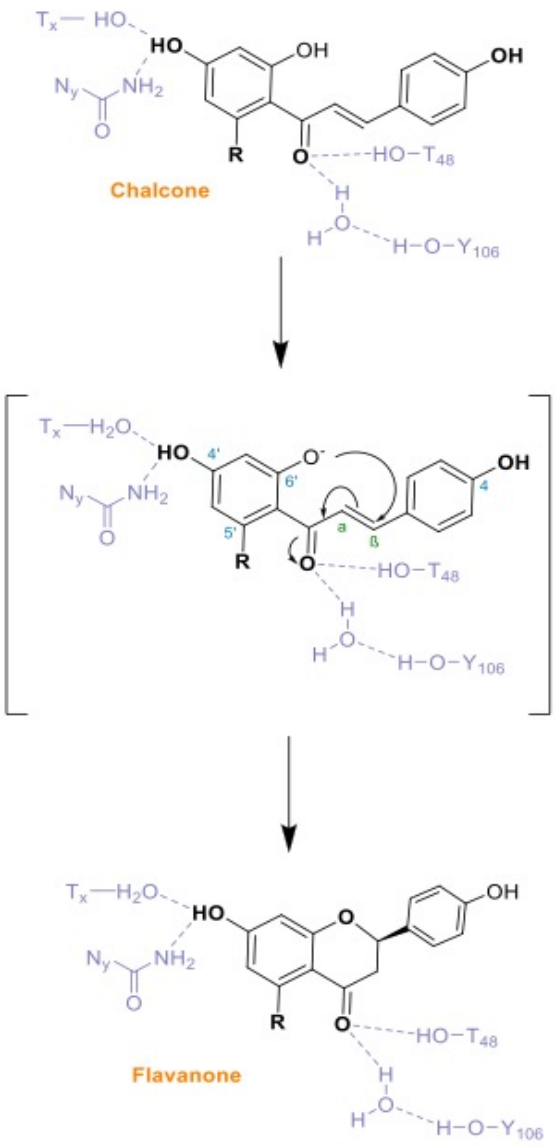


Figure 3. Proposed reaction mechanism for the cyclization of chalcone to flavanone through chalcone isomerase (CHI) activity.

Diagram highlights the involvement of conserved residues via hydrogen bonding in CHI active site. The reaction begins with deprotonation of the chalcone 2'-OH group, followed by the resultant 2-oxyanion nucleophilic attack on the α,β-unsaturated double bond, stabilized by a water molecule and forming an intermediate that produces the flavanone product through tautomerization.

Recently it has been shown in hops that CHI-L proteins are involved in demethylxanthohumol (DMX/XN) production. Although they do not possess catalytic activity,

they directly interact directly with chalcone synthase (HlCHS1) synthase via protein-protein interactions to modulate the activity of the PKS.[39] XN/DMX biosynthesis occurs in the glandular trichomes of female hops flowers, where trichome-specific genes facilitate the conversion of phenylalanine and the prenylation of intermediate molecules with dimethylallyl diphosphate (DMAPP) into chalcones DMX and XN.[33,45,46] CHI-L proteins were theorized to act in the DMX/XN biosynthetic pathway based on inefficiencies in the pathway when constructed in a biosynthetic host, as well as prevalence of CHI-L genes in the trichome expressed sequence tag (EST) libraries. HlCHI-L2 protein was shown to interact directly with the CHS in the XN/DMX pathway and resulted in a 2.3-fold increase in DMX production when included in engineered strains.[39] In this same study, the HlCHI-L1 protein was shown to be involved in the storage of chalcone-product intermediates, where it binds and stabilizes open-ring conformation of DMX and naringenin chalcone.[39]

Cannabis and specialized metabolites phytocannabinoids

Today, cannabis is a household name, referring to the annual flowering herbs renowned for their multitude of uses. Cannabis plants are a genus of the family *Cannabacea*, which encompasses eleven other genus groups including *Humulus*, used as hops in beer production, and *Celtis*, grown mainly as decorative and ornamental shrubs. Cannabis has an extremely long and colorful history of use by human civilizations, and in present-day is globally regarded as a plant of high economic value. It is believed that cannabis and hops diverged from a common ancestor approximately 28.7 million years ago, and originated near Qinghai Lake of the Tibetan Plateau.[47] From there, cannabis spread both westward into Europe and eastward towards China. Cannabis has been an important plant in many ancient civilizations, used for medicinal, industrial and recreational purposes in China, India and Tibet.[48] In China, the first records of cannabis use were

discovered, and archaeological findings indicate that the plant has been cultivated there since 4000 B.C.[49] Cannabis use was also widespread in India as a large component of Avurveydic medicine, in religious ceremonies, and as a source of hemp textiles. Cannabis is noted as a sacred plant in text The Atharva Veda and was believed to be a source of happiness, donator of joy, and bringer of freedom.[50] The use of cannabis as a medicine was first reported in the world's oldest pharmacopoeia, the pen-ts'ao ching, based on oral traditions passed down from the Emperor Shen-Nung in 2700 B.C.[51]

Differentiation in the global dispersal of cannabis plants lead to cultivation of different strains (often confused with speciation). Much of the differential taxonomy of cannabis has been contested; proper current denotation is to refer to all cannabis as *C. sativa*, where varying sub-types may occur (i.e. indica, ruderalis). Utility of cannabis plants comes in many different forms. Hemp fibres, acquired from the stems and branches of the cannabis plant, are important in production of textiles and cloths. Seeds, cultivated long ago as important food crops, are often now hulled and sold as "hemp hearts", touted for their rich nutritive profiles that include fatty acids and other bioactive components. [49] Specialized metabolites phytocannabinoids are renowned for their physiological and psychotropic properties.

Cannabinoids are naturally occurring terpenophenolic compounds, produced both as lipid-based neurotransmitters in animals (endocannabinoids) and as part of the secondary metabolism of plants (phytocannabinoids). Tetrahydrocannabinol (THC) and cannabidiol (CBD) are the most abundant and widely known of these compounds, derived from precursor phytocannabinoid cannabigerolic acid (CBGA) in cannabis plants.[52,53] Cannabis plants are known to produce over 140 different alkyl cannabinoids, including the lesser known cannabinol, cannabigerol and cannabichromene.[54–56] The production of phytocannabinoids is also noted in

several other species, including in algae, bryophytes and monilophytes.[57] *Helichrysum umbraculigerum*, a perennial herb native to South Africa, produces resorcinol derivatives that have a close relation to CBG/CBGA.[58] *Radula marginata*, a New Zealand liverwort, synthesizes cannabinoid-like bibenzyl compounds.[59] These cannabinoids can differ greatly from species to species, in both profile and in molecular basis. In *C. sativa*, cannabinoids are derived from aliphatic ketide starters, where phytocannabinoids in other organisms are often sourced from an aromatic ketide starter.[55] Phytocannabinoids are most abundantly produced and accumulated in cannabis when compared to other sources.

In cannabis, phytocannabinoids are produced in the glandular trichomes that are abundant on female flower tissue but also occur on leaves and stems. Trichomes, a type of epidermal outgrowth, are specialized in their function and commonly found in many plants. Cannabis has four types of trichomes: three secretory glandular trichomes (capitate-sessile, capitate-stalked and bulbous) as well as non-glandular, unicellular hair-like trichomes.[60,61] Glandular trichomes have a secretory disk, comprised of 8-13 cells arranged into a rosette that functions as the site of synthesis for many compounds, including cannabinoids, fatty acids and terpenes.[61] These compounds accumulate in a subcuticular cavity, creating the bulbous appearance of the trichomes.[61] Cannabinoid accumulation is the highest in capitate-stalked trichomes. Due to the prodigious amount of cannabinoids and terpenoids produced by these trichomes, mature flowers take on a frosty or crystalized appearance. Tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) are the most abundant cannabinoids in fresh trichomes, followed by cannabinol (CBN) and neutral cannabinoids, THC and CBD.[62] CBGA is also present in large amounts in trichomes, co-existing with THCA and CBDA during the flowering period when the largest production of cannabinoids occurs.[52]

Cannabinoids are monoterpenoid compounds consisting of a resorcinol core with a varying side chain that can contribute to pharmacological effects.[55] Characterization of the active constituents of cannabis plants began with the elucidation of the structure of Δ9-THC, followed shortly by the determination of the structure of cannabinolic, cannabidolic and cannabigerolic acids.[63,64] Determination of the active constituents structure lead way for molecular cloning and identification of the main cannabinoid target, CB1 primarily located in the central nervous system .[65] A second peripheral cannabinoid receptor, CB2, was later identified and proposed to have a strong immunomodulatory role.[66] Both CB1 and CB2 belong to the class A G-protein coupled receptor (GPCRs), with residues extending into the binding site which mediate interactions with different ligands.[67,68] Although similar, differences noted between the binding poses and side-chain conformation of ligands in the receptors may be accountable for high CB1/CB2 selectivity.[67] With the discovery of CB1 and CB2, it was possible to identify the endogenous ligands of these receptors that operate mammalian physiology. The so-called 'endocannabinoids' consist of arachidonic acid, anandamide , and their related derivatives.[69] These ligands, together with their receptors, form the endocannabinoid system (ECS) that is present in chordates. The main function of this system in the CNS is to reduce overstimulation of the nervous system. Phytocannabinoids exert their effects in the body via cannabinoid receptors (CBRs), CB1 and CB2.[70,71] The multiplicity of ligands and subsequent signalling cascades from CB receptors provides numerous opportunities for cannabinoid therapeutics to treat a myriad of ailments. [72] Cannabinoids have shown efficacy as a treatment for epilepsy; related to their anticonvulsant and ant-epileptiform properties. One study demonstrates that Δ^9-tetrahydrocannabivarin (Δ^9-THCV) acts upon the endocannabinoid system in rats, activating a CB1 receptor-mediated mechanism and increases the latency of seizure occurrence in a

mechanism theorized to be similar to the gamma-aminobutyricacidergic system.[73] Much of the therapeutic potential of cannabinoids is related to their anti-inflammatory action, a result of the interaction between the ECS and the immune system. Cannabinoid type 2 receptors can be regulated by CB2 agonists, such as cannabinoids and their derivatives, to reduce inflammation and infection.[74] Both airway inflammation and systemic immune response were greatly reduced by treatment with the CB2 agonist THC, providing relief of respiratory conditions such as asthma and COPD.[75] CBR agonists directly reduce the number of inflamed microglia in the brain, and show an effective treatment for chronic brain inflammation.[76] Reduction in inflammation also contributes to neuroprotective properties, as demonstrated by several cannabinoids and derivatives, including THC and CBD.[77] An aminoquinone derivative of cannabigerol increases the recovery of mice subjected to lipopolysaccharides (drivers of inflammation and neuronal deterioration) and reduces the autophagy impairment usually imparted by lipopolysaccharide lesions.[78] Additionally, overexpression of ECS receptors exerted neuroprotection against Parkinson's Disease and reduced neuroinflammation in Alzheimers Disease in animal models.[77] Cannabinoid receptors are localized in many regions involved in transmission of neurological pain, including periaqueductal gray, rostral ventromedial medulla, spinal trigeminal tract and nucleus, dorsal horn and lamina X of the spinal cord.[70,79] These locations are suggestive of the endocannabinoid systems ability to modulate pain via the central nervous system. More recently, cannabinoid efficacy in pain management has been proposed to be correlated to the inhibition of canonical Wnt/β-catenin signaling .[80] Cannabinoids also have the potential to attenuate the undesirable effects of morphine through CB-related mechanisms while enhancing the antinociceptive potency of opiates.[81,82] The aforementioned studies and additional ongoing research shows that cannabinoids exert effects and interact with a large

spectrum of macromolecular targets including transporters, channels, enzymes and receptors outside of CB1/CB2.[83] Medicinal plants such as cannabis are therefore undeniably important; both directly as therapeutic agents and indirectly as models for the synthesis of other pharmacologically active compounds. Synthetic cannabis-based medications such as Dronabinol, Nabilone and Nabiximols are currently used in the UK, USA and Canada. Dronabinol (Marinol®) and Nabilone (Cesamet®) are prescribed mainly to treat cancer patients experiencing nausea or appetite loss, with secondary use for pain management. Dronabinol has also demonstrated analgesic effects when used in concert with opiate treatment for chronic pain.[84] An increase in demand related to both the pharmaceutical potential and recreational uses of cannabinoids has created a need for an efficient and low-cost source of cannabinoids. Current sourcing relies on synthetic production of cannabinoids or on crop cultivation and harvesting of plant material.

In planta production of phytocannabinoinds

In general, phytocannabinoids are produced by the enzymatic conjugation of a resorcinol core (featuring an alkyl side chain) and a monoterpene moiety derived from geranylpyrophosphate (GPP). The biosynthetic pathways responsible for the production of both compounds have largely been determined in the past two decades. GPP precursors are produced by the 2C-methyl-D-erythritol-4-phosphate (MEP) pathway, where the majority of enzymatic components have been previously described in plant model species. Briefly, isopentyl diphosphate (IPP) is combined with dimethylallyl diphosphate by geranyl disphosphate synthetase.[85]

The biosynthesis of the alkylresorcinol component of phytocannabinoids begins with an acyl activating enzyme (CsAAE1) catalyzing the conversion of hexanoic acid to hexanoyl-CoA

(Figure 4). [86] The origin of hexanoate in cannabis plants is still under investigation, theorized to be derived from the cleavage of longer-chain fatty acids.[86] Hexanoyl-CoA is then extended with acetate units derived from three malonyl-CoA's to produce a linear tetraketide; this reaction is catalyzed by the type III polyketide synthase CsTKS (Figure 4).[87] CsTKS is hypothesized to be a homodimeric enzyme, where each unit is a 385 amino acid polypeptide to create an 89 kDa protein.[88] The malonyl-CoA necessary for this reaction is produced via the carboxylation of acetyl-CoA. The linear tetraketide intermediate is then cyclized via Claisen condensation by olivetolic acid cyclase (CsOAC) (Figure 4).[89] CsOAC is instrumental in this reaction; without it, the CsTKS alone will only be able to produce olivetol and two derailment products, α-pyrone pentyl diacetic acid lactone (PDAL) and hexanoyl triacetic acid lactone (HTAL: Figure 4).[89] CsOAC is an α + β barrel (DABB family) protein that facilitates the C2-C7 aldol condensation of the tetraketide intermediate to produce olivetolic acid, OA.[89] OA is then prenylated with GPP to create cannabigerolic acid, which functions as a substrate for various oxidocyclases that produce the phytocannabinoids Δ9-THCA, CBDA, and other less-abundant phytocannabinoids (Figure 4).[53,90–92] Expression of all genes utilized in both this pathway and the aforementioned methylerythritol phosphate (MEP) pathway have been noted to be highly expressed cannabis trichomes, where the majority of phytocannabinoids are produced in cannabis plants.[52,86,93]

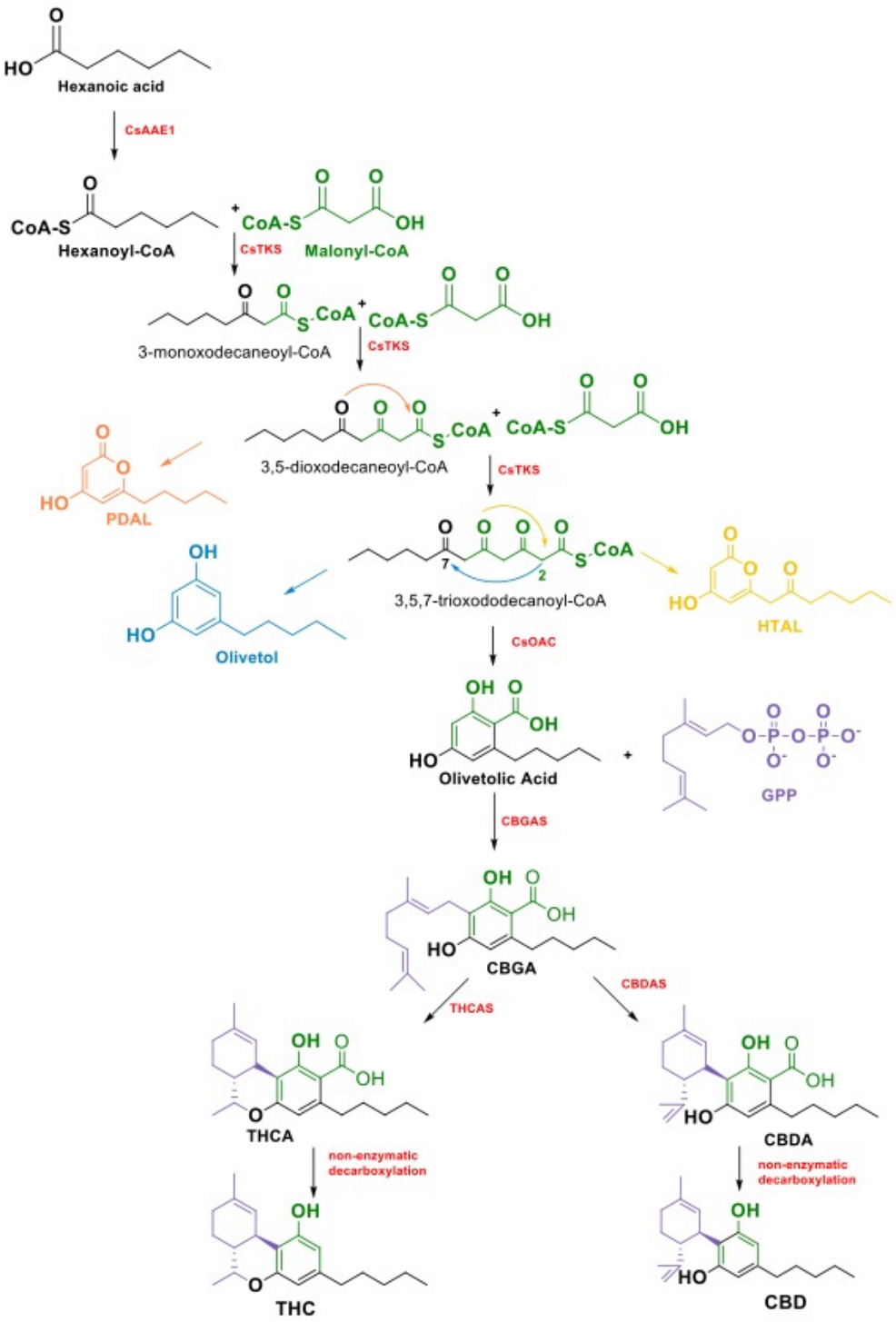


Figure 4. Cannabinoid biosynthetic pathway.

Detailed schematic of the proposed production of Δ^9-tetrahydrocannabinol (Δ^9-THC) and cannabidiol (CBD) as seen in *Cannabis sativa*, starting with hexanoyl-CoA and malonyl-CoA

and including pathway derailment products such as hexanoyl triacetic acid lactone (HTAL), pentyl diacetic lactone (PDAL) and olivetol.

Cannabis produces copious amounts of THCA and CBDA, and as such is currently the main source for phytocannabinoids. However, agricultural production of cannabinoids does face several challenges such as uncontrollable climate, susceptibility to diseases, varying content of cannabinoids, chemical processing for extraction, issues related to legality regarding cannabis illicit use, and environmental impact. From seed to flower, cannabis crops have an average growing period of three to five months. Cannabinoid production *in planta* begins around week four, increasing until week 28 but achieving a relative plateau around week 16.[62] Operations growing cannabis plants, even on a commercial scale (in warehouse or greenhouse) are therefore only able to achieve a maximum of four harvests per year. Growth of cannabis plants requires a significant amount of physical space to cultivate these rather large shrubs, as well as labourers to regiment watering, light cycles, and general care of the plants. As recreational cannabis is illegal in most of the world (and still a controlled substance in countries that have made recreational consumption legal) the large-scale growth of cannabis plants requires security and monitoring measures to ensure the protection of the crop. While certain climates may be able to take advantage of outdoor growth to provide less expensive options via agricultural farming, marijuana plants require specific light cycles to both trigger flowering and to reach maximal growth and cannabinoid production. Additionally, variation in mineral and nutrient availability has been demonstrated to affect specialized metabolite production. Phosphorus availability has been shown to affect phytocannabinoid biosynthesis, where phosphorus concentration lower than 15 mg/L did not support optimal plant function and concentrations of 30-90 mg/L were required for optimal cannabinoid content development.[94] Nutrient availability therefore must be adjusted

and tightly regimented (with considerations for the needs and nutritional profile required for production goals as well as the genetic requirements of the strain), providing an added layer of difficulty to phytocannabinoid production. Cannabis plants grown as crops outdoors generally have substantially lower THC and CBD content compared to plants grown with the proper regulated light cycles. Accidental fertilization of female plants with male pollen is also a concern, given that fertilization diverts most of the carbon from the cannabinoid pathway to seed production, resulting in 75 and 60% decreases in THC and CBD content, respectively[95]. These factors, combined with the need for security has largely directed most cannabis growth to indoor facilities. The sustainability of cannabis farming can be brought into question, as it requires many resources and expenses to produce physically large plants that are broken down and processed for molecular compounds. Cannabinoid production is limited to the glandular trichomes, with very low numbers of trichomes located on leaves and stems, rendering much of the produced biomass useless in terms of phytocannabinoid harvesting. In the United States, estimates for the greenhouse gas (GHG) emissions of indoor cannabis cultivation are considerable, and range from 2,283 to 5,184 kg CO_2/kg of dried flower.[96] These emissions are largely attributed to the consumption of external CO_2 to stimulate cannabis accelerated growth, energy required for high-intensity grow lights, and other environmental controls such as heat and water. These estimates are restricted to GHG emissions related to growth and harvest of cannabis flower, and do not include the further steps required to process, store, transport and package the product. Furthermore, processing of the plant and extraction of cannabinoids for specific pharmaceutical use requires chromatographic separation and purification of individual cannabinoids, even if they are to be recombined for full-spectrum use. This is a time consuming step requiring highly specialized equipment and techniques. When looking at production of

phytocannabinoids for medical applications, it is feasible to think that an alternate route of production may eventually provide a better (or at least competitive) sourcing of these compounds in terms of specificity and sustainability.

Alternative production of cannabinoids

Synthetic chemistry approaches have been successful in producing analogues of cannabinoids, termed 'synthetic cannabinoids'. However, chiral products are common in these syntheses and often one enantiomer possesses reduced efficacy or off-target effects. Enantiomer separation is often difficult and cost prohibitive in large-scale productions. Synthetic production results in compounds that, while similar, are not identical to natural phytocannabinoids. These compounds differ in their affinity for CB receptors, often resulting in a skewed potency and potential for more severe effects (including toxicity) when compared to naturally produced THC and CBD. The increased potency available in synthetic cannabinoids makes them popular for recreational use and abuse. Synthesis of the first synthetic cannabinoid was achieved several decades ago in Israel, by the production of enantiomers of 7-hydroxy-delta6-tetrahydrocannabinol.[97] These two synthetic cannabinoids demonstrate the extreme stereospecificity of the CB receptors and endocannabinoid system – the 3R, 4R enantiomer has a cannabimimetic profile, is highly psychoactive and 70-800 times more potent than THC, whereas 3S, 4S enantiomer demonstrates no cannabimimetic properties, but can function as an antiemetic and N-methyl-D-aspartate (NMDA) receptor blocker.[97] This difference outlines the necessity for an extremely high enantiomer specificity when producing synthetic compounds of this nature. During this time, Pfizer also directed studies towards the synthetic production of cannabinoid analogues, producing a series of cyclohexylphenols or 3-arylcyclohexanols (CP) compounds. Synthesis of a particular CB1 agonist (CP 47,497) was achieved from acetonitrile in 9 steps and

showed equal potency to morphine as an analgesic compound.[98] A group of synthetic cannabinoids were also produced by John W. Huffman, an American chemist in search of cannabimimetic compounds for pharmaceutical therapeutics.[99] The most famous of these compounds, JWH-018, exerts the classic effects of cannabinoids but is far more potent. This increased potency can lead to detrimental effects including respiratory depression and death [100]. Aminoalkylindoles (such as the JWH compounds) possess a high affinity for CBR's, where binding affinity has been demonstrated to be at least threefold higher than that seen in THC for both CB1 and CB2. [101] Synthetic cannabinoids can impart a wide variety of negative symptoms, including psychosis, tachycardia, vomiting, seizures and central nervous system depression.[102,103] Recreational use of synthetic cannabinoids results in a number of negative physiological and psychological impacts, due to the compounds high potency. Acute kidney damage and nephrotoxicity has been noted with use of synthetic cannabinoids.[104,105] Cardiovascular abnormalities, such as tachycardia and cardiac arrythmia are also correlated to use of synthetic cannabinoids.[106] There is evidence that synthetic cannabinoids lower seizure thresholds and induce convulsions, where natural cannabinoids demonstrate anti-convulsant effects.[107,108] Synthetic cannabinoid prevalence and abuse poses a significant psychological threat as well – risk of psychosis increases with cannabinoid compound potency, possibly correlated to increasing CB1 agonist activity of synthetic cannabinoids such as JWH-018 compound.[109] Synthetic cannabinoids have been shown to differ in their signalling pathway compared to their natural counterparts, specifically that they possess cAMP stimulation properties and efficacy in Beta-arrestin translocation.[110] These differences in signalling mechanism can correspond to a variety of downstream effects that have negative implications for the therapeutic potential compared to natural cannabinoids.

A technique utilized for the production of high-value compounds is plant-tissue culture, which has been used commercially to produce diosgenin-derived steroid hormone precursors, digitalis glycosides, berberine isoquinoline alkaloid, taxol, paclitaxel, and several other toxoids. Large-scale plant tissue or organ culture to produce secondary metabolites is an attractive alternative approach to traditional methods of cultivation of drug plants. The advantages of this method are many; it is independent of soil, paste, climatic interference, geographical location, and it can ultimately provide a continuous and reliable source of natural products. However, cannabis cell culture systems do not accumulate high concentrations of cannabinoids, , potentially due to phytotoxic effects of THCA and CBGA.[111] Potential toxicity of THCA and CBGA also noted *in planta*, where the phytocannabinoid content decreases in the last two weeks of the flowering period to an assumed degradation to release volatile compounds.[62]

Synthetic biology systems to produce cannabinoids

In more recent years, attention has shifted to synthetic biology approaches, wherein genes of interest are incorporated into microorganisms to produce specialized metabolites. This type of *in vivo* production, where all genes involved in biosynthesis of metabolites are co-expressed in a microbial system [112], can provide an efficient and competitive method for high-yield biosynthesis of targeted compounds. A full understanding of the *in planta* biosynthetic pathway is needed to achieve this sort of production. Following identification of participatory enzymes and proteins, hosts are engineered by the reconstruction of metabolic pathways to produce specialized metabolites. Efficient production of these compounds requires strategic planning, formatting and modification of the microbes due to strict regulations of metabolic pathways.[113] *E. coli* and *S. cerevisiae* are commonly used host systems, due to their strong biosynthetic capacity and the level of knowledge accumulated about their intrinsic metabolism and genetics.[114,115] Synthetic

biology systems (SBS) have shown to be advantageous in many ways. Terpenoid production reliant on cytochrome p450-based oxidation, which is difficult to scale using chemical synthesis, has been achieved at high titre using *E. coli* hosts.[116] Promiscuity of certain pathways can also be exploited in microbial systems, where pathways are directed towards producing diverse or rare compounds. Starter molecules have been converted to diverse stilbene polyketides using recombinant *E. coli* cells.[117] Microbial hosts are able to facilitate the conversion of starter products to high-value end compounds by the inclusion of multiple enzymatic steps and conversions, as demonstrated in the production of benzylisoquinoline alkaloids for pharmaceutical use in a 10-gene plant pathway incorporated into host *S. cerevisiae.*[118] Utilizing enzymatic components from multiple sources to create a recombinant pathway, where the promiscuity of protein action on substrates can produce a larger array of compounds, also has a role in the field of synthetic biology.[119] The identification of biosynthetic pathways is the largest restricting factor when it comes to natural product biosynthesis in host organisms.[113] Elucidation of the enzymes involved in the plant biosynthetic pathway enables the reconstruction of the synthetic pathway in the host. As the entire genome for *C. sativa* has been archived and most of the cannabinoid biosynthetic pathway has been elucidated, *in vivo* cannabinoid production is now feasible. However, to date the incorporation of these genes into microbial hosts has resulted in low-yielding platforms that are not economically competitive compared to growing cannabis plants. A successful study producing cannabinoids and intermediates in yeast suggested that production of said compounds is limited by the conversion of malonyl-CoA and hexanoyl-CoA to OA.[120] This bottleneck was revealed when the production of THCA was increased following overexpression of the CsOAC/CsTKS enzymes, but no discernible difference was noted when *C. sativa* prenyltransferase (CsPT4) and THCAS enzymes were overexpressed.[120] Additionally, a

biosynthetic platform created to produce OA in *E. coli* hosts was ultimately successful but experienced difficulties achieving a high OA yield.[121] These results are indicative that the pathway needs to be more efficient in the production of OA.

Hypothesis

CHI-L proteins have been demonstrated to interact with CHS/PKS enzymes to increase their activity. I hypothesize that a CHI-L protein is working in the cannabinoid biosynthetic pathway to support CsTKS/CsOAC activity for the production of OA from hexanoyl-CoA and malonyl-CoA. When CsOAC was characterized, three gene candidates were identified that could possess the OA-cyclase activity, of which *CsOAC* was the correct gene.[89] One of the other candidates encoded a CHI-L protein, which was selected based on the catalytic relationship of CHS/CHI and due to its representation in the trichome EST dataset. An idea first put forward by Taura et al., it was theorized that the CHI-L could provide auxiliary support to the CsTKS for the formation of OA.[88,89] Even though CsOAC is the *bona fide* cyclase that functions with CsTKS to produce OA, there is still a possibility that the CHI-L also functions in this pathway. CsTKS synthesises an intermediate product that is acted upon by CsOAC, but the two enzymes do not physically interact. It is plausible that the CHI-L protein acts an intermediate shuttle, guiding the released tetraketide intermediate from CsTKS to CsOAC to by cyclized. The CHI-L protein could also act at other points in the pathway; interacting with the prenyltransferase or working towards THCA/CBDA stabilization and storage.

Chapter 2: Building Synthetic Biology Expression Systems

Introduction

Synthetic biology approaches have become revolutionary in the field of molecular biology. They allow for the high-yield, low-cost sourcing of high value compounds, as well as provide the opportunity to direct or tailor pathways for specific or altered products. A large area of interest relating to construction of a phytocannabinoids production system using synthetic biology approaches is the ability to direct the system to produce rare or novel phytocannabinoids. Upwards of 150 phytocannabinoids have been identified and are produced in cannabis plants. [55] However, most of these bioactive compounds are not present at levels high enough to be extracted or sourced from plant biomass in the same manner as other more abundant phytocannabinoids, such as THCA and CBDA. These compounds include (but are not limited to) cannabinol (CBN), cannabigerol (CBG), cannabidiphorol (CBDP), cannabichromene (CBC), tetrahydrocannabivarin (THCV), cannabidivarin (CBDV) and cannabicyclol (CBL). Extracts from *Cannabis* plant matter often contain a multitude of other specialized metabolites that have a varying composition dependent on chemotypes or cultivars.[122] Presence of these other compounds in extracts require expensive purification processes to achieve a high purity and yield of phytocannabinoids. Even in floral tissue the concentration of the minor phytocannabinoids is low – a recent whole-plant profiling of *Cannabis* phytocannabinoid content showed the concentration of CBN and and CBDVA to be 0.01% and 0.05% by weight (respectively) in inflorescence samples.[123] An additional point of concern regarding extraction is that the acidic forms of cannabinoids are often lost or decreased by spontaneous decarboxylation via heating during the process. Many of the acidic forms can be more therapeutically potent or active than the neutral forms of the cannabinoids.[124] Some of the minor phytocannabinoids are being investigated for many different pharmaceutical and therapeutic applications. CBN has been

identified as an analgesic and anti-inflammatory agent, providing synergistic pain relief when used in combination with CBD.[125] CBN also demonstrates use as an anticonvulsant, and has shown to delay the onset of neurogenerative disease.[126,127] CBC shows promising ant-inflammatory effects as well, with inflammation and pain reduction noticed with treatment for osteoarthritis as well as gastrointestinal inflammation.[128,129] As with CBN, CBC shows a synergistic increase of anti-inflammatory properties when combined with THC.[130] CBG has shown pre-clinical evidence for treatment of inflammatory bowel disease.[131,132] THCV demonstrates potential as both an anticonvulsant, and for treatment of neurogenerative disease.[133] Should these minor phytocannabinoids be adopted as pharmaceutical compounds, a suitable system for their production will be required. Using metabolic engineering strategies, specific phytocannabinoids and metabolites of interest can be produced by microbial hosts.[120] Furthermore, should the enzymes of the pathway exhibit sufficient substrate promiscuity, the introduction of synthetic or unique starter and intermediate molecules can result in the production of novel cannabinoids.[134] Heterologous production can also be combined with chemical synthesis to produce synthetic cannabinoids that have similar physiological effects to their *in planta* counterparts.[135,136]

While SBS are powerful tools for sourcing of high-value compounds, they also afford investigative opportunities when looking into the mechanisms and components of a biosynthetic pathway. In addition to synthesizing specialized metabolites, synthetic biology approaches can be used to identify unknown enzyme and protein function.[113] In such situations, candidate genes are included in said microbial systems to either alter the activity of the pathway, or to produce new metabolic compounds via the activity of a candidate enzyme. This can help directly identify the biochemical function of an unknown gene. Candidate genes are selected based largely upon

similar homology to a previously investigated sequence.[137] In this experimentation, I incorporated the previously identified CsCHI-L protein into the basic enzyme scaffolding of the *C. sativa* phytocannabinoid biosynthetic pathway. I hypothesized that CsCHI-L, operates as an accessory protein that works in concert with the CsTKS and CsOAC enzymes to produce intermediate compound OA. The ORF of this CHI-L protein shares a 69% homology to a CHI-L protein identified in hops that demonstrated to have direct protein-protein interaction with a CHS enzyme to facilitate higher internal production of prenylchalcones.[39] The close relation of Cannabis and Hops (both belonging to the single family Cannabinaceae) and the similar production of prenylated polyketides in both plants gave me reason to believe the CHI-L proteins could be functioning in a similar manner. To investigate the role of the proposed CHI-L in the phytocannabinoid biosynthetic pathway, I first reconstructed the OA-forming steps in a suitable host system. Although transgenic plants could offer a route for this type of investigation, it remains difficult to engineer plants for improved specialized metabolite content or directed production of intermediate products. Additionally, there are many legal restrictions around *C. sativa* research, given that phytocannabinoids are a controlled substance and work with active plant materials requires specialized licensing and security measures. Microbial hosts offer an attractive alternative to the plant host, as they grow quickly and with minimal resources, and are easily transformed to incorporate genes for heterologous expression. Both *E. coli* and *S. cerevisiae* were used as host organisms for this experiment to increase my chance of successful OA production.

Materials and Methods

Experimental overview (designing and constructing the synthetic biology systems)

All pertinent enzymes for OA production from starter molecules malonyl-CoA and hexanoyl-CoA were introduced through heterologous recombination into microbial hosts *E. coli* and *S. cerevisiae*. Multiple vectors were used to incorporate all necessary genes (CsAAE1, CsTKS and CsOAC alongside our candidate protein CsCHI-L). The expression vectors pCOLADuet-1 and pCDFDuet1 were selected for the *E. coli* system, and pESC-HIS and pCEV-G4-KM were selected for the *S. cerevisiae* system. Through a series of cloning efforts and techniques the genes were properly incorporated into different BL21 or INVSC1 host strains. These microbial strains were then subjected to a multitude of *in vivo* feeding experiments to determine the capability of the systems to convert starter molecules to compound of interest, OA. Proteins were also individually expressed and isolated using HIS-Tag purification to determine *in vitro* functionality through enzymatic assays.

Cannabis candidate sequence identification

The *CHI-L* nucleotide sequence was first identified from a cannabis trichome EST library.[86] This candidate gene was codon-optimized for expression in *E. coli* and *S. cerevisiae* using the Integrated DNA Technologies (IDT) codon optimization tool (https://www.idtdna.com/pages/tools/codon-optimization-tool) and the optimized sequences were purchased from IDT.

Additional sequences used included the *Cannabis sativa* genes for the tetraketide synthase (*CsTKS*), olivetolic acid cyclase (*CsOAC*), and the acyl-activating enzyme (*CsAAE1*), and a *Primula obconica* polyketide synthase *PoPKS*. *E. coli* codon-optimized versions of these genes had previously been cloned and characterized by fellow Masters student Kyle MacDonald.

These genes were subjected to codon-optimization for expression in *S. cerevisiae*, and were also purchased from IDT for use in this experimentation. The gene encoding malonyl-CoA synthetase from *Bradyrhizobium japonicum* (*BjMCS*) was only optimized for use in the *E. coli* expression system, *as S. cerevisiae* have endogenous malonyl-CoA synthetase activity and therefore is not necessary to supplement the yeast system with this activity. Finally, genes encoding the prenyltransferases CsPT1 and CsPT4 were also codon-optimized and edited to remove the thylakoid membrane targeting tag for inclusion into the system.

Determination of vector cassettes and in silico construction

Incorporation of all enzymes into *in vivo* expression systems requires the use of several vector cassettes to harbour the sequences coding for the enzymes. Compatible dual expression cassettes were identified for both *E. coli* and *S. cerevisiae* systems. For the *E. coli* system, vectors pCOLA-Duet1 and pCDF-Duet1 were selected. pCOLA and pCDF are part of the Duet vector system (Novagen) designed for heterologous co-expression of multiple genes in *E. coli.* [138] Each vector encodes two multiple cloning sites (MCS) each of which is preceded by a T7 promoter, *lac* operator, and ribosome binding site. pCDF carries the pCloDF13 replicon, *lacI* gene and streptomycin/spectinomycin resistance marker. pCOLA carries the COLA replicon from ColA(1) as well as antibiotic resistance to kanamycin. An additional cassette used to incorporate *BjMCS*. For this, pLATE52 was used which carries an ampicillin resistance gene. For the *S. cerevisiae* expression system, pESC-HIS and pCEV-G4-KM were used. pESC-HIS contains an auxotrophic selectable marker gene (HIS3) to select and maintain the expression vector in yeast cells as well as a pBluescript II SK(+) phagemid origin of replication. pCEV-G4-KM contains an antibiotic resistance for G418 and a yeast 2μ origin of replication.

Plasmids for sequence amplification, protein expression and *in vivo* OA-expression system were first modeled using VectorNTI software (Vector NTI Software, ThermoFisher Scientific). Vector maps for plasmids were acquired from Addgene, and primer or adaptor sequences containing designated restriction sites were designed with Primer3Web (https://primer3.ut.ee/). These models were used to confirm that sequences of interest were properly incorporated and oriented in the plasmids, as well as to check for presence of any restriction sites that could interfere with cloning strategies. Restriction enzymes for each plasmid were chosen such that they only have one singular cut site, meaning that the insert will only have one possible site to fit into.

PCR amplification of acquired gene sequences

Codon-optimized genes were amplified by PCR to create a single DNA amplicon in a reaction mixture consisting of 0.2 mM forward and reverse primers, 0.2 mM GeneBlock template, 0.2 mM dNTP's, 10 µL 5X Phusion buffer (ThermoFisher) for final concentration of 1X buffer, and 1 Unit Phusion polymerase (ThermoFisher) in a final volume of 50 µL. PCR techniques throughout this experimentation were either standard TAQ or PHUSION thermal cycling, or Touchdown PCR. PHUSION polymerases were used for amplification and creation of all sequences of interest, while TAQ polymerase was used for insert confirmation and colony PCR. Annealing temperatures used were three degrees higher than the lower of the two melting temperatures of the respective primers (Table 1). Touchdown PCR, starting from a temperature 10 degrees higher than the annealing temperature, was used when standard thermal cycling generated more than one PCR product. Gel electrophoresis confirmed amplification of the geneblock by visualization of amplicons of the correct approximate size. PCR products were separated on a 1.25% agarose gel and compared to a 1 kb+ DNA ladder (ThermoFisher).

Agarose gels were made by dissolving 0.625 g agarose in 50 mL sodium borate (SB, 2.5 mM NaOH, 50 mM boric acid, pH 8.5) buffer facilitated by microwave heating. One µL of Invitrogen SYBR Safe DNA Gel Stain was added to the molten liquid mixture as a pre-stain. All agarose gels were submerged in SB buffer and separated at 90 V for approximately 40 minutes. For visualization, the gel was illuminated with blue light imaging (BLooKTM LED Transilluminator - GeneDireX, Inc.) to determine the size and purity of the PCR product(s).

Success of the PCR was determined according to the presence of bands corresponding to the correct number of base pairs found in the geneblock sequence. If additional PCR products were observed, indicating off-target amplification due to primer interaction or misfiring, the DNA amplicon of the proper size was gel extracted and purified using a GeneJet extraction kit (ThermoFisher). Briefly, the remaining PCR products from the PCR reaction was separated on a 1.00% agarose gel and imaged using blue light. Bands of the appropriate size were excised from the gel. The concentration of the recovered DNA was measured using a NanoDrop™ Spectrophotometer (Thermo Fisher).

Table 1. Primers used in construction of synthetic biology systems

Sequence optimization denoted beside Target sequence. Restriction Enzyme (RE), Melting temperature (Tm), *S. cerevisiae* (*S. cer*).

Primer Sequence	Target	Tm °C	Purpose
GGATCCATGGCGGTCAAACACTTAATC	CsOAC (E. coli)	64.33	Clone into pCOLA, BamHI RE site
CTGCAGCTACTTTCGTGGTGTGTAGTC	CsOAC (E. coli)	65.52	Clone into pCOLA, PstI RE site
GAATTCATGGCTGTTAAGCATCTTATAGTG	CsOAC ((S. cer)	61.16	Clone into pESC-HIS, EcoRI RE site
GAGCTCCTACTTCCTGGGTGTG	CsOAC (S. cer)	62.05	Clone into pESC-HIS, SacI RE site
AGATCTATGAACCACCTTCGCGCC	CsTKS (E. coli)	64.83	Clone into pCOLA, BglII RE site
CTCGAGTTAGTATTTAATTGGAACGCTACGTACC	CsTKS (E. coli)	65.19	Clone into pCOLA, XhoI RE site
GGATCCATGAATCACTTACGTGC	CsTKS (S. cer)	59.75	Clone into pESC-HIS, BamHI RE site
GTCGACTTAGTACTTGATCGGGAC	CsTKS (S. cer)	60.5	Clone into pESC-HIS, SalI RE site
AGATCTATGATCCCTCAGCCAAACGTTGC	PoPKS (E. coli)	66.66	Clone into pCOLA, BglII RE site
CTCGAGTTAATTACTCGTGATGGGACAGGAGC	PoPKS (E. coli)	67.77	Clone into pCOLA, XhoI RE site

GGATCCATGATTCCCCAGCC	PoPKS (S. cer)	60.25	Clone into pESC-HIS, BamHI RE site
GTCGACTTAGTTGGACGTAATCG	PoPKS (S. cer)	59.22	Clone into pESC-HIS, SalI RE site
GAATTCATGGGTAAGAATTACAAGTCCCTGGAC	CsAAE1 (E. coli)	65.33	Clone into pCDF, EcoRI RE site
GCGGCCGCTCATTCAAAGTGAGAAAATTG	CsAAE1 (E. coli)	67.49	Clone into pCDF, NotI RE site
ACTAGTATGGGTAAGAATTACAAGTCCC	CsAAE1 (S. cer)	55.8	Clone into pCEV-G4-KM, SpeI RE site
TTAATTAATCATTCAAAGTGAGAAAAT	CsAAE1 (S. cer)	52.3	Clone into pCEV-G4-KM, PacI RE site
AGATCTATGGCAACCTCAGACGG	CsCHI-L (E. coli)	61.81	Clone into pCDF, BglII RE site
CTCGAGTTAGAACAGTGACAGCATAGATT	CsCHI-L (E. coli)	62.76	Clone into pCDF, XhoI RE site
CCCGGGATGGCAACCTCAGACGG	CsCHI-L (S. cer)	60	Clone into pCEV-G4-KM, SmaI RE site
GGTACCTTAGAACAGGGATAACATTGACTGCC	CsCHI-L (S. cer)	61.1	Clone into pCEV-G4-KM, KpnI RE site
ATGACAACGAATCAAACGGAA	CsPT1 (E. coli)	60.4	Amplify CsPT1
GATAAAAACGTAGACTAGATACTCCGC	CsPT1 (E. coli)	59.3	Amplify CsPT1
ATGTCTGATCAAATTGAGGGAAG	CsPT4 (E. coli)	59.5	Amplify CsPT4
AAATACATATACGAAGTATTCGGCAT	CsPT4 (E. coli)	58.5	Amplify CsPT4
CGCCCTTTAGTGAGGGTTGCTACTTCCTGGGTGTGTAATC	CsOAC (S. cer)	60.6	Add overlap (Gibson)
GTAAGAATTTTTGAAAATTCGATGGCTGTTAAGCATCTTATAG	CsOAC (S. cer)	58.4	Add overlap (Gibson)
CTCTGGCGAAGAATTGTTAATTAAGAGCTCTACTTCCTGGGTGTG	CsOAC (S. cer)	64.9	Add overlap (Gibson)
GACGACGATAAGATCTGAGCTATGGCTGTTAAGCATCTTATAG	CsOAC (S. cer)	58.4	Add overlap (Gibson)
GGTTGGGAATTGCAAATGAACCGCGCCGCC	BjMCS (E. coli)	60.57	Add adaptor to clone into pLATE52
GGAGATGGGAAGTCATTATTTCTTCGCGTAAATATCCTTGTAGGT	BjMCS (E. coli)	60.85	Add adaptor to clone into pLATE52

Initial cloning of sequences into an entry vector

In traditional restriction enzyme-based cloning, it is possible to include the proper restriction sites on the primers used to produce a PCR amplicon. Once amplified, the PCR fragment can be cut with the appropriate set of enzymes and then directly cloned into an expression vector. However, this approach often does not work due to many restriction enzymes having difficulty cutting at the ends of DNA fragments. Thus, I first cloned the codon-optimized genes into the pJET1.2 vector (CloneJet, Thermofisher) to both verify the correct sequence of the cloned PCR products, and to generate sufficient quantities of DNA for subsequent use in generating the final expression vectors. A reaction mix comprised of 10 µL reaction buffer, 3 µL purified PCR product, 1 µL pJET plasmid, 1 µL T4 DNA ligase and 5 µL H20 was combined

and incubated briefly at room temperature (23 °C) for five minutes. Products of the optimized CloneJet reaction were transformed into chemically-competent *E. coli* Mach1 cells, resulting in *E. coli* colonies carrying independent vectors which may or may not harbour the insert. Chemical transformation was used to facilitate the successful uptake of pJET1.2+insert vectors by the Mach1 cells. Chemical transformation refers a period of heat shock during which the cell wall is more permeable for plasmid uptake. Chemical-competent Mach1 cells were thawed on ice and 2 µL of the CloneJet reaction was added to the cells and tapped lightly to facilitate mixing. The cells were left on ice for approximately 30 minutes before being subjected to a 30-second incubation at 42 °C. Following incubation, cells were placed briefly back on ice after which 975 µL of 37 °C Super Optimal broth with Catabolite repression (SOC) media was added to the cells. Mach1 cells were then shaken for 30 minutes at 37 °C at 220 rpm in the SOC medium. Three aliquots from the Mach1 solutions (10 µL, 100 µL and the remainder) were spread using glass beads on pre-warmed lysogeny broth (LB) agar plates supplemented with 50 µg mL^{-1} ampicillin (AMP). The plates were then left to incubate at 37 °C overnight. Individual colonies were harvested the following morning and grown in five mL of liquid LB+AMP media (again while being shaken at 37 °C/220 rpm) for approximately six hours.

Plasmid insert and verification

All pJET1.2 plasmids potentially harboring inserts were subjected to the same mini prep protocol and restriction digest to isolate independent vectors and verify the presence of insert, respectively. Plasmids from each independent culture were isolated using a GeneJet Miniprep Kit (ThermoFisher). From each culture, 1 mL aliquots were removed and stored at 4 °C to allow for later re-growth if necessary. The remaining cultures were centrifuged at 12,000 *g* for 10 minutes to pellet the bacterial cells. Following centrifugation, supernatant was decanted and the

pelleted cells re-suspended in 250 µL of Resuspension Solution by mixing with a pipette. Resuspended cells were transferred to a 1.5 mL centrifugation tube and 250 µL of Lysis solution was added to facilitate break down of cellular membranes and release bacterial proteins, followed shortly by 350 µL of Neutralization Solution. This process resulted in the lysis of bacterial cells and plasmid release without degradation of the plasmids themselves. The tube was centrifuged for five minutes to pellet the membranes and other various cell structural components. Supernatant was removed and applied to the GeneJET spin column (Thermo Fisher) and centrifuged for another minute. After flow-through had been discarded, the column was treated with 500 µL wash buffer followed by an additional centrifugation. All flow-through was once again discarded. Finally, the column was treated with 25 µL of elution buffer and centrifuged once more. The final plasmid-containing flow-through was collected.

To determine the presence of inserts in these vectors, 5 µL of the plasmid was digested using restriction enzymes in a solution with buffer and loading dye for a final reaction volume of 20 µL to cut out the sequence of interest from the plasmids. Restriction enzyme combinations for individual cloning sites and sequences can be found in Table 2. Approximately one unit of each enzyme was used per 1000 ng of DNA. Following a 5 minute incubation at 37 °C, digested plasmids were loaded onto a 1.25% agarose gel and ran alongside a 1kb+ ladder (Invitrogen) for determination of insert presence. Plasmids with inserts of the proper size were sent for Sanger sequencing at the MICB DNA Sequencing Service (University of Manitoba) to verify the identity of the insert and ensure that no PCR-induced errors were present in their sequences.

Attempted cloning of CsPT1 and CsPT4

Sequences coding for the cannabis prenyltransferase, active in the CBGA-forming step of the cannabinoid biosynthetic pathway, were also acquired in the manner outlined above and it was attempted to include these sequences in the *in vivo* expression system. These genes were edited for removal of the thylakoid-targeting tag, resulting in alteration of the sequence.

Insertion of sequences of interest into HIS-tag protein expression plasmids

To facilitate isolated production of each enzyme used, sequences coding for the GOIs were introduced into pLATE52 expression cassettes. pLATE52 plasmids are designed for bacterial protein expression and include a WELQut-cleavable N-terminal 6xHis tag for protein purification. Purified PCR products of each *E. coli* codon optimized sequence (*CsAAE1*, *CsTKS*, *CsOAC* and *CsCHI-L*) were cloned into individual pLATE52 vectors. pLATE52 vectors were acquired as part of the aLICator kit (ThermoFisher). All DNA amplicons were inserted into the pLATE52 vector using ligation-independent cloning (LIC). Amplicons were prepared for insert into the vector by treatment with T4 DNA polymerase in a 10 µL reaction consisting of 2 µL 5X LIC buffer, 1 µL purified PCR product, 1 µL T4 DNA polymerase, and 6 µL nuclease-free water. The reaction was incubated at room temperature for five minutes before being stopped by the addition of 0.6 µL 0.5M EDTA and mixed thoroughly. The annealing reaction was then instigated by the addition of 1 µL pLATE52 LIC-ready vector. The annealing mixture was incubated for approximately five minutes. Bacterial electrical transformation of BL21 *E. coli* cells was performed immediately following LIC procedure (transformation procedure outlined below).

Protein expression

To verify proper gene incorporation and protein expression, all proteins were expressed and collected with HIS-tag purification for subsequent experimentation. Individual proteins were expressed from their respective pLATE52 vectors. Cultures were started in 50 mL of LB supplemented with antibiotics. Following overnight growth in 50 mL, cultures were added to 500 mL of LB with no antibiotics for growth and protein expression. When 500 mL cultures achieved an OD_{600} of 0.6-0.8, protein expression was induced with 0.5 mM IPTG. Cultures were allowed to express for 16 hours. Cultures were then spun down at 12,000 *g* to pellet cells. Cells were then frozen overnight. To extract soluble proteins from the cells, the cellular pellet was resuspended in 10 mL of LB. The cells were then subjected to a 10:10 sonication:rest pulse for three minutes at a pulse strength of 65%. Lysate was then centrifuged at 12,000 *g* to pellet cellular debris. TALON affinity resin (90-150 µL) was washed thrice with Wash Buffer (50 mM Tris pH 7, 500 mM NaCl, 2.5 mM imidazole, 25% glycerol v/v, 10 mM β-mercaptoethanol; WB). Lysate was then added to TALON in falcon tube and allowed to agitate on shaker at 10 ˚C for 20 minutes. The mixture was spun down at 12,000 *g* to pellet the TALON-protein complex. TALON-protein complex was then washed and spun down again thrice with WB. TALON-protein complex was then applied to gravity column and eluted by addition of 2.5 mL Elution Buffer (50 mM Tris pH 7, 500 mM NaCl, 250 mM imidazole, 5% v/v glycerol, 10 mM β-mercaptoethanol; EB). Protein elution was then applied to a PD-10 column, previously equilibrated by slow addition of five column volumes of Storage Buffer (2 mM TRIS pH 7 in 50% v/v glycerol, SB). Final protein elution from PD-10 column was achieved by one final addition of a column volume of SB. Proteins were then either stored at -20 ˚C in storage buffer, or divided into 60 µL aliquots and preserved with liquid nitrogen for long-term storage in -80 ˚C.

In vitro assays

Reaction assays were run using a combination of all proteins present in our synthetic biology system. Relevant protein concentrations were determined using Protein820 on NanoDrop. Control assays were run using only CsTKS and CsOAC enzymes, where experimental assays were run using CsTKS/CsOAC/CsAAE1, CsTKS/CsOAC/CsCHI-L, and CsTKS/CsOAC/CsAAE1/CsCHI-L. Assay reactions consisted of 0.4 mM Hex-CoA, 1.2 mM Mal-CoA, 25 mM ADA 6.5 buffer combined with ~1.5 µg of each protein, to a final volume of 25 µL. 50% glycerol was used to make up for volume in reactions with less than four proteins used. Assays were allowed to react overnight at room temperature. Stopped-time assays were run using same reaction mixture listed above, where assays were allowed to run for 10 minutes before they were halted with the addition of 1 drop HCl to denature proteins. 1 µL of 10 µM ($1.0e^{-11}$ moles) Cinnamic acid was added to each assay as an internal standard before analysis on HPLC. Assays were extracted for two rounds with equal-volume EtAc. During each round of extraction, EtAc was mixed thoroughly with assay components using a vortex speed mixer and centrifuged for 5 minutes to achieve separation of aqueous and organic layers. Organic layers were carefully transferred to final collection 1.5 mL centrifugal tubes. The collected EtAc extracts were then dried using a speed Vacufuge at 45 °C before resuspension in 50:50 MeOH:Water and transferred to high-performance liquid chromatography-ultraviolet (HPLC-UV) sample vials for analysis.

HPLC-UV analysis of in vitro assay samples

Samples extracted from *in vitro* experimentation were analyzed on an Ultimate 3000 HPLC system with UV detection (Dionex UltiMate 3000 diode array detector, Thermo Fisher Scientific) using a Phenomenex Luna Omega C18 column (bead size 1.6uM, column 50 x 2.1

mm) and Luna Omega C18 guard column. A dual solvent system was used for the separation of sample compounds, where acetonitrile facilitated organic solvent separation and a 0.01% formic acid (v/v; prepared with nuclease-free water) solution comprised the polar solvent. Separation method(s) were developed using the Chromeleon7 software (ThermoFisher).

Samples retrieved from in vitro assays were run starting with a flow rate of 0.3 mL/minute with a gradient of 65% ACN dropping to 49% ACN over the course of 6 minutes, with 49% ACN to 0.5% ACN over 1.5 minutes and 0.5% ACN back to 65% ACN for the next 1.5 minutes.

Compounds detected in samples were quantified by comparison with internal standard (1 µL 10 mM cinnamic acid) where the largest integration value of cinnamic acid was used to create an adjustment factor to normalize the peak integration values of OA, HTAL and PDAL present in samples (Equation 1).

Equation 1. Calculation of adjustment factor for samples by comparison with known internal standard

$$\boldsymbol{adjustment\ factor} = \frac{\boldsymbol{internal\ standard\ value\ in\ sample}}{\boldsymbol{largest\ internal\ standard\ value\ in\ all\ samples}}$$

$$\boldsymbol{compound\ in\ sample} = \boldsymbol{peak\ integration\ value} * \boldsymbol{adjustment\ factor}$$

Statistical analysis of data

Data sets were subjected to statistical analysis to determine significance of results. The Shapiro-Wilk test was used to determine normalcy of the data distribution before subjected to further parametric measures. One-way ANOVA was used for all data sets, corrected for unequal standard deviation with Brown-Forsythe and Welch and supplemented with Dunnet's multiple comparison test for individual variance. Analysis was run using GraphPad (PRISM) software.

Cloning into specialized expression vector for E. coli expression

In general, the enzyme-encoding genes were excised from the pJET1.2 vectors via restriction digest and subsequently gel extracted in preparation for transformation into the appropriate expression vector. pJET1.2 vectors housed all appropriate *E. coli* optimized genes with flanking restriction sites. These restriction sites were added by previous PCR amplification with PCR primers, and the same combination of RE's can be used to excise the genes of interest in both the pJET1.2 and in their final destination vectors (Table 1, Table 2). Restriction enzymes are as follows: pJET1.2+*CsOAC* with BamHI and PstI, pJET1.2+*CsTKS* with BglII and XhoI, pJET1.2+*PoPKS* with BglII and XhoI, pJET1.2+*CsAAE1* with EcoRI and NotI and pJET1.2+*CsCHI-L* with BglII and XhoI (Table 2). The pCDF vector contains multiple expression sites and thus was constructed in a sequential manner, incorporating first *CsAAE1* to produce pCDF+*CsAAE1*, and then adding *CsCHI-L* to produce pCDF+*CsAAE1*+*CsCHI-L* (Table 2). These two vectors served different purposes: the one containing both the *CsCHI-L* and *CsAAE1* sequences was used to test *CsCHI-L* function, whereas the other containing solely the *CsAAE1* sequence served as a control.

To prepare the pCDF expression vector for cloning, a restriction digest was performed on pCDF using the same restriction enzymes (EcoRI and NotI) used to excise *CsAAE1* from the pJET1.2 vector (Table 2). The pCDF restriction digest reaction was then run on an agarose gel and purified via gel extraction. The restriction digested pCDF and the gel-purified *E. coli* optimized *CsAAE1* were ligated together consisting of a 5:1 Insert:Vector ratio, as calculated using the BioNEB calculator and concentrations retrieved from the nanodrop, 1 U of T4 ligase, and 2 µL of T4 ligase buffer in a final reaction volume of 20 µL This reaction was incubated at -4 °C overnight. After incubation the reaction mixture was heat inactivated and transformed into Mach1 cells. Transformations were plated, and individual colonies were picked and grown in

liquid LB media as previously described. After purification by miniprep, the isolated plasmids were restriction digested with EcoRI and NotI to confirm the presence of the *CsAAE1* insert. The newly produced pCDF+*CsAAE1* was then restriction digested with BglII and XhoI in preparation for cloning *CsCHI-L* into the second cloning site (Table 2).The above process was repeated; where restriction digested and purified pCDF+*CsAAE1* and *CsCHI* were combined together in a ligation reaction, heat inactivated and transformed into Mach1 cells. Presence of both *CsAAE1* and *CsCHI* inserts were confirmed by a restriction digest and PCR of the vector. Samples were sent for sequencing, which later confirmed the presence of the proper insert(s) with no PCR-induced mutations.

Prior to this work Kyle MacDonald had already created pCOLA-Duet1 *E. coli* expression plasmid harbouring the *CsTKS* and *CsOAC* sequence inserts. Additionally, experimentation was conducted using an alternate polyketide synthase, PoPKS from *Primula obconica*. As CsTKS and PoPKS are proposed to serve the same function in the biosynthetic pathway, two separate dual expression cassettes were created: pCOLA+*CsTKC*+*CsOAC* and pCOLA+*PoPKS*+*CsOAC* (Table 2). These pCOLA plasmids alongside the pCDF plasmids make up most of our *E. coli* expression system.

Also included in the *E. coli* microbial expression system is a malonyl-CoA synthetase (MCS) from *Bradyrhizobium japonicum* that allows for higher internal production of malonyl-CoA. Incorporation of the BjMCS enzyme into microbial expression systems for higher internal production of malonyl-CoA has been successfully demonstrated[121]. *BjMCS* was cloned into PLATE52 using ligation independent cloning (LIC) and carried out using components and protocol from the aLICator kit (ThermoFisher) (Table 2). Briefly, 5 µL PCR amplified and purified BjMCS was combined with 2 µL LIC buffer, 2 µL H2O and 1 µL (I Unit) T4 DNA

Polymerase. This reaction was incubated at 23 °C for five minutes. 0.6 µL EDTA was then added, mixed thoroughly via pipetting, and 1.0 µL pLATE52 vector added. Mixture was incubated for another five minutes at 23 °C before being directly used for cellular transformation.

Table 2. Plasmids used in synthetic biology systems

All expression cassettes used in final construction of *E. coli* and *S. cerevisiae* synthetic biology systems for incorporation of relevant genes of interes (CsOAC, CsTKS, PoPKS, CsAAE1, BjMCS and CsCHI-L). Multiple cloning site (MCS), Antibiotic (Anti.) origin of replication (Ori), *S. cerevisiae* (*S. cer*).

Plasmid	MCS1	MCS2	RE	Anti.	Ori
pCOLA+CsOAC (*E. coli* opt)	CsOAC		BamHI, PstI	KAN	ColA
pCOLA+CsTKS (*E. coli* opt)		CsTKS	BglII, XhoI	KAN	ColA
pCOLA+CsOAC+CsTKS (*E. coli* opt)	CsOAC	CsTKS	BamHI+PstI; BglII+XhoI	KAN	ColA
pCOLA+PoPKS (*E. coli* opt)		PoPKS	BglII, XhoI	KAN	ColA
pCOLA+CsOAC+PoPKS (*E. coli* opt)	CsOAC	PoPKS	BamHI+PstI; BglII+XhoI	KAN	ColA
pCDF+CsAAE1(*E. coli* opt)	CsAAE1		EcoRI, NotI	STREP	CloDF13
pCDF+CsAAE1+CsCHI-L (*E. coli* opt)	CsAAE1	CsCHI-L	EcoRI+NotI; BglII+XhoI	STREP	CloDF13
pESC-HIS+CsOAC (*S. cer* opt)	CsOAC		EcoRI, SacI	AMP	F1
pESC-HIS+CsTKS (*S. cer* opt)		CsTKS	BamHI, SalI	AMP	F1
pESC-HIS+CsOAC+CsTKS *(S. cer* opt)	CsOAC	CsTKS	EcoRI+SacI; BamHI+SalI	AMP	F1
pESC-HIS+PoPKS (*S. cer* opt)		PoPKS	BamHI, SalI	AMP	F1
pESC-HIS+CsOAC+PoPKS (*S. cer* opt)	CsOAC	PoPKS	EcoRI+SacI; BamHI+SalI	AMP	F1
pCEV-G4-KM+CsAAE1 (*S. cer* opt)	CsAAE1		SpeI, PacI	G418	pBR322
pCEV-G4-KM+CsAAE1+CsCHI-L (*S. cer* opt)	CsAAE1	CsCHI-L	SpeI+PacI; SmaI+KpnI	G418	pBR322
pLATE52+BjMCS (*E. coli* opt)	BjMCS	N/A	N/A	AMP	ColE1
pLATE52+CsAAE1 (*E. coli* opt)	CsAAE1	N/A	EcoRI, NotI	AMP	ColE1
pLATE52+CsTKS (*E. coli* opt)	CsTKS	N/A	BglII, XhoI	AMP	ColE1
pLATE52+CsOAC (*E. coli* opt)	CsOAC	N/A	BamHI, PstI	AMP	ColE1
pLATE52+CsCHI-L (*E. coli* opt)	CsCHI-L	N/A	BglII, XhoI	AMP	ColE1
pJET1.2+CsOAC (*E. coli* opt)	CsOAC	N/A	BamHI, PstI	AMP	ColE1
pJET1.2+CsTKS (*E. coli* opt)	CsTKS	N/A	BglII, XhoI	AMP	ColE1
pJET1.2+PoPKS (*E. coli* opt)	PoPKS	N/A	BglII, XhoI	AMP	ColE1
pJET1.2+CsAAE1 (*E. coli* opt)	CsAAE1	N/A	EcoRI, NotI	AMP	ColE1
pJET1.2+CsCHI-L (*E. coli* opt)	CsCHI-L	N/A	BglII, XhoI	AMP	ColE1
pJET1.2+CsOAC (*S. cer* opt)	CsOAC	N/A	EcoRI, SacI	AMP	ColE1
pJET1.2+CsTKS (*S. cer* opt)	CsTKS	N/A	BamHI, SalI	AMP	ColE1

pJET1.2+PoPKS (*S. cer* opt)	PoPKS	N/A	BamHI, SalI	AMP	ColE1
pJET1.2+CsAAE1 (*S. cer* opt)	CsAAE1	N/A	SpeI, PacI	AMP	ColE1
pJET1.2+CsCHI-L (*S. cer* opt)	CsCHI-L	N/A	SmaI, KpnI	AMP	ColE1

S. cerevisiae expression construct construction

A very similar process was used to create the expression system for *S. cerevisiae*. Although the eventual host of the *S. cerevisiae* plasmids is, of course, yeast, the initial cassettes were constructed using *E. coli* cells and transformation. The following pJET1.2 vectors containing the *S. cerevisiae* optimized versions of the gene sequences and added restriction enzyme sites were created: pJET1.2+*CsOAC* with EcoRI+SacI, pJET1.2+*CsTKS* with BamHI+SalI, pJET1.2+*CsAAE1* with SpeI+PacI and pJET1.2+CsCHI-L with SmaI and KpnI (Table 2). After verification of the insert sequences by Sanger sequencing, I performed restriction digestions to cut out the sequence inserts, and said inserts were gel extracted and purified. pESC-HIS and pCEV-G4-KM vectors were also restriction digested with restriction enzymes for corresponding cloning sites, gel extracted and purified such that their cloning sites were "open" for ligation .Restriction enzymes used for pESC-HIS include EcoRI and SacI at MCS1 and BamHI and SalI at MCS2, where restriction enzymes for pCEV-G4-KM were SpeI and PacI at MCS1 and SmaI and KpnI at MCS2 (Table 2). These vectors were also sequentially constructed, with the *CsOAC* inserts being cloned in first to the pESC-HIS vectors, and *CsAAE1* cloned in first for the pCEV-G4-KM vectors. Following confirmation of the appropriate *CsOAC*/*CsAAE1* presence in these vectors, they were once again digested open and ligated with the second insert at the second cloning site – *PoPKS*/*CsTKS* for pESC-HIS and *CsCHI* for pCEV-G4-KM (Table 2). Diagnostic restriction digests and PCRs were performed to confirm the presence of the sequences in these vectors.

Gibson assembly for plasmid cassette creation

I initially was not successful in producing OA in the yeast system, and I subsequently found that the freezer stock of the pESC-HIS+*CsOAC*+*CsTKS* was in fact lacking the *CsOAC* and *CsTKS* genes. I thus had to remake this vector and chose to use Gibson assembly to re-introduce the CsOAC sequence into the expression vector. pESC-HIS+*CsTKS* was linearized with EcoR1 which cut in the multi cloning site. Primers were designed to amplify the *CsOAC S. cerevisiae* optimized insert with 40 base pair 5' extensions that are complimentary to the EcoR1-digested linear vector ends (Table 1). The insert was amplified using these primers using Touchdown PCR produce the *CsOAC* amplicon for Gibson assembly. The cut backbone and the insert with extensions were then gel and PCR purified with ThermoFisher purification kits (protocol outlined previously) and measured on the NanoDrop to ascertain purity and DNA concentration. 80 ng (0.016 pmoles) of pESC-HIS+CsTKS vector and 10 ng (0.048 pmoles) of CsOAC insert were then combined with NEB Gibson Master Mix and incubated at 50 ˚C for 15 minutes to allow the fragments to assemble. The re-created pESC-HIS+*CsOAC*+*CsTKS* was verified by Sanger sequencing.

Transformation of E. coli bacterial cells

E. coli expression plasmids were transformed into BL21 cells, an *E. coli* strain used for high-level recombinant protein production, by electroportation. Electrocompetent BL21 cells were produced in the laboratory by inoculation of 500 mL LB media with an overnight culture and growth to an OD_{600} of 0.4-0.6. Cells were harvested by centrifugation and washed repeatedly with 10% glycerol before a final resuspension in 1mL 10% glycerol. Aliquots of 60 µL of the cells were stored at -80 °C. An aliquot of 0.5 µL of each plasmid (~100ng DNA) was combined on ice with electro-competent BL21 *E. coli* cells for five minutes after which the mixture was

pipetted into electroporation cuvettes. 1.8 kV shocks were applied to the reaction mixture for 4 ms. The cells were then shaken in 975 μL of pre-warmed SOC media for one hour at 37 °C with shaking at 220 rpm. Cultures were then plated onto selection plates 50mM ampicillin (AMP), 50 mM spectinomycin (SPEC) and 30 mM kanamycin (KAN). Transformations were initially done to produce the system containing only the pCDF and pCOLA expression cassettes (pCDF+*CsAAE1*+*CsCHI*-L or pCDF+*CsAAE1* and pCOLA+*CsOAC*+*CsTKS*/*PoPK*S) where lack of CsCHI-L in the pCDF vector would serve as a control. Following production of the *BjMCS* pLATE52 expression cassette, triple transformations were performed to produce the final system(s) with pLATE52+*BjMCS*, pCDF+*CsAAE1*+*CsCHI*-L and pCOLA+*CsOAC*+*CsTKS*/*PoPKS* plasmids. These plates were then incubated at 37 ˚C overnight and monitored for growth the next day. Colonies were picked and grown up in LB media with the corresponding antibiotics for further experimentation.

When using the Gibson assembly created plasmid, Gibson mixture was diluted 1:10 in water and 0.5 μL was then used to transform electro potent BL21 cells in the manner described above.

Confirmation of cellular transformations with expression cassettes

Presence of all vectors and their corresponding gene sequences in cultures were confirmed using PCR. Cultures were miniprepped using a commercial kit to obtain plasmids. Plasmids were then combined in PCR reactions with primers specified to amplify each sequence (the same primers used for initial sequence amplification). PCR reactions were then run on the thermocycler and sequence presence checked using an agarose gel. Samples from the miniprepped plasmids were also sent along with respective primers for Sanger sequencing to determine proper insert sequence presence.

Transformation of S. cerevisiae yeast cells

Plasmids were transformed into INVSc1 yeast cells using electroporation and plating on selective media, a similar approach to the construction of the *E. coli* expression system. Electrocompetent INVSc1 cells were produced by inoculation of 500 mL YPD media with INVSc1 overnight culture and grown to an OD_{600} of 1.3-1.5. Cells were harvested via centrifugation and washed with 500mL cold water. This process was repeated with 250 mL water and 20 mL 1 M sorbitol before a final resuspension in 1 mL 1 M sorbitol to yield a highly concentrated amount of yeast cells. Cells were divided into 60 µL aliquots and stored at -80 °C. When ready for use in transformations, cells were thawed on ice and combined with 0.5-1 µL (100-300 ng) of plasmid DNA. The cell and DNA mixtures were incubated on ice for five minutes before transfer to electroporation cuvettes. Cells were then exposed to a ~4-6 ms shock of 1.5 kV. Cells were immediately resuspended in 1 mL ice-cold 1 M sorbitol and allowed recovery growth for 1 hour at 30 ˚C while shaking at 220 rpm. 100 µL of transformed cells were then plated onto selection plates. pESC-HIS plasmids growth on HIS deficient media, whereas pCEV-G4-KM plasmids contain kanamycin/G418 antibiotic resistance. G418 at a concentration of 400uM/mL was used for the INVSc1 selection plates. Sorbitol was incorporated into said plates at a concentration of 1 M to facilitate transformant growth. Plates were allowed to grow for 4-6 days at 30 ˚C while being monitored for culture growth. Resultant colonies were picked and grown in selective/antibiotic media overnight. Due to the selective nature of media use for yeast cells, it is assumed that successful growth of colonies on the plates corresponded to a successful transformation, however colony PCR was performed to confirm plasmid presence.

Selectivity of media and antibiotics with INVSc1 cellular growth

Experimentation was conducted with the different types of selective media used in the transformation of INVSc1 cells with constructed plasmids. This was done to verify the validity of the use of selective media and antibiotics, and to demonstrate proper incorporation of plasmids into host cells following transformation. 100 µL aliquots of electrocompetent INVSc1 cells were plated onto selective dropout (HIS deficient) media supplemented with 400 µM G418. 100 µL of INVSc1 cells transformed with 0.5 µL pESC-HIS and 0.5 µL pCEV-G4-KM plasmids were also plated onto selective dropout (HIS deficient) media supplemented with 400 uM G418 to serve as a control group, demonstrating cells that have properly incorporated the plasmids. Plates were incubated at 30 ˚C for up to 6 days to ensure proper time allowed for development of any *S. cerevisiae* colonies.

Results

Selective media is capable of restricting colony growth

No growth was noted for the non-transformed INVSc1 cells when plated on the His-deficient and antibiotic supplemented plates, whereas transformed cells that had incorporated the appropriate expression plasmids (pCEV-G4-KM and pESC-HIS, as verified by colony PCR), exhibited normal and expected growth on these plates (Figure 5). INVSc1 cells grew normally on a plain agar plate with no selective qualities, demonstrating the true capabilities of using selective media (Figure 5). Use of selective media (LB agar supplemented with 30-50 µM antibiotics) in the *E. coli* system also demonstrated tight regulation of transformed colonies, where growth of cells that had not properly incorporated the required plasmids were unable to

grow (visual representation not shown). Antibiotics used for *E. coli* selection include AMP, SPEC and KAN.

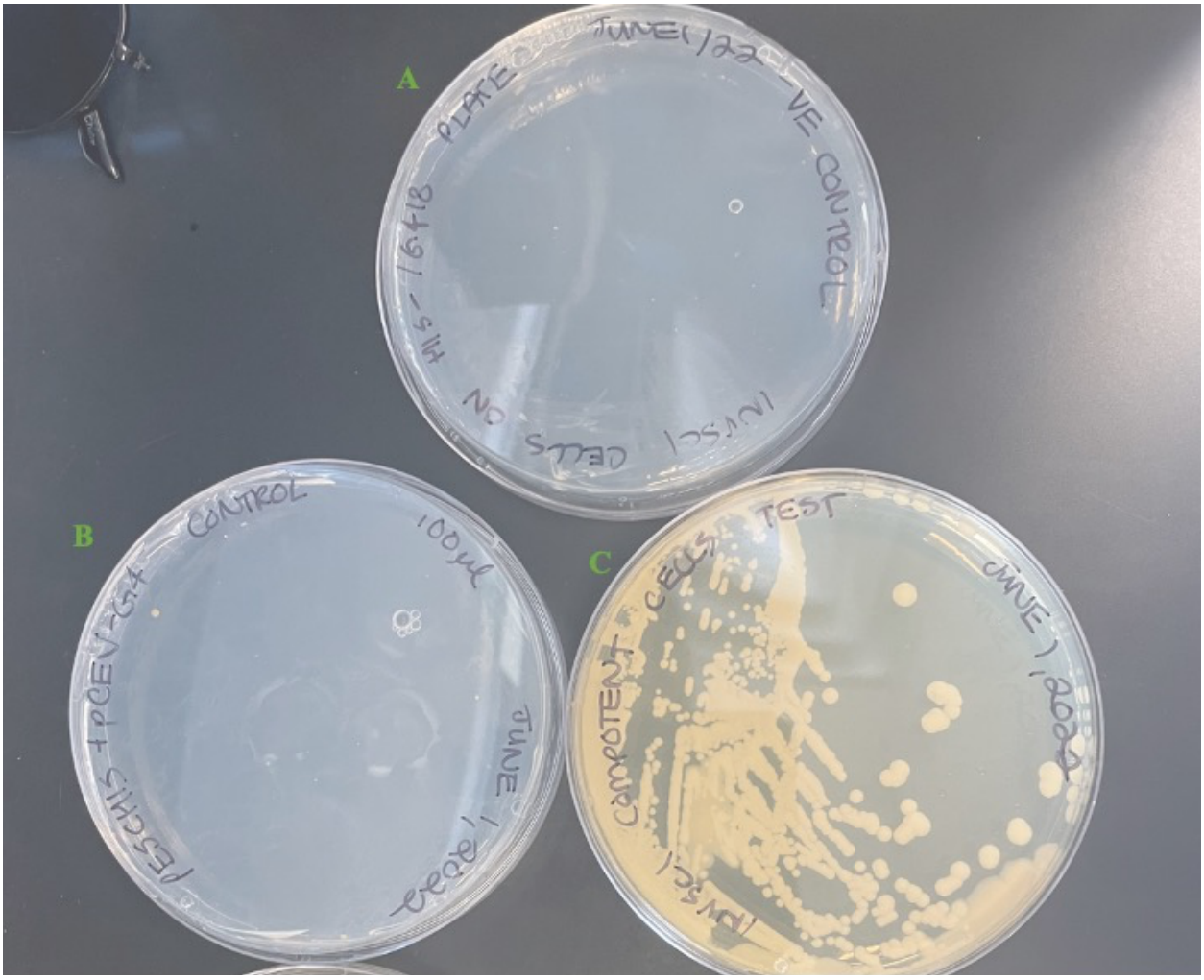


Figure 5. Selective dropout media and antibiotic selection

Demonstration of selectivity of SD and antibiotic supplemented media. A) INVSc1 competent cells plated on selective media (HIS-) supplemented with 400µM G418, no growth noted. B) Colony growth on selective media (HIS-) with 400µM G419 following transformation with pESC-HIS and pCEV-G4-KM plasmids. C) Regular growth of INVSc1 competent cells when plated on YPD agar plate.

CsCHI-L inclusion is not correlated with higher OA production in in vitro assays

In order to test the viability of our GOI's and to start exploring the activity of the CsCHI-L enzyme*, in vitro* assays were conducted. This facet of experimentation was done for the purpose of a.) confirming the OA-producing capabilities of the enzymes, and b.) to test the hypothesis that CsCHI-L may increase the production of OA *in vitro*. Different enzymatic combinations of all proteins (CsAAE1, CsTKS, CsOAC and CsCHI-L) were incorporated into assays with starter substrates hexanoyl-CoA and manolyl-CoA. These starter substrates were chosen for the *in vitro* assays in place of hexanoic and malonic acid due to the inability of the *in vitro* environment to produce malonyl-CoA. CsAAE1 enzymes were still included in these *in vitro* assays (despite not being directly involved in hexanoyl-CoA production from hexanoic acid) to confirm there are no potential negative interactions between it and the other proteins, as well as to monitor for involvement of the CsCHI-L with CsAAE1, where the proposed accessory protein may act as a shuttle for hexanoyl-CoA between CsAAE1 and CsTKS.

Once the GOI's had been incorporated into pLATE52 vectors, protein expression of each enzyme/protein was possible. The proteins were expressed individually in *E. coli* and subsequently purified via HIS-tag purification. The isolated proteins were then separated and visualized using SDS-PAGE to confirm the presence and purity. Following confirmation of successful proteins, the proteins were then added in various combinations with the substrates hexanoyl-CoA and malonyl-CoA in *in vitro* assays. Two sets of assays were run; one with a stopped-time reaction at 10 minutes and another set allowed to react for 24 hours (assumed to be "to completion" for the reaction). Both reactions were run at room temperature. Stopped time assays were halted with the addition of HCl to denature the proteins. Four different protein combinations were run (n=3) to determine the relative activities of the enzyme combinations:

CsTKS/CsOAC, CsTKS/CsOAC/CsAAE1, CsTKS/CsOAC/CsCHI-L and finally CsTKS/CsOAC/CsAAE1/CsCHI-L. These combinations were chosen to monitor any potential increased OA production between the CsCHI-L and both the CsAAE1 protein and CsTKS/CsOAC proteins, as well as to see how all proteins worked together. A cinnamic acid internal standard (1 μL of 10 mM or 1.0e-^{11} moles) was added to each assay following completion to allow for creation of an adjustment factor for potential product losses during the processing of the samples prior to analysis.

Following sample preparation, high performance liquid chromatography-ultraviolet (HPLC-UV) analysis of the samples provided chromatographic data indicating the presence of compounds of interest that were confirmed by retention time and UV-VIS spec (Figure 6). Production of HTAL, PDAL and OA were noted in all assay samples (Figure *7*). The amount of each product in a reaction was quantified using normalization of the signal to the internal standard of cinnamic acid and was then compared to regression analysis of authentic standard curves for each compound (Figure 8).

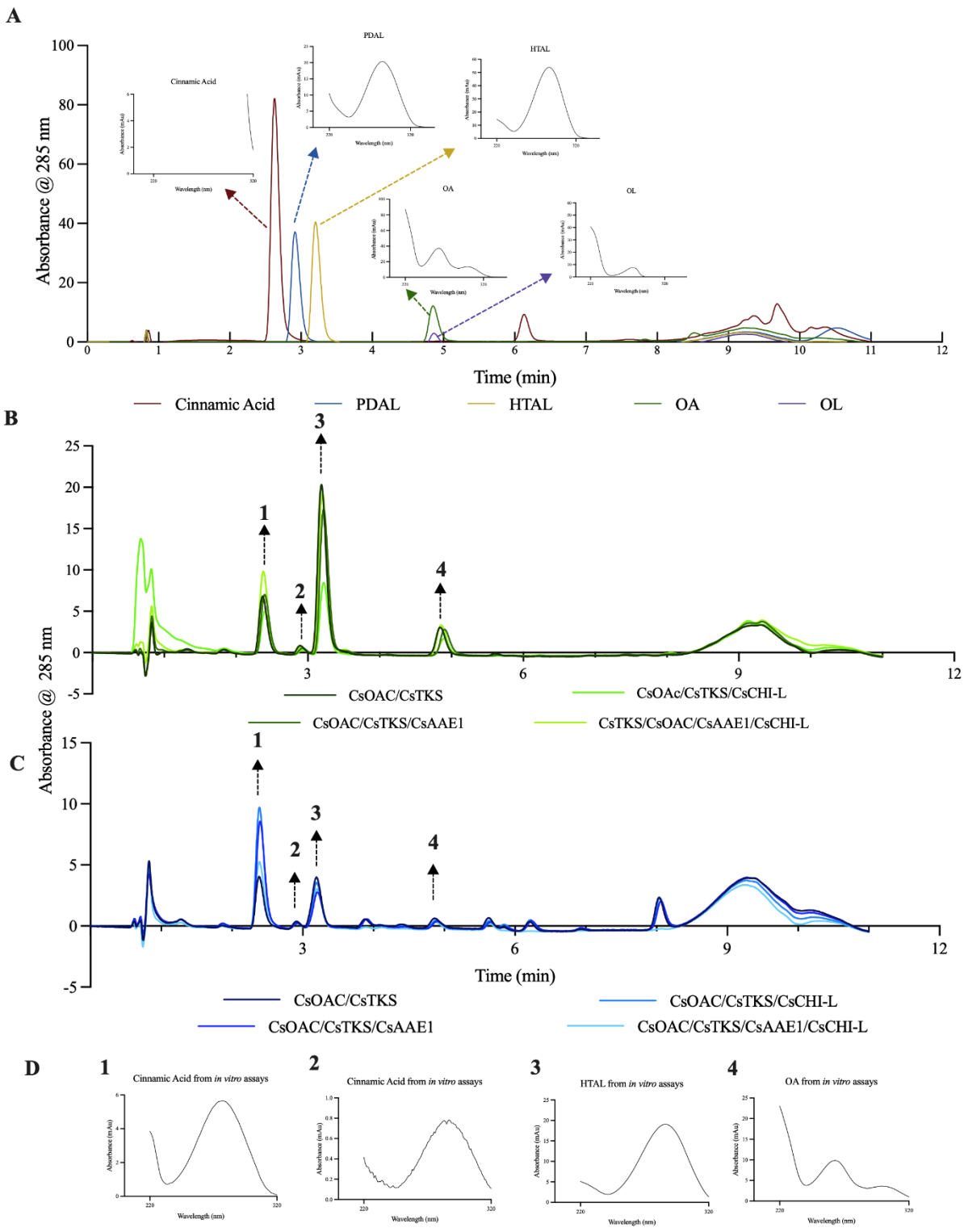


Figure 6. Chromatograms and UV-VIS spectra from *in vitro* assays

A. Chromatograms of authentic standards of compounds of interest, including Cinnamic Acid, hexanoyl triacetic acid lactone (HTAL), pentyl diacetic lactone (PDAL), olivetol (OL) and

olivetolic acid (OA). UV-VIS spectra displayed next to corresponding peak of standard on chromatograph. B. Chromatograms resulting from HPLC-UV analysis of in vitro 10 minute assay samples. C. Chromatograms resulting from HPLC-UV analysis of in vitro 24 hour assay samples. Internal standards of 1.0e-^{11} moles cinnamic acid were added to each assay prior to extraction. D. UV-VIS spectra from in vitro assay chromatograms confirming compound presence.

Data interpretation proved the presence of OA, as well as bi-products HTAL and PDAL, in all assay samples, both at the 10 minute and 24 hour timepoints (Figure 6). No OL was noted in the in vitro assays (Figure 6). Compound amounts were reported following conversion of response data from the HPLC to concentration through standard curve analysis for each compound (Figure 8).

In the 10 minute assays, HTAL and OA accumulation was the highest in the CsOAC/CsTKS/CsAAE1 experimental group, with averages of 71.704±7.689 pmole (HTAL), and 687.246±57.857 pmole (OA; Figure 7). The highest average PDAL accumulation was in the CsOAC/CsTKS/CsAAE1/CsCHI-L experimental group, with a value of 3048.920±126.232 pmoles PDAL (Figure 7). The highest individual HTAL and OA values were detected in the same experimental group, with values of 80.134±3.959 pmole (HTAL) and 746.643±64.122 pmole (OA; Figure 7). The highest individual value for production of PDAL was seen in experimental group CsOAC/CsTKS/CsAAE1/CsCHI-L with a value of 3191.486±126.232 pmoles PDAL (Figure 7). On average, the lowest HTAL production was noted in the CsOAC/CsTKS/CsAAE1/CsCHI-L experimental group with a value 58.981±3.959 pmoles (Figure 7).Lowest average PDAL and OA values were seen in the CsOAC/CsTKS experimental group, with values of 2669.300±16.827 (PDAL) and 552.572±67.877 pmoles (OA; Figure 7). The lowest individual HTAL value was detected in the CsOAC/CsTKS/CsAAE1/CsCHI-L experimental group with a value of 54.972±3.959 pmoles (Figure 7). The lowest individual

PDAL and OA values were seen in the CsOAC/CsTKS experimental group with a value of 2659.489±16.827 pmoles (PDAL) and 483.845±67.877 pmoles (OA; Figure 7).

In the 24 hour samples, both OA and PDAL accumulation was the highest in the CsOAC+CsTKS experimental group, with average compound productions of 700.242±84.877 pmoles (OA) and 3662.787±64.008 pmoles (PDAL; Figure 7). HTAL average production was highest in the CsOAC/CsTKS/CsAAE1/CsCHI-L experimental group with a value of 142.624±26.155 pmoles HTAL (Figure 7). The highest individual values for HTAL was detected in the CsOAC/CsTKS/CsAAE1/CsCHI-L group with a value of 167.940±26.155 pmoles HTAL, where the highest individual PDAL value was produced in the CsOAC/CsTKS/CsAAE1 with a value of 3760.281±311.703 pmoles PDAL. The highest individual OA value was seen in the CsOAC/CsTKS experimental group, with an value of 798.204±84.877 pmoles OA. The lowest average HTAL and OA production was seen in CsOAC+CsTKS+CsAAE1 samples with 126.315±35.225 pmoles (HTAL) and 665.265±31.913 pmoles (OA), whereas the lowest average PDAL production was seen in CsOAC+CsTKS+CsCHI-L with a value of 3328.362±85.599 pmoles (PDAL; Figure 7).The lowest individual values were 87.186±35.225 pmoles of HTAL and 3151.317±311.703 pmoles PDAL in the CsOAC/CsTKS/CsAAE1 expeirmetnal group and 632.853±69.362 pmoles OA in the CsOAC/CsTKS/CsCHI-L experimental group (Figure 7).

There was no significant difference in the amount of OA produced between the CsOAC/CsTKS/CsAAE1/CsCHI-l and the CsOAC/CsTKS/CsAAE1 group (used for comparison as it is the full enzymatic combinations as proven *in planta*). There was no significant difference noticed between these two groups for derailment (HTAL and PDAL) production in either the 24 hour or 10 minute assays. In the 10 minute assays, we can note a statistically significant increased presence of PDAL in the CsTKS/CsOAC/CsAAE1 (p value: 0.0183, 10.00% increase)

and the CsTKS/CsOAC/CsAAE1/CsCHI-L (p value: 0.0330, 14.24 % increase) group when compared to the CsOAC/CsTKS. There is also a statistically significant lower amount of PDAL in 24 hour of CsTKS/CsOAC/CsCHI-L when compared to PDAL production in the CsOAC/CsTKS (p value of 0.0070, 10.07 % increase) and the CsTKS/CsOAC/CsAAE1/CsCHI-L (p value of 0.0204; 8.92% increase) groups.

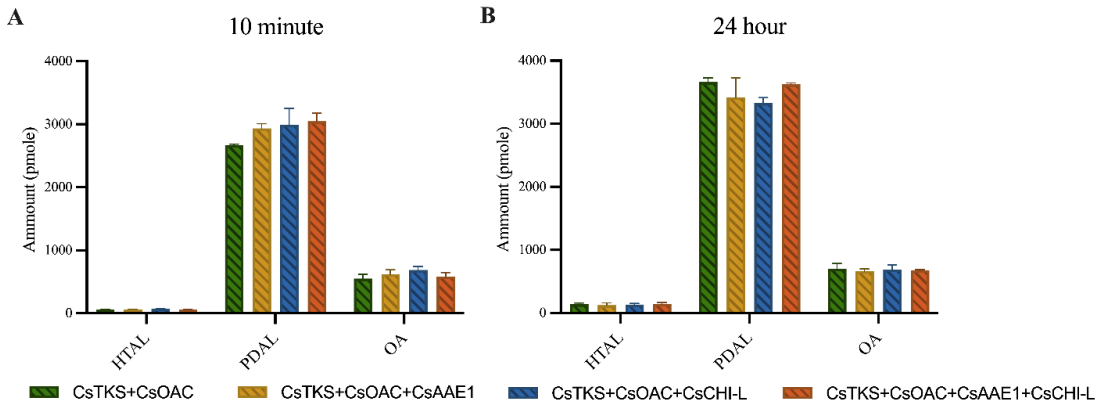


Figure 7. Compound production in *in vitro* assays

CsOAC, CSTKS, CSAAE and CsCHI-L enzymes combined with hexanoyl-CoA and malonyl-CoA. A. 10 minute assays. B. 24 hour assays. Reaction products hexanoyl triacetic acid lactone (HTAL), pentyl diacetic lactone (PDAL) and olivetolic acid (OA). Data subjected to one-way ANOVA (Brown-Forsythe and Welch test; Dunnett's multiple comparisons test), and subsequent individual t-test with Welch's correction. $n = 4$, error bars: CI = 95 %, p value ≤ 0.05. Enzymes and proteins: Cannabis sativa olivetolic acid cyclase (CsOAC), Cannabis sativa tetraketide synthase (CsTKS), Cannabis sativa acyl-activating enzyme (CsAAE1) and Cannabis sativa chalcone isomerase-like (CsCHI-L).

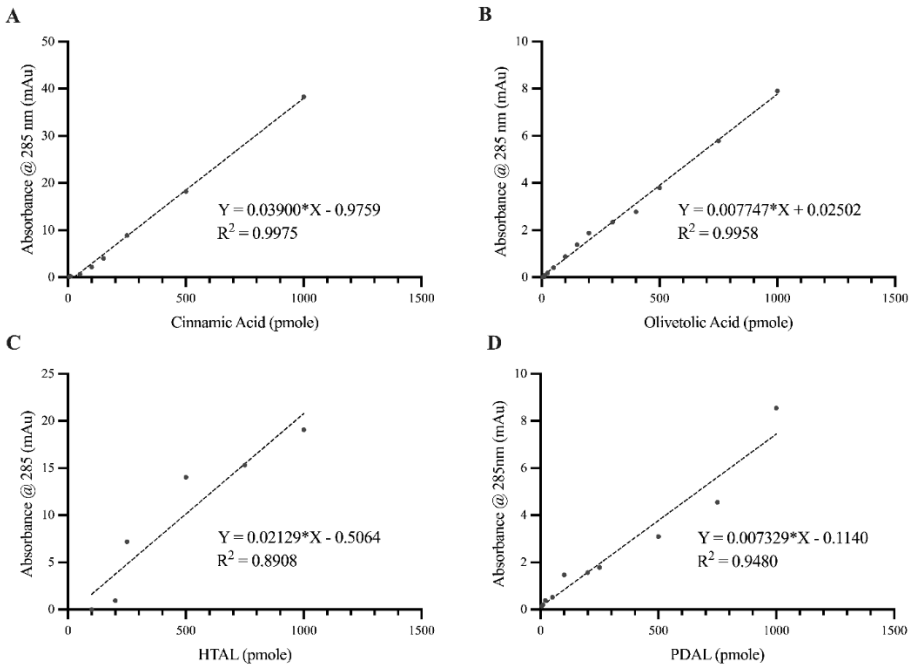


Figure 8. Calibration curves of authentic analytical standards

Analytical standards were resuspended in 50:50 MeOH:Water and analyzed in triplicate using HPLC/UV. Sample injections were 1 µL prepared standard. A. Cinnamic Acid. B. Olivetolic Acid (OA). C. Hexanoyl Triacetic Acid Lactone (HTAL). D. Pentyl Diacetic Lactone (PDAL). HPLC/UV (high performance liquid chromatography/ultraviolet absorbance).

Building of SBS (E. coli and S. cerevisiae)

Initially, all codon-optimized sequences were cloned into pJET1.2 vector (CloneJetPCR cloning kit, Thermofisher). Once enough of each sequence had been produced by bacterial growth and subsequent replication of the pJET1.2 plasmid, the genes were excised from pJET1.2 with restriction enzymes and cloned into final destination vectors. After additional consideration, I decided that due to the lack of endogenous malonyl-CoA production in the *E. coli* hosts, a malonyl-CoA synthetase from *Bradyrhizobium japonicum* (BjMCS) should also be included in this system. Unlike canonical malonyl-CoA synthesis that occurs through the carboxylation of acetyl-CoA, BjMCS catalyzes the synthesis of malonyl-CoA directly from CoA and malonate using ATP as an energy source [139]. Increased malonyl-CoA production has previously been noted

in *E. coli* using BjMCS in OA-producing synthetic biology systems [121]. In the *E. coli* system the BjMCS gene was cloned into the pLATE52 expression vector.

Following successful protein production, all GOI's were cloned sequentially into the final expression cassettes using a DNA-ligase based approach (**Table *2***). For the E. coli system *CsTKS* was cloned first into pCOLADuet-1 at the MCS-1, followed by the incorporation of *CsOAC* into MCS-2. *CsAAE1* was cloned first into the pCDFDuet1 vector at MCS-1, followed by *CsCHI-L* at MSC-2. For the yeast system, *CsTKS* was cloned first into the MCS-1 of pESC-HIS, followed by *CsOAC* at MSC-2. *CsAAE1* was incorporated into the MSC-1 of pCEV-G4-KM, and *CsCHI-L* was into MCS-2. This allowed for production of an entire suite of plasmids that could be transformed into cells to produce systems with different enzymatic activities. The final step in construction of the complete synthetic biology systems was to transform the microbial host (BL21 *E. coli* or INVSc1 *S. cerevisiae*) with the final vector combinations for the *in vivo* feeding experimentation. Regardless of the eventual destination (*E. coli* or yeast), all constructed recombinant plasmids were transformed into BL21 *E. coli* cells directly following cloning (ligation, Gibson assembly or otherwise). This initial transformation was done to verify the legitimacy of the constructed plasmid, either through restriction digest or PCR to check for insert(s) presence. Additionally, plasmids from these transformations were prepared and sent for sequencing to ensure that our genes of interest were properly incorporated into said plasmid. Double and triple transformations were then performed using purified DNA for each plasmid needed in the system. Both systems were successfully constructed. *E. coli* hosts incorporated pCOLA-Duet1+*CsOAC* (*E. coli* opt.)+*CsTKS* (*E. coli* opt) and pCDF-Duet1+*CsAAE1* (*E. coli* opt)/pCDF-Duet1+*CsAAE1* (*E. coli* opt)+*CsCHI-L* (*E. coli* opt) (Figure 9). *S. cerevisiae* hosts incorporated pESC-HIS+*CsOAC* (*S. cerevisiae* opt)+*CsTKS* (*S. cerevisiae* opt) and pCEV-G4-

KM+*CsAAE1*(*S. cerevisiae* opt)/pCEV-G4-KM+*CsAAE1* (*S. cereviaise* opt)+*CsCHI-L*(*S. cerevisiae* opt) (Figure 10).

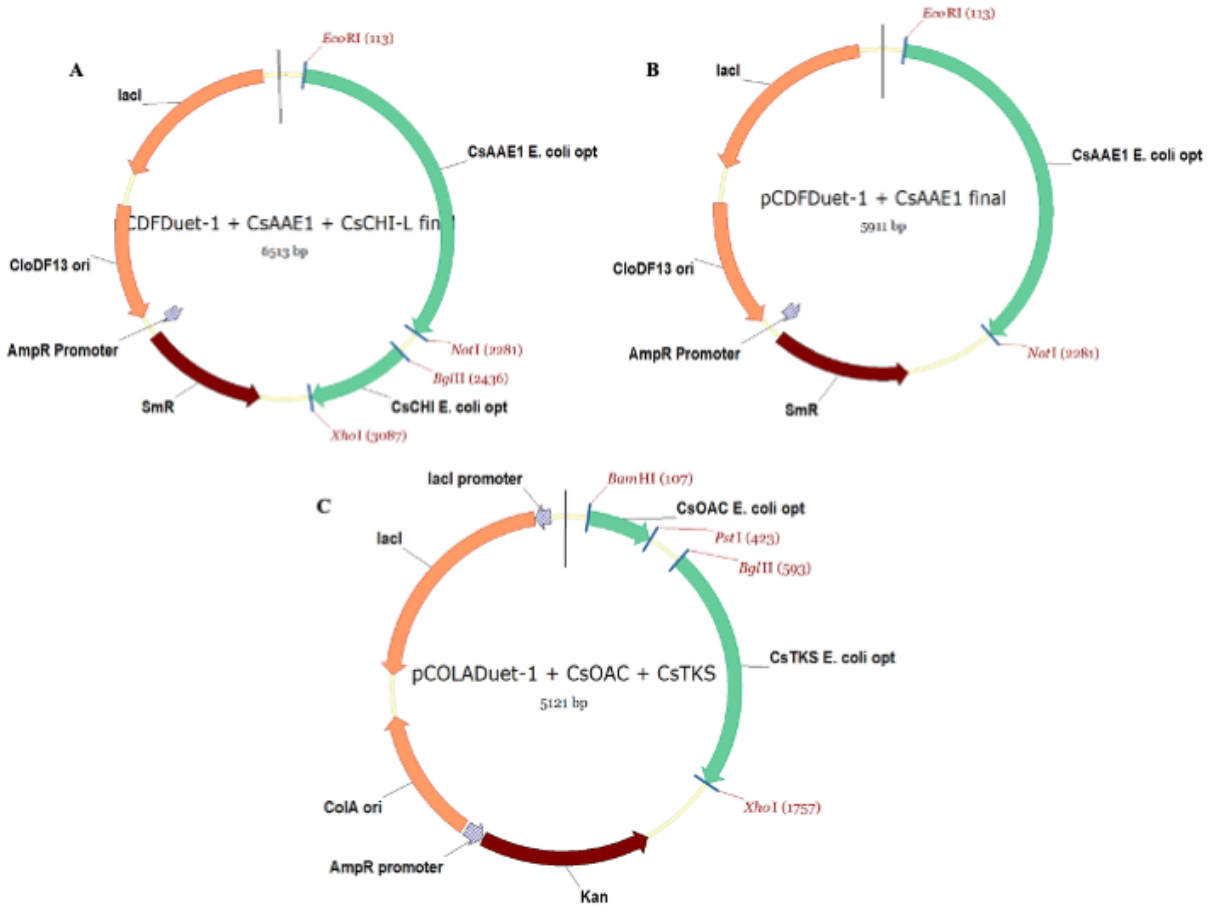


Figure 9. *E. coli* SBS plasmid expression cassettes.

A) pCDF-Duet1(CloDF13 ori, lacI gene, streptomycin resistance) with *CsAAE1* at MCS1. B) pCDF-Duet1(CloDF13 ori, lacI gene, streptomycin resistance) with *CsAAE1* at MCS1 and *CsCHI-L* at MCS2. C) pCOLA-Duet1 (ColA ori, lacI gene, kanamycin resistance) with *CsOAC* at MCS1 and *CsTKS* at MCS2. D) pLATE (pMB1 ori, lacI gene, ampicillin restistance) with *BjMCS*. Note: pLATE52+*BjMCS* plasmid was eventually removed from system and not included in later *in vivo* testing and LC-MS/MS analysis.

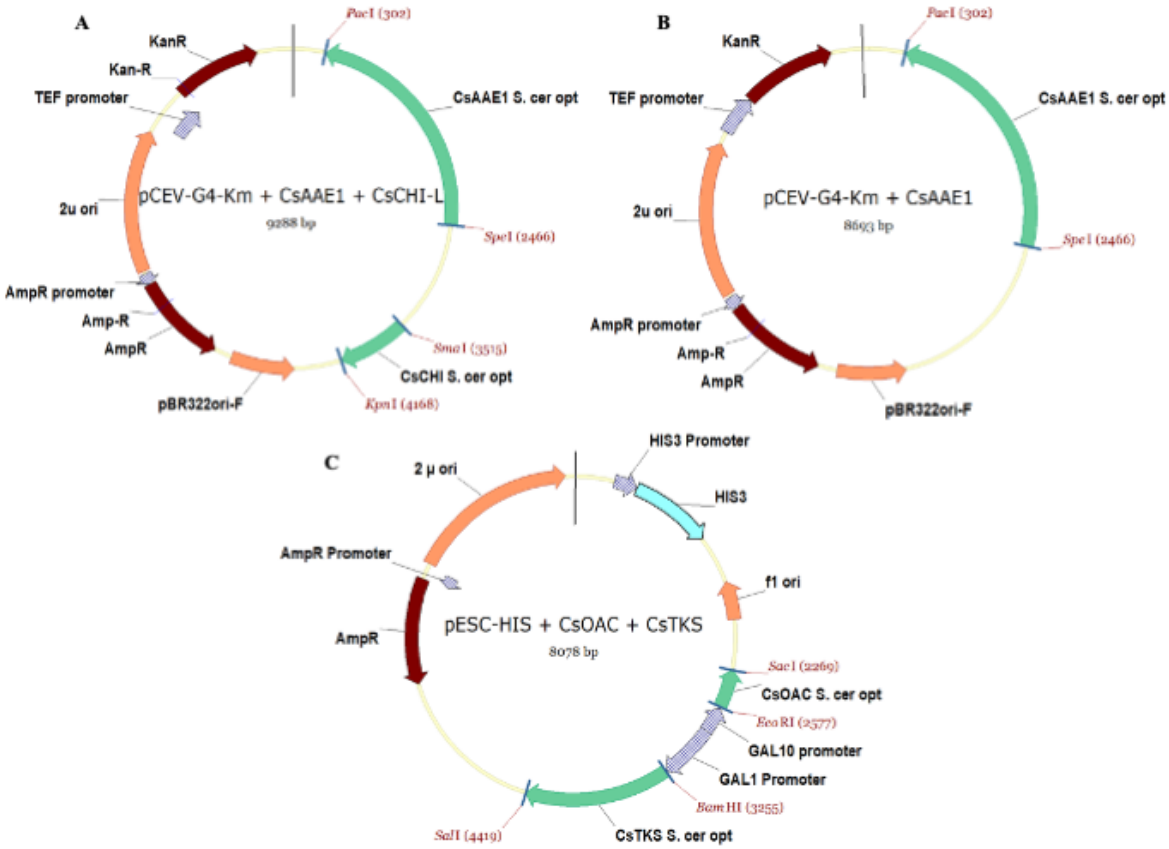


Figure 10. *S. cerevisiae* SBS plasmid expression cassettes

A) pCEV-G4-KM (pBR322-f ori, galactose-regulated, G418/kanamycin resistance) with *CsAAE1* at MCS1. B) pCEV-G4-KM (pBR322-f ori, galactose-regulated, G418/kanamycin resistance) with *CsAAE1* at MCS1and *CsCHI-L* at MCS2. C)pESC-HIS (F1 ori, galactose-regulated, HIS3 marker) with *CsOAC* at MCS1 and *CsTKS* at MCS2.

Discussion

Building of the synthetic biology system(s)

Synthetic biology systems, where heterologous incorporation of biosynthetic genes or gene components results in the directed production of desired compounds, has become an excellent strategy for the sourcing of high-value molecules. However, a large number of issues can still be faced when designing and constructing these systems. Undesired or unpredictable interactions between the host and the introduced transgenes that can result in hampered output.[140]

Natural substrates and other elements (genes or promoter) can behave unpredictably in different combinations, and can exhibit a multitude of metabolic connectivity with the host metabolism, which makes interactions hard to predict and control.[141] Cellular complexity is a major component of this difficulty, where incorporated genes can function in the cellular environment in ways that are not intended or pertinent to the pathway of interest. On the other hand, endogenous machinery can be beneficial – and in many cases, necessary – to facilitate the functioning of genes and pathway that are being incorporated into the host. Synthetic biology is still an emerging field, and the success and appropriate application of multiple strategies are still under development and investigation. In this experimentation, my chassis consisted of the host (*S. cerevisiae* or *E. coli*) with introduced vector cassettes (plasmid DNA) to be replicated and maintained through selection in the host. A different approach used in synthetic biology is genomic integration, where relevant genes – alongside appropriate promoter and terminator sequences – are permanently transposed into the hosts genomic DNA, thus eliminating the need for selection and plasmid maintenance and providing a more intrinsically stable microbial host with incorporated GOI's. This type of chromosomal integration is superior in synthetic biology systems for long-term production and large-scale operations, largely because this approach negates the need for plasmid maintenance in the expression systems.[142] Examples of genomic integration include Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) based approaches for efficient, single-step integration of sequences with homology arms to distinct loci.[143] Following difficulties experienced during the system construction process, genomic integration was considered as an alternate approach for our *S. cerevisiae* system. This approach would have used Gibson-assembly generate fragments alongside CRISPR and guide-RNA to integrate our sequences into the yeast DNA via a homology-mediated end joining approach. Due

to time constraints such a yeast genomic integration was not attempted, although I did use the Gibson-assembly approach for the re-construction of the pESC-HIS+CsTKS+CsOAC yeast expression plasmid.

In my experimentation, the building of both synthetic biology systems in microbial hosts *S. cerevisiae* and *E. coli* were ultimately successful. Our genes of interest (*CsOAC*, *CsTKS*, *CsAAE1* and *CsCHI-L*) were properly cloned into plasmid vectors, which in turn were successfully incorporated into their respective microbial hosts.

Codon optimization of sequences

Following determination of the sequences needed for experimentation (and the vectors to incorporate them with), codon optimization was needed to allow for proper protein production in the microbial hosts. The basis for this process is the internal bias that organisms have for specific codons, where preferred codons are more accurately and efficiently translated.[144] Often, preferred codons are related to the diversity and predominance of tRNA's required to translate an mRNA transcript into a protein, where low concentration tRNAs correspond to "slow" codons. In every cell, tRNA populations are correlated to reflect the bias of the mRNA population and codon frequencies.[145,146] Ill-suited tRNA pools can lead to halting of translation, premature termination, frameshifting and amino acid incorporation.[147] Codon-optimization refers to the gene editing technique of synonymous codon changes that result in an increased protein production. This gene engineering practice takes advantage of the redundancy of the genetic code, where different nucleic acid sequences can be translated into the same amino acid sequence for protein production.[148] Codon-optimization can dramatically increase protein expression levels in microbial hosts, and has become an integral part of modern biotechnology.[149]

Cloning techniques used in construction of SBS

Codon optimized sequences were cloned into the multiple different vectors on the journey to incorporate them into their final destinations. Molecular cloning is a method to insert a gene (or multiple) into circular plasmid DNA, to create a recombinant DNA molecule. Cloning techniques used in this experimentation were both ligase-based and ligation-independent. In DNA replication, DNA ligase joins together newly synthesized fragments of DNA. This same activity is utilized in cloning, where pieces of DNA have been excised using restriction enzymes (to create identical cut sites) and are then stitched back together using the DNA ligase. Alternatively, ligation-independent cloning utilizes the 3'→5' exonuclease activity of the T4 DNA polymerase to create complimentary overhangs, which are then repaired by endogenous *E. coli* machinery following transformation.

Protein expression and the lac operon

Protein expression in *E. col*i hosts was achieved using a lac promoter regulated system, with isopropyl β-D-1-thiogalactopyranoside (IPTG) driven expression. The lactose operon is responsible for the metabolism and transport of lactose in *E. coli*, where beta-galactosidase allows for the catabolism of lactose when the preferred glucose carbon source is not available. Briefly, the lac repressor (lacI), regulates the transcription of the lac operon by inhibitory binding. When lactose – or allolactose – is present in the system, a conformational change occurs which changes the binding affinity of LacI for the operon and allows for transcription to occur. IPTG, a chemical analogue of galactose, is an allolactose mimic that can also bind to the tetrameric lacI repressor and induce transcription and subsequent protein expression. Unlike allolactose, IPTG is non-hydrolyzable by the cell, and therefore its concentration will remain constant and protein expression will continue at the same elevated rate throughout the

experimental process. Decreased biosynthesis has been noted in *E. coli* cultures induced with high concentrations of IPTG, due to the cellular toxicity of protein inclusion bodies following high levels of protein expression.[150,151] As determined in our lab for Type III CHS enzymes CsTKS and PoPKS (by Kyle MacDonald), optimal IPTG concentration is 0.4 mM at 23 °C for 22 hours. As confirmed by SDS-PAGE imaging, all proteins were properly expressed from induced BL21 *E. coli* cultures.

GAL-induced protein expression

Protein expression in yeast cultures was driven by galactose induction. The galactose (GAL) network found in yeast cells is a well-characterized regulatory system, with multiple regulators.[152] It allows the yeast cells to change their utilization of sugars for growth. When glucose is present and being used as the main carbon source for the yeast cells, the Mig1 protein suppresses the GAL genes.[153] When galactose is present (and glucose absent), transcriptional regulator Gal4p binds to and enhances multiple promoters to drive GAL gene expression. Galactose enters the cell through the transporter Gal2p, and is activated by galactokinase Gal1p.[154–156] Galactose will then bind to signal transducer Gal3p and activate it, causing Gal3p to form a complex with Gal80p.[157] This sequestration of Gal80p releases Gal4p and allows it to active the GAL network.[158,159] This mechanism can be utilized in synthetic biology approached for expression of recombinant proteins, where the Gal1/Gal10 promoter(s) are used at the cloning sites in the expression vectors.

The yeast expression cassettes used in the *S. cerevisiae* synthetic biology system are pCEV-G4-KM, which has TEF1/GAL10 and HXT7 promoters for cloning sites, and pESC-HIS which has GAL1 and GAL10 divergent promoters. GAL1 and GAL10 promoters are galactose-dependent promoters, where they are repressed in the presence of glucose and switched on in the

presence of galactose. HXT7 is an additional inducible promoter that works in a complementary manner to the GAL genes; HXT7 is up-regulated when glucose concentration is low and repressed when it is high. HXT7 is therefore another promoter that will drive gene expression in low or no glucose media, including the galactose media used for the GAL-driven protein expression.[160]

Multiple-vector transformations of hosts

Different approaches were used for *E. coli* and *S. cerevisiae* hosts; although largely electroporation was the most successful for both. For both systems, multiple attempts were required to facilitate successful transformants. As the electro-competent host cells are in a very delicate state, this was to be expected. When working with the *E. coli* system, a variety of DNA amounts were used to transform the cells, ranging from about 150-1000 ng of plasmid DNA. Following a shock of 1.8kV, pre-warmed liquid media was added to the cells and they were allowed to grow in recovery before being plated. Electrocompetent *E. coli* cells were initially made by washing log-growth phase cells with 10% glycerol at cold temperatures (0 °C or ice-cold). This methodology followed standard procedure for *E. coli* transformations [161,162].

Development of a successful yeast transformation protocol was challenging, and I adjusted my methods very frequently throughout the process. Although many protocols advocate for the linearization of DNA and the presence of carrier DNA to transform *S. cerevisiae* cells, I found these methods unnecessary to facilitate the transformation. Electroporation is sufficient to introduce plasmid DNA into intact yeast cells, as initially demonstrated by Hashimoto et al. [163] Electrocompetent yeast cells were prepared by various rounds of washing and harvesting, with a final resuspension in 1M sorbitol. Electrocompetent cells were then aliquoted and frozen to -80 °C, in imitation of the method developed by Suga and Hatakeyama.[164] In my experimentation, *S.*

cerevisiae hosts were combined with 100-300 ng of DNA, and allowed to incubate briefly before being transformed. Initially, larger amounts of DNA were used (~1000 ng), but lower amounts of DNA proved more successful. My protocol used a lower electrical voltage for their shock treatment than was used for the *E. coli* transformants (1.5 kV vs the 1.8 kV for bacterial cells), and yeast cells were incorporated with 1 M sorbitol directly after said shock was applied. The use of sorbitol to provide osmotic support was discovered by Becker and Guarente, and improves transformation efficiency so drastically that it has since been incorporated into standard protocol for yeast transformations.[165] I initially directly plated the sorbitol-cell mixture onto selective media plates as advocated for in other yeast protocols. However, this did not lead to high transformation efficiency, and in some cases, there was no colony growth on plates even after incubation for four to six days. I then adjusted the protocol to allow the yeast cells a recovery period, by combining the sorbitol-cell mixture with yeast extract-peptone-dextrose (YPD) media and allowing them to shake at 30 °C for a period of 1-3 hours. One hour proved to be sufficient to allow supportive growth for the transformed cells. Additionally, I began to add 1 M sorbitol to the selective media plates to provide additional support. This dramatically improved the transformation efficiencies of the yeast and allowed for success in the transformations.

Use of selective media and antibiotics in SBS construction

In order to facilitate isolated growth of colonies containing our expression constructs, plates were prepared with either selective dropout (SD) media, antibiotics or both. For *E. coli* constructs, antibiotics were used to limit growth of *E. coli* colonies that were unsuccessful in incorporating the plasmid DNA. This is done through the presence of dominant selective genes, which confer resistance to specific antibiotics. For yeast constructs, both auxotrophic selection markers and dominance resistance markers were used. Selection in yeast plasmids is often

maintained by the complementation or repair of an auxotrophic chromosomal mutation specific for an amino acid. In our system, SD HIS-deficient media and antibiotics were used to ensure growth of only plasmid-incorporated colonies. Due to the highly selective nature of the dropout media, supplemented with the presence of G418 antibiotic, colonies growing on yeast plates were assumed to have incorporated both plasmids as there would be no growth on said plates if they had not. G418 is an aminoglycoside and acts to restrict the growth of both prokaryotic and eukaryotic cells through inhibition of the elongation step in polypeptide synthesis and therefore demonstrates toxicity towards mammalian, higher plant, yeast, protozoan and bacterial cells. Due to its wide range of targets, G418 has demonstrated the ability to select for yeast cells without the use of auxotrophic markers. [166] This was demonstrated by a control study done with INVSc1 cells plated onto agar plates of YPD media, SD media and SD media with 400uM G418 (concentration used for plating transformants). Confirmation of the selectivity of the HIS-deficient media combined with 400uM G418 antibiotics was demonstrated by the lack of fungal colony growth for a strain lacking the appropriate plasmids on the selective plates (Figure 5). Additionally, confirmation of the ability of properly incorporated plasmid yeast strains to grown on these selective plates was also demonstrated (Figure 5).

OA is produced in in vitro enzymatic assays

The *in vitro* assays with enzymes produced by our GOI's (CsOAC, CsTKS, CsAAE1 and CsCHI-L) were able to produce OA when combined with starter molecules hexanoyl-CoA and malonylCoA. This proves that the codon-optimized sequences we have chosen to incorporate into our BL21 hosts are capable of producing the proper enzymatic components of the cannabinoid biosynthetic pathway. The most 'basic' enzymatic combination of OAC/TKS was enough to produce OA, where the addition of AAE1 and CHI-L enzymatic components did not

significantly change the outcome of the compounds produced. In these assays, the presence of the AAE1 enzyme is fairly redundant, as this enzyme functions to synthesize hexanoyl-CoA from hexanoate in the Cannabis plant.[86]. In my enzymatic assays, the hexanoyl-CoA and malonyl-CoA substrates were already present as acyl-activated compounds as there were no enzymatic components capable of producing malonyl-CoA present in the assay and it was required to be directly added. CsAAE1 was included in these assays to see if there was any kind of functionality noted between the CsAAE1 enzyme and the proposed CsCHI-L, as we are monitoring all potential protein interactions. However, there was no significant difference in the amount of OA produced between assays, even those that included CsCHI-L and those that did not in both the 24 hour and 10 minute assays (Figure 7). In the 10 minute assays, CsOAC/CsTKS had statistically signifigantly lower amount of PDAL produced than the CsOAC/CsTKS/CsAAE1 and the CsOAC/CsTKS/CsAAE1/CsCHI-L groups (Figure 7). In the 24 hour assyas, CsOAC/CsTKS showed statistically significant lower higher accumulation of PDAL when compared to the CsOAC/CsTKS group (Figure 7).These results seen in the 10 minute assays with respect to PDAL production were unanticipated, as our assumption is that inclusion of CsAAE1 and CsCHI-L proteins would make OA production more efficient and lead to lower production of derailment products. However, as these results are from in vitro assays, it is possible that the random distribution of substrates in the assays combined with lack of cellular environment to properly direct intermediates between enzymes in a constricted time frame lead to these anomalies. The results seen in the 24 hour samples with respect to PDAL production support the hypothesis that inclusion of AAE1 and CHI-L enzymes will reduce derailment product accumulation, however as differences between other groups were negligible, and the differences noted in the 10 minute samples directly contradictory, little insight can be gained

from these results. Differences in the production of HTAL were not demonstrated to be statistically significant between any experimental combinations, in both the 10 minute and 24 hour assays.

CsAAE1 has demonstrated its capability to activate hexanoate in vitro and form hexanoyl-CoA.[86] If additional assays in this line of experimentation were to be run then I would have like to have included starter molecules hexanoate and malonate alongside the CsAAE1 and BjMCS enzymes in vitro. However, the large goal for this line of experimentation was simply performed to monitor the enzymatic capabilities for the production of OA, which was achieved. I did not expect to see a large difference in the efficiency of OA produced between assays containing the CsCHI-L candidate and those that did not, as our candidate is an accessory protein and largely would function in an in vivo situation, where sequestration of the tetraketide intermediate is required to facilitate a more efficient reaction. CHI-L proteins that play an active role in H. lupus xanthohumol synthesis (HlCHIL1 and HlCHIL2) also do not exhibit action in vitro as they function in the glandular trichome and prokaryotic cellular environments. While there was an increased amount of OA produced in the 10 minute samples of CsTKS/CsOAC/CsAAE1, CsTKS/CsOAC/CsCHI and CsTKS/CsOAC/CsAAE1/CsCHI when compared to the CsTKS/CsOAC assays, which had been functioning as the control group, this difference was not statistically significant and therefore does not support our hypothesis (Figure 7).

Chapter 3: Olivetolic Acid production in *E. coli* and *S. cerevisiae* SBS

Introduction

Following successful construction of the two synthetic biology systems, I next tested the systems for i) their ability to produce OA, our main compound of interest, and ii) to see if the presence of our proposed CsCHI-L protein had any effect on the ability of the pathway to efficiently produce OA. Having already looked at how our enzymatic components function in *in vitro* assays, I needed to monitor the components *in vivo*, where cellular environment will be closest to that seen *in planta*. The concepts behind an *in vivo* culture feeding experiment are fairly simplistic, where an appropriate level of cells are produced in culture, protein expression induced and starter molecules or substrates introduced into culture for the enzymes to interact with. In this line of experimentation, cultures were either fed the starter molecules sodium hexanoate and sodium malonate, or hexanoic acid and malonic acid. Hexanoic acid, also known as caproic acid, is a short-chain fatty acid (C6H12O2). Sodium hexanoate is the ionized form of hexanoic acid, with a sodium molecule replacing a carboxyl R-group. Both molecules can be used to synthesize hexanoyl-CoA. Malonic acid is a dicarboxylic acid where two carboxyl groups are separated by a single methylene group (C3H4O4). Similarly, the ionized forms of this molecule are known as malonates, such as the sodium malonate used in our experimentation. Malonates and malonic acid are both used to synthesize malonyl-CoA. Sodium malonate and hexanoate were used for the feeding of the *E. coli* systems, where hexanoic and malonic acid were used in the *S. cerevisiae in vivo* experimentation.[120,121] Compounds produced from *in vivo* experimentation are harvested or extracted from culture.

Materials and Methods

Experimental overview

Multiple approaches were used to facilitate the feeding experiments in this line of research. I began with the triple-transformed *E. coli* system as that was the first to be constructed and ready to test. This system had pCOLA+*CsOAC*+*CsTKS*, pCDF+*CsAAE1* or pCDF+*CsAAE1*+*CsCHI-L* as well as pLATE52+*BjMCS*. Beginning with overnight (5 mL) cultures in selective media, cells were grown up to a volume of 50 mL in generic (non-selective) media, where they were allowed to reach appropriate growth phases before being induced for protein expression with IPTG. Cultures were allowed sufficient time to induce before being fed starter molecules malonic acid and sodium hexanoate. Although *E. coli* do not have endogenous hexanoyl-CoA synthase activity, the introduced CsAAE1 will produce hexanoyl-CoA, and similarly the activity of BjMCS was expected to increase the malonyl-CoA concentrations above the endogenous levels. Following the addition of the starter substrates, the cultures were allowed to react for 48 hours before being removed for extraction and analysis with HPLC-UV.

After the initial failure with the *E. coli* system, I set up a series of OA-spiked feeding experiments to verify both our experimental method and extraction method. This was done using the *E. coli* SBS with added authentic OA standard, where we could monitor for degradation of the compound. Briefly, 0.25mg/L OA was added to cultures both at a timepoint of 0 (Early stage spike) and 48 hours following culturing (Late stage spike). All other conditions – IPTG level, temperature, feeding of starter substrates – were done as in our feeding experiments. The same extraction process was also followed. Samples from the early and late stage spike experimental groups were analyzed using HPLC-UV to monitor for presence of OA. Successful detection of OA in these experiments verified my culturing and extraction process as viable, where produced OA should not be degraded by either process.

Following this verification, reconsideration of the *E. coli* system was necessary, and I decided that the inclusion of the BjMCS cassette (pLATE52) may be hindering protein expression of the other vectors as it is not designed in the same Duet expression system and may have competing or non-compatible origins of replication. Previous experimentation done with *E. coli* has shown that incorporation of the BjMCS protein will help with production of OA, but is not necessary for its biosynthesis due to the presence of endogenous malonyl-CoA in the cell[121]. Removal of the BjMCS from the *E. coli* system resulted in a simpler, more streamlined approach to OA production.

Feeding experiments were then done using only the dual-transformed BL21 hosts (with the system containing pCOLA+*CsOAC*+*CsTKS* and either pCDF+*CsAAE1* or pCDF+*CsAAE1*+*CsCHI-L*). These feeding experiments were conducted following the protocol established by Tan et al. as detailed in methods. [121] Briefly, cultures are allowed to grow to an appropriate OD_{600}, before allowing IPTG-based induction for 15 hours. Cells are then harvested, washed and resuspended in M9Y media to a much higher OD_{600} of ~3. These cultures are then used for subsequent feeding experiments, where sodium malonate and hexanoate are added and allowed to incubate for 48 hours. These feeding experiments were extracted and analyzed using liquid chromatography with tandem mass spectrometry (LC-MS/MS).

The yeast expression system feeding experiments followed a very similar trajectory to the *E. coli* feeding experiments. Yeast cultures were started in SD minimal media (supplemented with G418 for cultures containing the p-CEV-G4-KM plasmid) and grown overnight. A galactose-containing media (YPG), which drives the expression of the GOI's as they are driven by galactose-inducible promoters, is then inoculated with the overnight cultures. Following sufficient cell growth, the cultures were supplemented with malonic acid and hexanoic acid as

starter substrates and left to react for 48 hours. The cultures were supplemented with galactose throughout the experiment, to facilitate continued protein expression. Cultures were then treated with lyticase, frozen and sonicated to release our compounds of interest from the yeast cells prior to extraction. Initial samples were analyzed by HPLC-UV, however it became apparent that the quantity of target molecules produced by the cultures were possibly below the limit of detection for this instrument. Thus, subsequent *S. cerevisiae in vivo* samples were analyzed by LC-MS/MS using a triple quadrupole mass detector, which is one of the most sensitive analytical techniques for detecting and quantifying small molecules.

Following an unsuccessful initial round of experimentation with the *S. cerevisiae* SBS, constructs were examined by restriction digest and it was determined that the pESC-HIS vector used did not contain either *CsOAC* or *CsTKS*. As these enzymes are essential to the OA forming steps of the pathway, the pESC-HIS vector was rebuilt using a Gibson assembly method. Subsequently, the *S. cerevisiae* SBS was completely rebuilt using this new vector and used for additional *in vivo* feeding experimentation.

E. coli in vivo assays/ culture growth and feeding experiments

Various approaches were used to find success with *E. coli* feeding experimentation. Initially, all combinations of expression cassettes were used. This gave four experimental groups as follows. Group one was pCDF+*CsAAE1*/pCOLA+*CsOAC*+*CsTKS*/pLATE52+*BjMCS*. Group two was pCDF+*CsAAE1*+*CsCHI-L*/pCOLA+*CsOAC*+*CsTKS*/pLATE52+*BjMCS*. Group three was pCDF+*CsAAE1*/pCOLA+*CsOAC*+*PoPKS*/pLATE52+*BjMCS*. Group four was pCDF+*CsAAE1*+*CsCHI-L*/pCOLA+*CsOAC*+*PoPKS*/pLATE52+*BjMCS*. All cultures for *E. coli in vivo* experimentation were grown up into 50 mL cultures in LB liquid media with 50mM AMP/50 mM SPEC/30 mM KAN until an OD_{600} of 0.4-0.6 was achieved. Cultures were

supplemented with 400 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) and allowed to induce at 22 °C for 15 hours. Cells from culture were then harvested by centrifugation, washed and resuspended in M9 liquid media (93.0 mM Na^+ 22.1 mM K^+, 18.7 mM NH_4, 1.0 mM Ca^{2}+, 0.1 mM Mg^{2+}, 29.2 mM Cl^-, 0.1 mM SO_4^{2-}, 42.4 mM PO_4^{3-}, 10% w/v glucose). Cultures were then fed 4 mM hexanoic acid and 12 mM malonic acid and allowed to react. Samples (1 mL) were pulled from the cultures at 24, 48 and 72 hours. Media and cells were then extracted for analysis.

Cultures were also grown using TB media and induced with varying concentration of IPTG in an attempt to investigate the efficacy of different levels of protein induction with respect to IPTG concentration. Overnight cultures were grown in 5 mL LB with 50uM SPEC/50uM AMP and 30 uM KAN before being used to inoculate 25 mL of TB media (72 mM K_2HPO_4, 17 mM KH_2PO_4 with 12 g/L bacto tryptone, 24 g/L yeast extract and 4% w/v glycerol) supplemented with the same antibiotics. These cultures were grown until an OD_{600} of 0.8 before inducing with 0.5 mM, 0.25 mM, 0.1mM and 0.05 mM IPTG (n=4). Cultures were left to induce for 24 hours at 20 °C. Cultures were then fed 4 mM hexanoic acid and 12 mM malonic acid and left to react for 48 hours.

Finally, a simplified version of the experimental system was used. This was comprised of either pCOLA+*CsOAC*+*CsTKS* alone or pCOLA+*CsOAC*+*CsTKS* and pCDF+*CsAAE1* or pCOLA+*CsOAC*+*CsTKS* and pCDF+*CsAAE1*+*CsCHI-L*. Strains containing the BjMCS plasmids were left out of this experimentation due to concern about pLATE52 competing with other plasmid origin of replication and causing incompatibility. These experiments were carried out in a combination of LB and M9 media following the protocol outlined by Tan et al.[121] 5 mL starter cultures were grown in antibiotics (50 mM Kan for cultures with only pCOLA plasmid or

50 mM Kan/Spec for cultures with both pCOLA and pCDF plasmids) overnight. 50 mL of LB was inoculated using 1 mL of the overnight culture and grown to an OD_{600} of 0.4-0.8. Cultures were induced with 0.5 mM IPTG and left to induce for 20 hours at 22 °C . Cells were then harvested via centrifugation and washed with 25 mL of M9Y media before a final resuspension in 50 mL M9Y. Resuspension was facilitated by gentle shaking on a platform. Cultures were then fed 4 mM hexanoic acid and 12 mM malonic acid and left to react for 48 hours at 22 °C at 220 rpm.

OA-Spike feeding experiments

Following concerns about compound degradation, I performed a series of OA-spiked feeding experiments to verify the presence of known OA at the end of the culturing and extraction process. The same protocol as identified above for culture with TB media was used. 5 mL pCDF+*CsAAE1*/pCOLA+*CsOAC*+*CsTKS* overnight culture in LB was transferred into 50 mL TB media and grown with shaking at 220 rpm at 37 ˚C to an $O.D._{600}$ of 0.8. Protein expression was induced with 0.5 mM IPTG. Cultures were left to express at 22 ˚C overnight. Following induction for 15 hours, experiments were fed the 4 mM hexanoic acid and 12 mM malonic acid. An OA spike of 0.25 mg/L or 1.1148 uM was also added to the "Early spike" cultures at this time. Cultures were left to react for 48 hours. At this time point 0.25 mg/L OA spike was added to separate "Late spike" cultures that served as controls to ensure that I could recover and detect this concentration of OA from *E. coli* cultures. Spent media from these cultures then underwent extraction and HPLC analysis (extraction protocol outlined below in "Sample extraction and preparation for analysis").

S. cerevisiae feeding experiments

Yeast feeding experimentation was initially carried out using strains harbouring four experimental groups, which are as follows. Group one: pESC-HIS+*CsOAC*+*CsTKS* and pCEV-G4-KM+*CsAAE1*. Group two: pESC-HIS+*CsOAC*+*CsTKS* and pCEV-G4-KM+*CsAAE1*+*CsCHI-L*. Group three: pESC-HIS+*CsOAC*+*PoPKS* and pCEV-G4-KM+*CsAAE1*. Group four: pESC-HIS+*CsOAC*+*PoPKS* and pCEV-G4-KM+*CsAAE1*+*CsCHI-L*. Overnight cultures of each group were grown in 5 mL of SD Media (His-) with 400 µM G418 antibiotics. 1 mL of these starter cultures were used to inoculate 25 mL of yeast extract peptone galactose (YPG) medium. Cultures grown at 30 ˚C until they reached an OD_{600} of 0.2. Cultures were then fed with hexanoic acid and malonic acid to a final concentration of 4 mM and 12 mM respectively. Cultures were incubated with shaking for 48 hours at 30 ˚C, with one supplementation of 2% (w/v) galactose at the 24 hour timepoint.

E. coli culture extraction

As our compounds of interest would mainly be exported out of the cell due to toxicity levels and into the culturing media, our focus when analyzing the samples was directed towards the media. [121] With the initial round of experimentation, cell mass was also extracted for analysis to verify the lack of compounds of interest. Cultures were spun down by centrifugation at 16,000 *g* to pellet the cells. Spent media was decanted and subjected to three rounds of ethyl acetate (EtAc) extractions. For small volume extractions (1mL), equal volumes of EtAc were used for all three rounds of extraction. For larger volumes (25-50 mL's), half volumes of EtAc were used for the initial two rounds followed by a final extraction done with 5 mL of EtAc. During each round of extraction, EtAc was mixed thoroughly with spent media using a vortex speed mixer and centrifuged for 5 minutes to achieve separation of aqueous and organic layers. Organic layers

were carefully transferred to glass collection tubes using pipettes. The EtAc extracts were then dried using different approaches were used depending on the volume of the extraction: small volume extractions were dried down using a speed Vacufuge at 45 °C , whereas large volume extractions were dried down with under a nitrogen gas stream while held in a heating block set at 30 - 50 ˚C).

Although not the primary target, cells from the *E. coli* platform were tested to see if they contained any compounds of interest. Pelleted cells from 50 mL cultures were first frozen at -20 ˚C, which acted as a physical lysis. Cells were then resuspended in Lysis Buffer (50 mM Tris pH 7 500 mM NaCl, 2.5 mM imidazole, 5 % v/v glycerol, 0.5% v/v Tween 20, 0.75 mg/mL lysozyme, 10 mM β-mercaptoethanol; LB) and sonicated using a 10:10 sonication:rest pulse for three minutes at a pulse strength of 65% to disrupt the cellular membrane and release internal compounds. Following these cellular lyses, cellular debris was pelleted and the remaining aqueous solution was subjected to EtAc extraction as described above.

S. cerevisiae culture extraction

Due to the nature of the yeast cells and the presence of a chitinous cell wall, the INVSc1 cultures required additional steps prior to extraction of compounds of interest. Each of the 50 mL cultures were directly treated with 1 Unit of lyticase per each unit of the culture OD_{600}. Lyticase treatment was carried out for two hours at 30 °C while shaking at 220 rpm. Cultures were then frozen for subsequent extraction. Cultures were thawed and subjected to sonication for cellular lysis. Following lysis, cellular debris was pelleted via centrifugation and the remaining supernatant extracted with EtAc. Volumes of 5 mL, 5 mL and 2.5 mL were used for each round of the extraction. Extractions were dried using nitrogen gas and a heating block until roughly 1mL of extraction remained. This was evaporated fully using the Vacufuge.

HPLC sample preparation

All samples used for HPLC analysis were resuspended in 10-100 µL of 50:50 MeOH: Water.

LC-MS sample preparation

All samples used for LC-MS analysis were resuspended in 1 mL of 50:50 LCMS grade Acetonitrile (ACN): Water. These suspensions were then syringe filtered using a µM filter (BioRad). All LC-MS/MS samples included the addition of 1 µL 10 mM cinnamic acid as an internal standard.

Compound analysis using HPLC-UV

Samples retrieved from extracted *in vivo* and *in vitro* experimentation were analyzed on an Ultimate 3000 HPLC system with UV detection (Dionex UltiMate 3000 diode array detector, Thermo Fisher Scientific) using a Phenomenex Luna Omega C18 column (bead size 1.6uM, column 50 x 2.1 mm) and Luna Omega C18 guard column. A dual solvent system was used for the separation of sample compounds, where acetonitrile facilitated organic solvent separation and a 0.01% formic acid (v/v; prepared with nuclease-free water) solution comprised the polar solvent. Separation method(s) were developed using the Chromeleon7 software (ThermoFisher).

For *E. coli* SBS experiments, samples were run starting with a flow rate of 0.3 mL/minute with a gradient of 65% ACN dropping to 49% ACN over the course of 6 minutes, with 49% ACN to 0.5% ACN over 1.5 minutes and 0.5% ACN back to 65% ACN for the next 1.5 minutes. *S. cerevisiae* SBS experiments were run with a flow rate of 0.3 mL/minute with a gradient of 75% ACN dropping to 45% ACN over the course of 45 minutes, with 45% ACN to 25% ACN over 3 minutes and 25% ACN back to 75% ACN for the next 3 minutes.

Samples were monitored for the presence of any compound of interest (HTAL, PDAL, Olivetol and OA) by comparison of retention times and UV-VIS generated spectra of compound peaks with those generate from authentic standard injections.

Compound analysis using LC-MS/MS

Samples obtained from *in vivo* SBS feeding experiments (from both *E. coli* and *S. cerevisiae* systems) were prepared and analyzed using LC-MS/MS. Samples were separated using liquid chromatography and identified based on mass as observed by a mass spectrometer. A dual solvent system was used to achieve compound separation, with 100% ACN and 100% water used as the respective solvents. A gradient method was created to base-line separate compounds. Beginning with 15% ACN and 85% water, the gradient was increased to 100% ACN over 0.5 min, where it was held for three minutes before ramping back down to 85% water and 15% ACN over one min. A final column flush was held at this gradient for four minutes. A flow rate of 0.450 mL/min was used for the entirety of the gradient, and injection volumes for each sample were 10 µL. Compounds were run through a HSS T3 column featuring a 1.8 µm bead size and measuring 50 x 2.1mm (ACQUITY). Compounds were subjected to analysis through tandem mass spectroscopy (MS/MS). A triple quadrupole was used analyze samples and standards, where mass and charge ratios are given for by inspection of each compound and its daughter ions (1260 quaternary pump and 6470 triple quadrupole, Agilent). Compounds were subjected to ionization as they passed through the first quadrupole, set in negative mode for OA, HTAL and PDAL standards while set in positive mode for the Cinnamic acid internal standard. Both positive and negative charges were set at a capillary voltage of 3500 V and a nozzle voltage of 500 V, with a ΔEMV (+) of 200 and a ΔEMV (-) 300. Samples were held constant at 42 ˚C. This same method was applied to increasing set concentrations (0.1, 1, 5, 10, 50 and 100 ppb) of

OA, HTAL, and PDAL standards to create a calibration curve for each compound. Data from the LC-MS/MS was analyzed using Mass Hunter and Optimizer software to give area under the curve, concentration, and signal to noise (S/N) ratios for each sample and standard injection. S/N ratios are important for the validity of the response received at each sample, where if they are too small, the signal can be considered invalid due to the similarity between the "signal" and the background noise of the instrument. Approximation of a valid S/N ratio is largely considered to be 10 or higher. However, calculation of the limit of detection (LOD) and limit of quantification (LOQ) where possible are considered to be the highest standard in determining the capacity of an instrument and method and are based upon S/N data.

To determine the LOD and LOQ, 10 sample injections of 1 ppb standard mix were run through the LC-MS/MS. Using these samples allows for the determination of the S_0, the standard deviation of the 10 injections. Using the S_0, we can calculate the S_0^1 (Equation 2). In turn, the S_0^1 can be used to calculate LOD and LOQ values (Equations 3 and 4). Equations for the calculation of LOD and LOQ values are shown below.

Equation 2. Calculation of the S_0^1 from S/N values of 10 sample injection of 1 ppb.

$$\boldsymbol{S_0^1 = \frac{S_0}{\sqrt{10}}}$$

Equation 3. Calculation of the limit of detection using the S_0^1

$$\boldsymbol{LOD = 3 \times S_0^1}$$

Equation 4. Calculation of the limit of quantification using the S_0^1

$$\boldsymbol{LOQ = 10 \times S_0^1}$$

Statistical analysis of data

Data sets were subjected to statistical analysis to determine significance of results. The Shapiro-Wilk test was used to determine normalcy of the data distribution before subjected to further parametric measures. One-way ANOVA was used for all data sets, corrected for unequal standard deviation with Brown-Forsythe and Welch and supplemented with Dunnet's multiple comparison test for individual variance. Analysis was run using GraphPad (PRISM) software. In the yeast SBS system, one of the data points for the CsOAC+CsTKS experimental group was above the highest concentration value in the calibration curve, and therefore possibly unreliable as a data point. All other measured values fell within the measurable range. Exclusion of this singular outlying data point changes the average production of OA in the CsOAC+CsTKS experimental group from 99.640±81.926 ppb to 59.607±21.256 ppb, and although this reduced the mean OA value, the reduction in variance caused this experimental group to be significantly different compared to the control.

Results

OA production in E. coli SBS in vivo experiments

Samples were extracted using EtAc from either spent media or pelleted cells pre-treated with physical freeze/thaw lysis. Previous work done with *E. coli* producing cannabinoids or cannabinoid-pathway intermediates has demonstrated that analysis of the spent media is sufficient for analysis of compound production.[121,167] Cells were extracted in my initial *E. coli* experimentation to confirm the lack of OA, HTAL and PDAL, but this was not repeated in subsequent experimentation as it was not necessary for evaluation. There was no difference noted between the varying levels of IPTG concentration used for induction and corresponding chromatographs produced on the HPLC. However, peaks corresponding to derailment products

HTAL and PDAL were noticed in multiple extractions, and compound identity was confirmed by UV-VIS (Figure 11). No OA was observed in any of the samples. Due to the failure of this line of experimentation, little insight was gained with respect to IPTG concentration for optimization. Previous experimentation conducted regarding IPTG concentration and the CsTKS protein was able to provide me with a baseline amount that would be suitable for induction. This amount was determined to be 0.4 mM, and this concentration was used for all subsequent *E. coli* experimentation and protein induction. Detection of other relevant compounds in the HPLC-UV analysis of these samples was determined by comparison of unknown compound peaks in the sample to known compound peaks, elution time and UV-VIS spectra. OA standard chromatograms show the compound with a retention time of 4.38 minutes (Figure 12). HTAL standard chromatograms show the compound having a retention time of 2.73 minutes (Figure 12). PDAL standard chromatograms show the compound to have a retention time of 2.43 minutes (Figure 12).

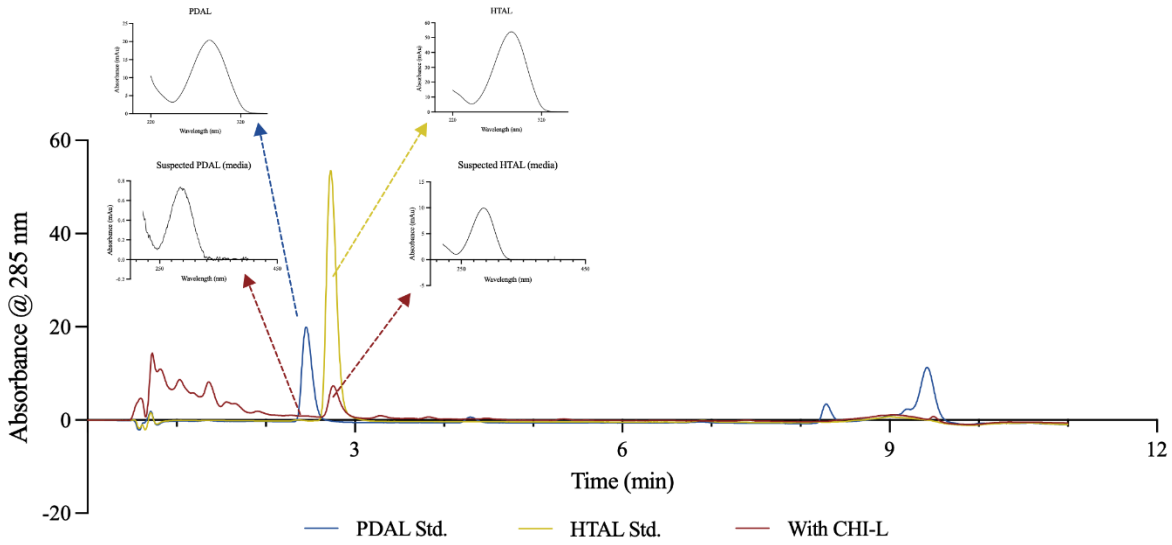


Figure 11. Chromatograms and UV-VIS to confirm presence of HTAL and PDAL from *E. coli* SBS feeding experiments

Chromatogram from HPLC-UV analysis of *in vivo E. coli* feeding experiment(s). Randomly selected *E. coli* feeding experiment sample to showcase confirmed presence of HTAL and PDAL in the sample (experimental group pCDF+CsAAE1+CsCHI-L/pCOLA+CsOAC+CsTKS/pLATE52+BjMCS). Included are authentic HTAL and PDAL standard injections and UV-VIS spectra, as well as UV-VIS spectra from relevant peaks in the extracted media sample.

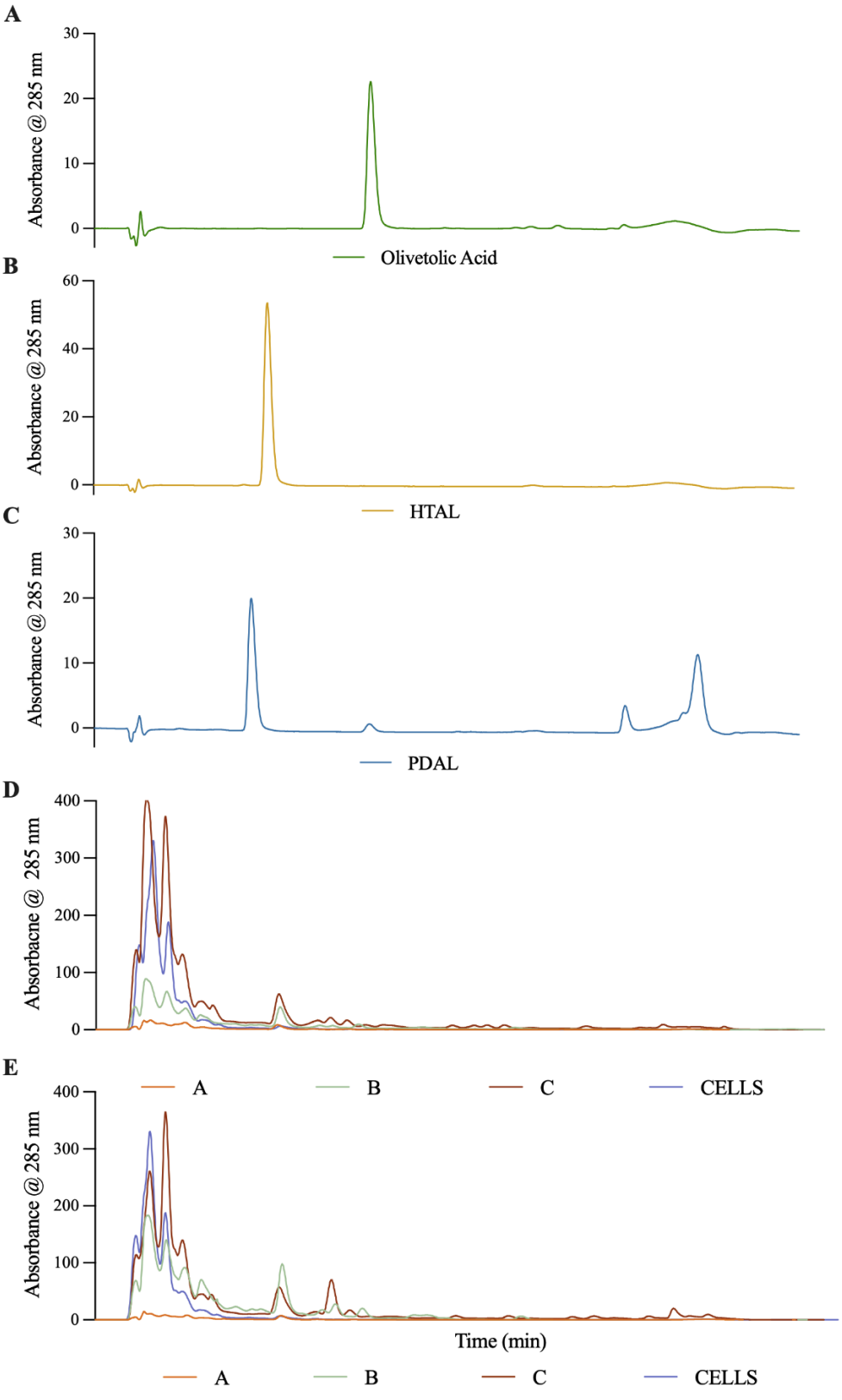


Figure 12. HPLC chromatograms from *in vivo E. coli* SBS experiments

A-C: Injections of authentic standards for hexanoyl triacetic acid lactone (HTAL), pentyl diacetic lactone (PDAL) and olivetolic acid (OA). Each standard corresponds to the production of one singular peak with consistent retention time. D: Experimental samples from *E. coli in vivo* feeding experiments expressing putative enzymatic proteins CsOAC, CsTKS, CsAAE1 and BjMCS. E: Experimental samples from *E. coli in vivo* experiments putative expressing enzymatic proteins CsOAC, CsTKS, CsAAE1 and BjMCS as well as expressing the proposed accessory protein, CsCHI-L. Experimental samples A to C (n=3) were extracted from spent culturing media as HTAL, PDAL and OA are all exported out of the bacterial cell. "Cells" are an extraction of the cell mass (n=1) to monitor for any OA accumulation in the bacterial cells. Enzymes and proteins: Cannabis sativa olivetolic acid cyclase (CsOAC), Cannabis sativa tetraketide synthase (CsTKS), Cannabis sativa acyl-activating enzyme (CsAAE1), Bradyrhizobium japonicum malonyl-CoA synthetase (BjMCS) and Cannabis sativa chalcone isomerase-like (CsCHI-L).

Feeding experiments conducted with a more simplified *E. coli* SBS (with relevant combinations of pCOLADuet1+*CsOAC*+*CsTKS*, pCDFDuet1+*CsAAE1*, pCDFDuet1+*CsAAE1*+*CsCHI-L*) lacking the additional pLATE52+BjMCS plasmid were then conducted. These experiments were analyzed using LC-MS/MS, and the presence of OA, HTAL and PDAL was detected in all samples (Figure 13). Included in SBS experimental combinations was an "unfed" group (CsOAC+CsTKS) that did not receive addition of the starter molecules to culture, and acted to demonstrate the ability of the systems to produce relevant compounds when starter molecules were not present.

The largest amount of OA was produced in a single run by CsOAC/CsTKS/CsAAE1/CsCHI-L A with a concentration of 0.225±0.00124 ppb and the smallest concentration was noted in CsOAC/CsTKS/CsAAE1 B at 0.124±0.00178 ppb (Figure 13). On average, the group with the largest accumulated OA was the CsOAC/CsTKS/CsAAE1/CsCHI-L experimental group, with an average of 0.184±0.00124 ppb and the lowest was observed in the CsOAC/CsTKS/CsAAE1 experimental group with an average concentration of 0.150±0.00178 ppb (Figure 13). Interestingly, the unfed control group

expressing CsOAC and CsTKS produced OA, with an average of 0.170±00.018 ppb. There was no significant difference noted in the amount of OA produced in the CsCHI-L-expressing samples when compared other experimental samples (Figure 13).

HTAL and PDAL were also detected in all *E. coli* SBS samples. The largest amount of HTAL produced was seen in the CsOAC/CsTKS unfed group, and was statistically significant when compared to the lower production of HTAL in the CsOAC/CsTKS/CsAAE1 experimental group. In all fed experimental samples, HTAL production was relatively similar with no significant difference. However, there was a significantly higher proportion of PDAL produced in the Unfed experimental group with an average value of 0.573±0.126 ppb when compared to the CsOAC/CsTKS/CsAAE1 experimental group average of 0.208±0.009 ppb. HTAL average concentration remained almost precisely the same across all groups, with no significant differences noted and averages of 0.155±0.00005 ppb in CsOAC/CsTKS Unfed, 0.155±0.0004 ppb in CsOAC/CsTKS, 0.156±0.002 ppb in CsOAC/CsTKS/CsAAE1 and 0.156±0.001 ppb in CsOAC/CsTKS/CsAAE1/CsCHI-L.

Signal to noise ratios for each samples were used to determine the validity of the concentrations reported by the LC-MS/MS (Supplemental tables 1-3).

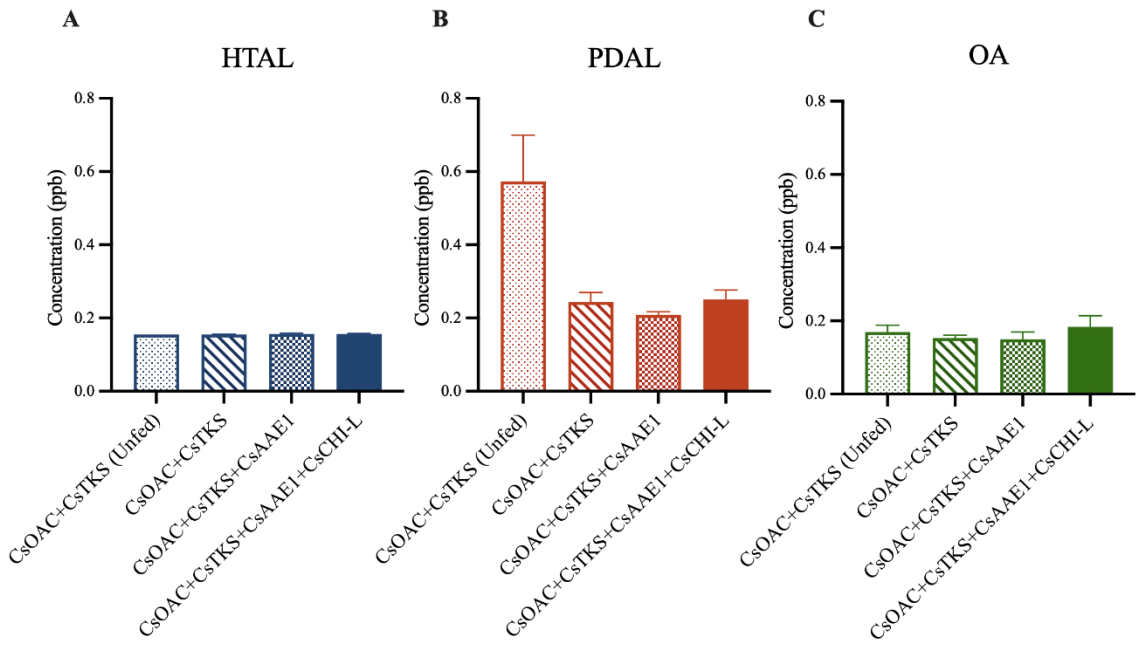


Figure 13. Compound production from *in vivo E. coli* SBS experiments.

LC-MS/MS analysis of extracted spent media from culture. A. Concentrations of hexanoyl triacetic acid lactone (HTAL) in different experimental scaffolds. B. Concentrations of pentyl diacetic lactone (PDAL) in different experimental scaffolds. C. Concentrations of Olivetolic Acid (OA) in different experimental scaffolds. Data subjected to one-way ANOVA (Brown-Forsythe and Welch test; Dunnett's multiple comparisons test), $n = 4$, error bars: CI = 95 %, p value ≤ 0.05Cultures fed 4mM sodium hexanoate and 12 mM malonic acid and allowed to react for 48 hours (n=4).

OA is not degraded by the culturing process and can be detected

In all samples, OA was detected on the HPLC as confirmed both by retention time and UV-VIS spectra. The OA standard has a retention time of 4.853 minutes, and OA in the spiked-fed samples had the same retention time (Figure 14). In the two random samples chosen to represent the OA-spike, the peak corresponding to the late-stage spike had a retention time of 4.803 and the peak corresponding to the early-stage spike sample had a retention time of 4.823 minutes (Figure 14). Compound identification was confirmed by UV-VIS absorbance spectra and comparison to the OA standard (Figure 14). Interestingly, a large peak can be observed

directly before the OA elutes off the column with a retention time of roughly 3.90 minutes. This peak is noticed in both the early and the late stage spike samples. UV-VIS spectra of the unknown component reveals that it is not OA or any other known compound of interest (). This large unknown helps to demonstrate the difficulty of achieving baseline separation in samples extracted from culture.

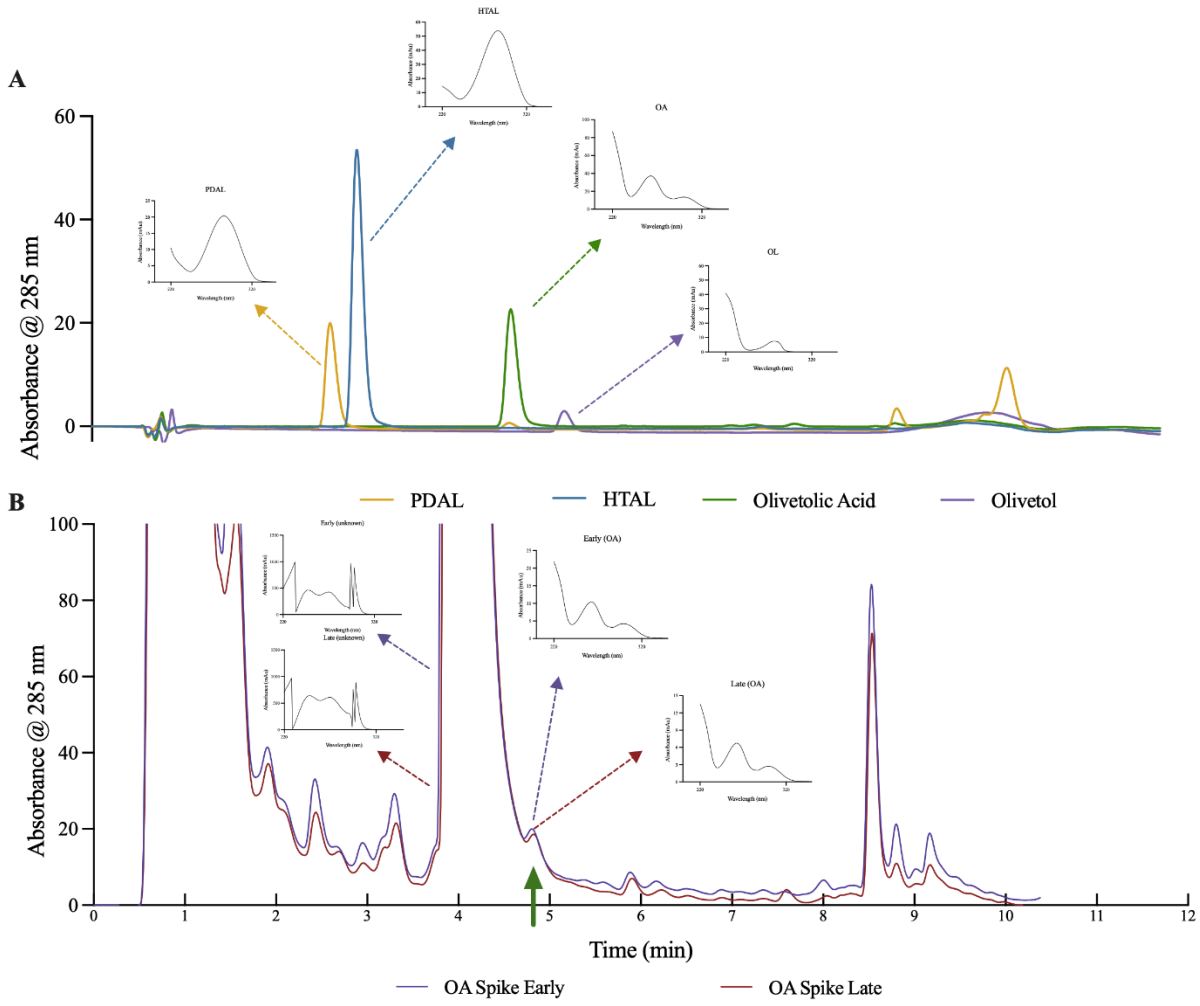


Figure 14. Presence of Olivetolic Acid in spiked experimental samples.

A. Injections of authentic standards for hexanoyl triacetic acid lactone (HTAL), pentyl diacetic lactone (PDAL) and olivetolic acid (OA). Each standard corresponds to the production of one singular peak with consistent retention time. Corresponding UV-VIS absorbance spectra of HTAL, PDAL and OA displayed next to appropriate standard peak. B. Chromatograms of cultures that received an olivetolic acid (OA) spike at 0.25mg/L^{-1} at the beginning of a 48 hour fed culture (OA Spike Early) or at the end (OA Spike Late) to ensure the ability of the extraction

and analytical process to correctly identify and monitor OA production. Green arrow indicates OA peak on chromatograph.

S. cerevisiae SBS are able to produce high levels of OA in vivo experiments

The first *in vivo* feeding experiments conducted with the yeast system did not yield the presence of any OA or relevant compounds. Through using only HPLC coupled to a diode array detector, I was not able to baseline separate spiked-in authentic standards of PDAL, HTAL, and OA with other unknown compounds produced by the yeast cells (data not shown). As such, I was not able to quantify any potential production of these compounds in the yeast system using this analytical approach.

To circumvent this problem, I analyzed further feeding experiments using LC-MS/MS, which detects and quantifies only compounds of interest while ignoring others and as such, baseline separation of all compounds is not necessary. Initially, this second round of feeding experiments with the yeast SBS did not yield the presence of any OA, HTAL or PDAL compounds (Figure 15). The legitimacy of these results is verified by the success of HTAL, PDAL and OA calibration curves run on the LC-MS/MS, where compound detection is possible at the 0.1 ppb level (Figure 15). All compound detection was below the values obtained for the "blank" runs on the LC-MS/MS, and values were well below the threshold that the 0.1ppb sample was observed at (Figure 15).

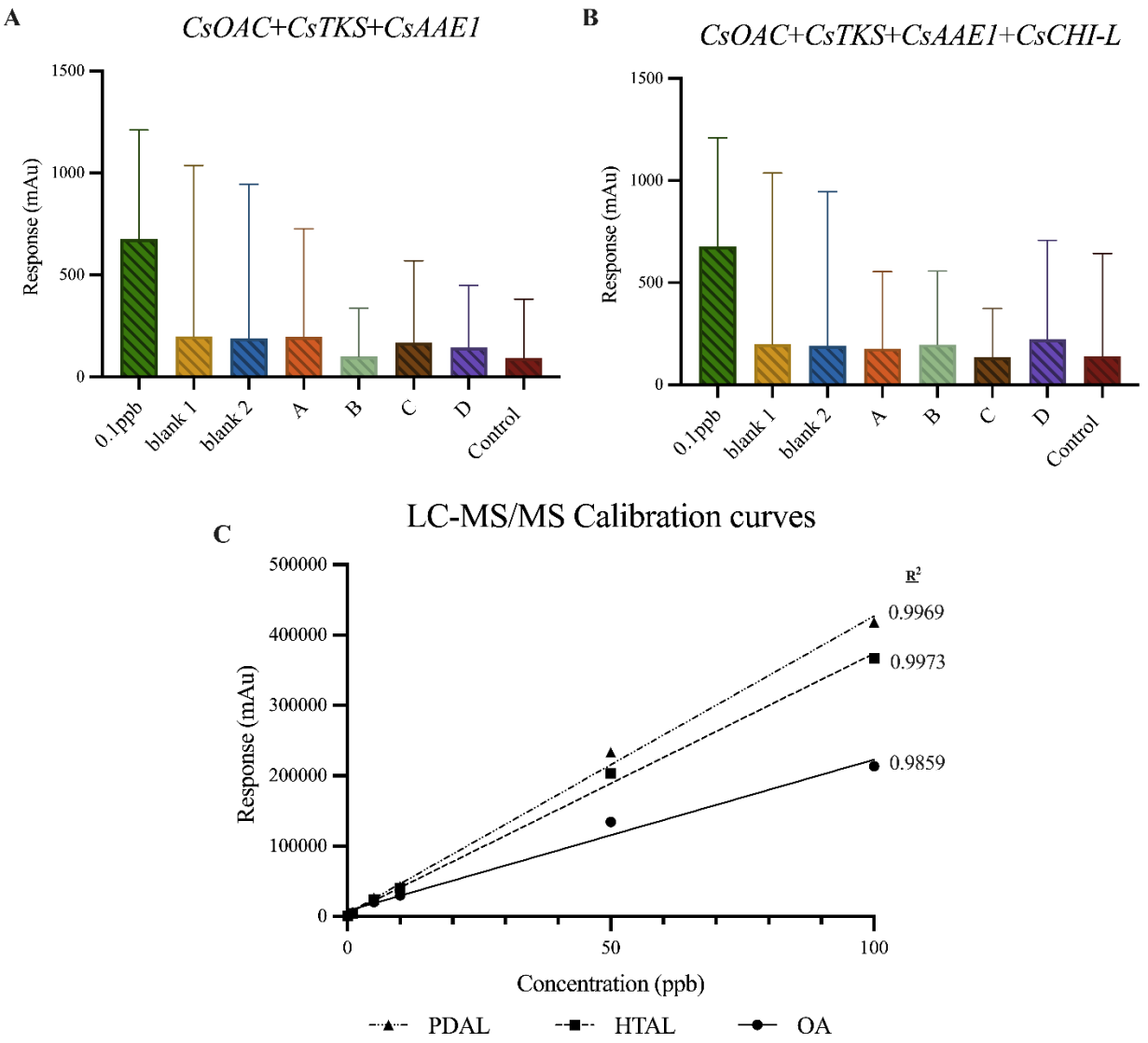


Figure 15. Response of LC-MS/MS to in vivo S. cerevisiae SBS culture extractions

A. Response from LC-MS/MS to varying samples or injections for *in vivo* experimental cultures harboring different gene combinations (*CsOAC+CsTKS+CsAAE1* and *CsOAC+CsTKS+CsAAE1+CsCHI-L*) to illustrate the sensitivity of the machine and verify the lack of OA production in culture by the *S. cerevisiae* SBS system. Cultures were fed 4mM hexanoic acid and 12mM malonic acid and allowed to react for 48 hours (n=4). Where 0.1 ppb is the lowest injected concentration in the calibration curve, blank1 and blank2 are blanks fed in between sample injections, A-D are experimental samples and Control is INVSc1 yeast cells fed in the same manner as expeirmetnal groups. B) Calibration curves of HTAL, PDAL and OA as determined by 0.1, 1, 5, 10, 50 and 100 ppb injections analyzed by LC-MS/MS. Ppb = parts per billions, HTAL is hexanoyl triacetic acid lactone, PDAL is pentyl diacetic lactone and OA is olivetolic acid.

It was due to this negative result that I re-assessed the plasmids and found that the pESC-HIS+CsOAC+CsTKS construct was lacking CsTKS and CsOAC inserts. After reconstructing this vector and re-building the system, I re-ran the feeding experiment. In this case, LC-MS/MS analysis revealed OA production in all experimental samples. In this set of experiments, a true control was introduced, where INVSc1 cells lacking the SBS constructs were cultured and fed alongside the experimental groups to monitor for any endogenous OA-producing activity of the *S. cerevisiae* cells. The largest amount of OA observed was in the CsOAC+CsTKS experimental group, with an average concentration of 59.6±21.2 ppb ppb OA, followed by 50.0±50.0 ppb in the CsOAC+CsTKS+CsAAE1 experimental group and ending with 43.0±32.9 ppb in the CsOAC+CsTKS+CsAAE1+CsCHI-L experimental group (Figure 16). These values are orders of magnitude higher compared to the amount of OA produced in *E. coli* by the experimental groups (Figure 13). OA production was not statistically significant between any of the experimental groups. The yeast control group produced very low levels of OA, with an average of 0.43±0.047 ppb, which was so low that the value does not yield a visually identifiable signal in Figure 16. The CsOAC+CsTKS experimental group was significantly different than the control group. Interestingly, the amount of OA produced in the yeast control is very similar to the amount of OA produced in the E. coli SBS, which was 0.17±0.1 ppb (Figure 13).

HTAL production in the CsOAC/CsTKS/CsAAE1 group was significantly higher when compared to the control, but relatively the same amounts of this bi-product was observed in all groups with the highest average value of 0.21±0.10 ppb in the CsOAC/CsTKS/CsAAE1/CsCHI-L group and the lowest average value of 0.11±0.02 ppb in the control group. These values are also similar to the amount of HTAL produced by the *E. coli* SBS in the unfed control and all experimental groups (Figure 16).

Both the CsOAC/CsTKS and the CsOAC/CsTKS/CsAAE1 groups showed statistically lower production of PDAL than the control group did (0.81±0.57 ppb and 1.22±0.71 ppb respectively, when compared with 5.56±2.82 ppb in the control). The CsOAC/CsTKS/CsAAE1/CsCHI-L group also accumulated much lower amounts of PDAL than the control (1.65±1.66 ppb) but the results between these two groups did not have any statistical significance. The highest average production of PDAL was in the CsOAC+CsTKS+CsAAE1+CsCHI-L experimental group and lowest in the CsOAC+CsTKS experimental group (Figure 16).

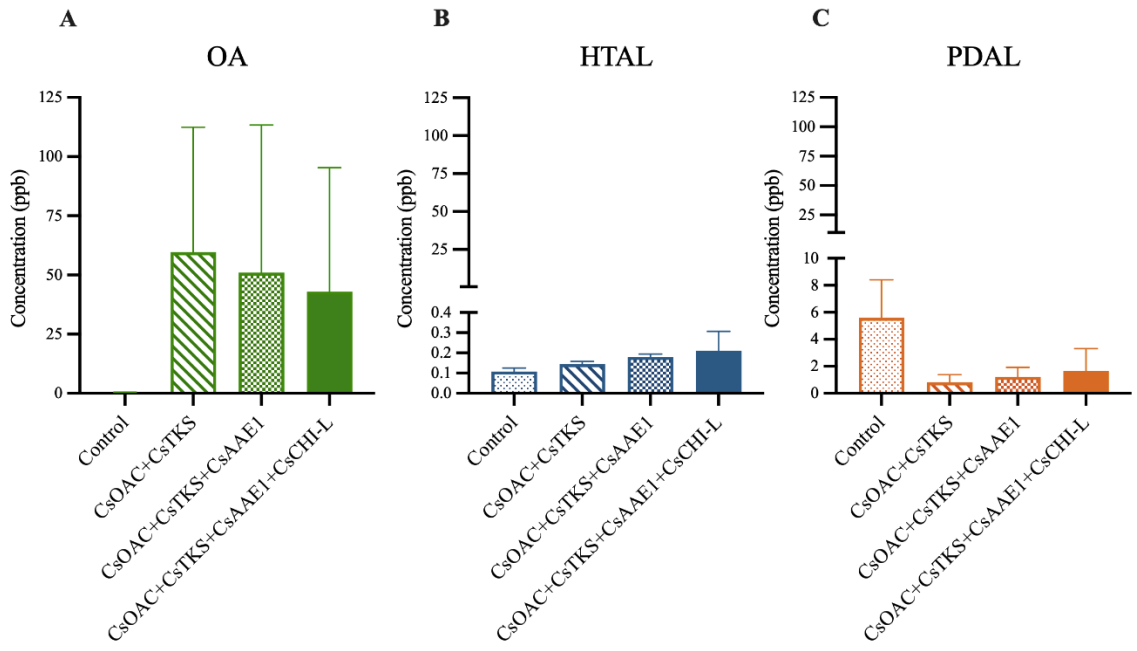


Figure 16. Compound production from *in vivo S. cerevisiae* SBS experiments.

LC-MS/MS analysis of extracted media and cells from culture. A. Concentration of Olivetolic Acid (OA) in different experimental scaffolds. B. Concentration of hexanoyl triacetic acid lactone (HTAL) in different experimental scaffolds. C. Concentration of pentyl diacetic lactone (PDAL) in different experimental scaffolds. Cultures fed 4mM hexanoic acid and 12 mM malonic acid and allowed to react for 48 hours. Data subjected to one-way ANOVA (Brown-Forsythe and Welch test; Dunnett's multiple comparisons test), $n = 4$, error bars: CI = 95 %, p value ≤ 0.05.

The accuracy and validity of these samples was confirmed by calculation of the limit of detection (LOD) and limit of quantification (LOQ) values that are noted to be below the smallest levels of compound concentration reported (Table 3). Signal to noise ratios for each samples were used to calculate the LOD and LOQ values (Supplemental tables 1-3).

Table 3. Limit of detection and quantification for LC-MS/MS.

Limit of detection (LOD) and limit of quantification (LOQ) values as determined for compounds OA, HTAL and PDAL in LC-MS/MS analysis. Limits determined from signal to noise ratio values from repeated known injections of 1 ppb (n=10). Values are given in ppb. LOD and LOQ values calculated using equations 2-4 detailed in Methods.

	OA	HTAL	PDAL
LOD	0.022	0.016	0.017
LOQ	0.074	0.055	0.056

LC-MS/MS Standard Analysis and MRM

Samples run on the LC-MS/MS were quantified using multiple reaction monitoring (MRM). In this process, the sample is pumped through a stationary phase via dissolution in a high pressure, mobile phase (liquid chromatography, LC). Variation in the migration rates through the LC column allow for initial separation of compounds. Once the compounds are eluted from the stationary phase, the analytes are subjected to electrospray ionisation (ESI), where the ionization source is able to nebulize and ionize the sample to create charged particles. The particles pass under high vacuum pressure through three quadrupoles. The first quadrupole is set to a specific mass/charge ratio to pass the desired ion into the next quadrupole. This step excludes all other particles from entering the quadrupoles. The next quadrupole, termed the collision cell, fractures the parent ion into daughter (or product ions) via collison with an inert

gas. The third quadrupole is then able to target the specific daughter ions, which are quantified with an electron multiplier. MRM refers to the ability of the LC-MS/MS to monitor specific daughter fragments from a parent ion, and it is this process that allows for such a high level of specificity and detection in the instrument. All compounds analyzed in this experimentation on the LC-MS/MS (Cinnamic Acid, OA, HTAL and PDAL) fragment from one parental ion into two daughter ions. MRM data for the fragmentation of these compounds is detailed in Table 4.

Table 4. LC-MS/MS analysis of HTAL, PDAL, OA and Cinnamic Acid standards

Compound	Precursor Ion	Product Ion 1	CE 1	Product Ion 2	CE 2	Cell voltage	Polarity
HTAL	223.1	125	12	57.1	32	5	Negative
PDAL	181.1	139.1	20	137.1	12	5	Negative
Olivetolic Acid	223.1	179.1	16	137.1	24	5	Negative
Cinnamic Acid	149.06	65.1	16	39.1	16	5	Positive

Discussion

Initial E. coli synthetic biology system in vivo experiments do not produce OA

Extracts from lysed cells and spent media from the first attempt to produce OA *in vivo* in the *E. coli* SBS were analyzed by HPLC (Figure 10). Although I noted the presence of derailment products HTAL and PDAL, no OA was present in the samples. It was already known from the *in vitro* assays previously conducted that the proteins being expressed are in fact capable of forming OA when given the appropriate starter molecules (Figure *7*).However, it was possible that the cells were producing OA, but a catabolic pathway was consuming the OA below the limits of detection by and HPLC instrument. To test this hypothesis I conducted a set of OA-spiked feeding experiments, following the same protocol used for the initial experimentation, to determine if *E. coli* has the ability to degrade. OA was observed in both early stage and late stage sample spikes (Figure 14), indicating that OA is not degraded by the

culturing/feeding process or by the extraction process. Given that there is likely no OA catabolic pathway in *E. coli*, the lack of an OA signal in the HPLC chromatograms could be explained by either that no OA was produced by the SBS, or that the amount of OA produced in the *in vivo* system was so small that it was below the limit of detection for the HPLC apparatus that I used. The amount of OA that I added to the OA-spiked cultures was at a concentration of 0.25 mg/L, which is comparable to concentrations produced in other *E.coli*-based *in vivo* SBSs.[121] This concentration was only achieved following optimization of the system, and it is realistic to think that my system would not yet be capable of producing OA at this level. Additionally, when looking at the OA-spike sample chromatograms, the OA peak is very small and almost imperceptible on the full-scale chromatogram (Figure 14). This is indicative that even at 0.25mg/L, the HPLC-UV has difficulty registering the signal from OA presence. Following this realization, subsequent *E. coli* feeding experimental samples were analyzed by LC-MS/MS, which has a much higher sensitivity and can provide more accurate data regarding compound presence and amount. .

Plasmid incompatibility in the E. coli SBS (Duet1 system)

Potential difficulties were encountered with respect to plasmid compatibility in the *E. coli* host. Following the failure of the first feeding experiments, I decided to remove the pLATE52+BjMCS plasmid from the system. pLATE52 was not part of the originally planned system, and is not designed to specifically work alongside the Duet1 expression vectors. pLATE52 has a pMB1 origin of replication (ori), which is a type of ColE1 replicon. Other plasmids in the *E. coli* system are pCOLA, which has a ColA ori and pCDF, which has a CloDF13 ori. Although ColE1, ColA and CloDF13 are not directly incompatible, plasmid interactions or issues due to copy number may have influence this iteration of the *E. coli* system.

Additional incompatibility issues may have arisen due to other characteristics of the plasmids. pCOLA and pCDF are both medium copy number plasmids, producing roughly 20 to 40 copies per cell. On the other hand, pLATE52 is a low copy plasmid, due to the presence of the repressor of primer (ROP) protein that regulates plasmid copy number. ROP is a small dimeric protein that increases the affinity between the complementary RNA I Replication inhibitor and the RNA II replication activator. The ROP protein functions as a global copy number regulator, where introduction into a multi-plasmid system will lower the copy number of each group. As we know that the Duet1 system plasmids (pCOLA and pCDF) are not high copy number plasmids to begin with, ROP presence may have effectively lowered the plasmid number in the cells below a threshold that is capable of substantial protein expression. This draws into question the validity of the first *in vivo* feeding trials performed with the *E. coli* system, as production of relevant proteins may have been hindered by competing for the same machinery required for DNA and plasmid replication. Subsequent experimentation was done with a simplified *E. coli* SBS, where a maximum of two expression vectors (pCOLA-Duet1 and pCDF-Duet1) were included. While the BjMCS enzyme may allow for a higher internal production of malonyl-CoA, it is not necessary for successful production of OA in *E. coli*.[121]

E. coli synthetic biology system in vivo experiments are able to produce OA. HTAL and PDAL

Following re-construction of the *E. coli* SBS without pLATE52-BjMCS, feeding experiments were once again conducted, extracted and analyzed. As mentioned previously, LC-MS/MS was used to analyze the results of these experiments. HTAL, PDAL and OA were all present in all samples analyzed at relatively similar levels with the exception of large amount of PDAL in the unfed experimental group (Figure 13). In this experiment, the “unfed” group consisted of cells carrying *CsOAC* and *CsTKS* that were induced with IPTG, but were not fed with hexanoate and

malonate. In experiments carried out by Tan et al., *E. coli* constructs lacking a fed starter substrate (hexanoate) but incorporating expression of the CsTKS and CsOAC enzymes also demonstrated the ability to produce OA *in vivo*, suggesting that there is sufficient pools of hexanoyl-CoA and malonyl-CoA *in vivo* that can be utilized by CsTKS and CsOAC to make small quantities of OA.[121] The addition of hexanoate or malonate to the cultures, nor the inclusion of CsAAE1 to produce extra hexanoyl-CoA, altered the abundance of any of the products. The lack of increased compound production when with fed starter substrates can possibly be attributed to the lack of available endogenous malonyl-CoA, as hexanoyl-CoA can be produced from fed starter molecule sodium hexanoate, but intrinsic malonyl-CoA production is limited due to lack of endogenous malonyl-CoA synthase activity.

The only statistically significant difference in this experiment was that the quantity of PDAL was higher in the unfed control group expressing CsTKS and CsOAC. This finding will be discussed in the yeast SBS discussion.

As there was no LOD and LOQ estimation when *E. coli* samples were run on the LC-MS/MS, approximation of signal to noise (S/N) ratios using the accepted standard of 10 or higher can be used in place. When looking at S/N values given by the *E. coli* samples, we can see that most S/N values are indeed >10, with two exceptions (Supplemental Figures 1-3). These values should therefore be counted as “no signal” in consideration for the calculations of OA, however removal of these values does not change the overall trend of the experiment or the significance. Due to the high level of S/N values observed for OA samples in the *E. coli* SBS samples, I can conclude that the concentrations generated by the LC-MS/MS for this compound are accurate. However, the amounts noted are extremely low, indicative that the system is operating in an extremely inefficient capacity, with high bi-product formation and low overall

production of intermediate OA. On average, the three experimental groups produced 0.000153 mg/L (CsOAC+CsTKS), 0.000150 mg/L (CsOAC+CsTKS+CsAAE1) and 0.000184 mg/L (CsOAC+CsTKS+CsAAE1+CsCHI-L). This amount of OA can be seen as almost negligible, as titres around 0.1mg/L are even considered low for this type of system.[121] We can also note that the production of OA in the *S. cerevisiae* SBS Control group (lacking GOI's and pertinent enzymes) is very similar to the level of OA produced in the *E. coli* experimentation.

There was no significant difference noted between the amount of compound present in cultures that had been expressing the CsCHI-L protein and those that had not. This is indicative that there is no beneficial interaction between the proteins, or with respect to the intermediate, that would be facilitated by the CsCHI-L. However, it is possible that the CsCHI-L does not exert this function in the *E. coli* cellular environment.

Macromolecular concentration, including proteins and RNAs, is often extremely high in the bacterial cytoplasm and can cause macromolecular crowding in *E. coli*.[168,169] This crowding can hinder protein diffusion and influence protein association or kinetics, as well as gene expression through translational limits imposed by tRNA hinderance.[170,171] The hypothesized accessory protein, CsCHI-L, could be experiencing restricted diffusion or altered protein-protein interactions in a prokaryotic cellular environment.[172]

Another disadvantage of expressing eukaryotic proteins in prokaryotic cells is the lack of post-translational modification – including phosphorylation, glycosylation, disfulfide bridge formation and even proper folding of the protein.[173] Multiple post-translational modifications are often require for eukaryotic proteins to achieve their active conformation.[174] Many characteristics of eukaryotic proteins can be effected by post-translational modification, including their hydrophobicity and charges, which in turn lead to changes in protein stability and solubility.[175] It

is possible that the CsCHI-L protein, although seemingly expressed properly in our BL21 protein expression experiments, is lacking important post-translational modifications that are necessary for the proteins action in the pathway. CHI and CHI-L proteins share a common folded structure.[37,42] A cleft fold, occupied by fatty acids, has been an important evolutionary step that lead to the Type III CHI/CHI-L proteins.[176] Subsequent evolutionary modification has given rise to other CHI-L enzymes, where an increase of sterically restriction surrounding the ligand-cleft and polar residues that facilitate hydrogen-bond interaction with substrates impart different action and specificity of the protein.[42] Alterations to the folding style of CsCHI-L accessory proteins is therefore very important to their subsequent actions in the cell, and improper folding may easily result in an inactive protein.

E. coli heterologous protein production is also prone to misincorporation of amino acids at particular codons, especially if that codon is rare in *E. coli*. For example, the AGA arginine codon (rare in *E. coli*) is often misinterpreted and lysine incorporated instead at an error rate of 36 to 42%.[177] This type of error was considered when designing the system and corrected for by codon-optimization, as discussed previously. However, it is still possible that misinterpretation of the heterologous coding sequence by the *E. coli* translational machinery occurred and impacted the activity of the protein.

As noted previously, OA production in the engineered *E. coli* SBS system was very low and can be considered an extremely inefficient pathway with negligible compound production. This could be attributed in part to the *E. coli* strain used to construct the system, BL21. BL21 is a bacterial strain commonly used for high-level recombinant protein production. However, our SBS relies heavily on the ability to convert fed starter compounds into intermediate or end products. While in theory BL21 cells should be able to convert the 4mM sodium hexanoate and

12mM malonic acid into hexanoyl-CoA and malonyl-CoA, this is not necessarily the case. An additional route of synthesis is to include either a fatty acid biosynthesis pathway or a β-oxidation reversal (r-BOX) pathway in the host strain for more efficient production of fatty acids (hexanoic and malonic acid in this case). In the experimentation done by Tan. et al, integration of the r-BOX pathways alongside the OA biosynthesis pathway into *E. coli* strain JST10 resulted in a 30-fold higher production of OA when compared to production using the same pathway in BL21 cells [121]. Additionally, OA production was increased 2.3 fold when only the OA biosynthesis pathway was included into JST10 cells when compared to the BL21 cells, indicative of a more efficient pathway[121]. Additionally, as the BjMCS was not properly planned and effectively incorporated into our SBS, production of malonyl-CoA would also have been minimal and hindering efficient OA production. For the *E. coli* SBS to function efficiently, these problems regarding the starter molecules and fatty acid production in the cells must be addressed.

Yeast cell degradation for analysis of compounds in S. cerevisiae SBS

Extraction of compounds from the yeast cultures requires some additional steps when compared to extracting the *E. coli* cultures. While bacterial cell walls are composed mainly of peptidoglycan, yeast cell walls are more complex and are comprised largely of β-1,3-glucan, with smaller amounts of chitin, β-1,6-glucan and mannoproteins.[178,179] Additionally, experimentation conducted with *E. coli* cultures has noted the presence of OA and other by-products of this pathway (HTAL, PDAL, OL) as being present in the media or supernatant following cells harvesting, not in the cells themselves. This is indicative that the compounds are either being transported out of the cells, or are being released due to cellular lysis in response to accumulation of said compounds.[121] Experimentation done with OA-producing steps facilitated

in yeast culture has required the physical lysis of the cell by both physical and enzymatic degradation to release compounds prior to extraction.[120] Following the feeding experimentation, I treated the yeast cultures with 2U/OD of zymolase (Lyticase, ThermoFisher) from *Arthrobacter luteus*. Zymolase degrades yeast cell walls through β-1,3-glucan laminaripentaohydrolase which hydrolyzes polymers with β-1,3-bonds.[180] Zymolase also has protease and mannase activities, which degrade additional components of the yeast cell wall.

To help with lysis of the yeast cell wall, sonication pulses were also applied to the culture. Sonication creates an intense shear action by high-frequency sonic waves. The mechanism behind this is cell wall disintegration through acoustic cavitation, in which the formation and expansion of gas and vapor bubbles are triggered by the high-frequency waves.[181,182] The collapse of said bubbles releases a large amount of energy, which is converted to mechanical energy in the form of shock waves, imparting a high pressure on the cell and cell wall. In my experimentation, cells were frozen following lyticase treatment, then thawed and sonicated thoroughly. The combination of enzymatic (lyticase) and mechanical (freeze-thaw and sonication) degradation of the cell wall is more than sufficient to release compounds of interest from the yeast cells.

S. cerevisiae synthetic biology system was not able to be analyzed by HPLC-UV

Due to equipment and software error, chromatogram and UV-VIS data from the *S. cerevisiae* feeding experiments run on the HPLC-UV were compromised and cannot be retrieved for visual representation in this project. However, the initial testing of the *S. cerevisiae* system did not produce any OA, as confirmed at the time of testing. Baseline separation of compounds on the HPLC was particularly difficult to achieve for extractions from the yeast cultures and experience a lot of "junk" or unidentified peaks eluting off the column within the first couple

minutes of the run. This led to the development of a new method for sample analysis on the HPLC-UV, with a longer gradient and slower flow rate. Due to problems related with HPLC reliability and ability to analyze the *S. cerevisiae* samples, all subsequent samples were analyzed by the LC-MS/MS

S. cerevisiae synthetic biology system produces OA in vivo

Analysis of initial *S. cerevisiae* feeding experiments by the LC-MS/MS revealed no data for the presence of any relevant compounds – HTAL, PDAL or OA (Figure 15). This was especially concerning, as even if the system was not optimally producing OA, we would expect to see some compound development, at least with respect to derailment products. Additionally, the LC-MS/MS has a much higher sensitivity for compound detection when compared to the HPLC, and even negligible amounts of compounds should have been detected. Following this realization, I re-checked all components of the *S. cerevisiae* SBS for any malfunctioning or unexpected parts. Constructs were examined to make sure they contain the proper inserts and were able to express our proteins of interest. During this investigation, it was discovered that the pESC-HIS+*CsOAC*+*CsTKS* plasmid was no longer harbouring the relevant *CsOAC* or the *CsTKS* sequences, meaning these proteins were never expressed in the experiments. I am currently unsure if the freezer stock was rendered not viable by the move undergone by our lab (and switching of -80 storage), or if the inserts were kicked out of the plasmid construct due to other causes. As these two enzymes are the major components of the OA-forming steps of the pathway, this was a major problem and required reconstruction of said plasmid to properly incorporate and express the CsOAC and CsTKS proteins. Difficulty was encountered whilst trying to rebuild the vector using the typical ligation procedure that I had used before, possibly due to reagent viability. A different approach, Gibson assembly, was then used to recreate the

pESC-HIS+*CsTKS*+*CsOAC* vector. Gibson assembly takes advantage of the natural enzymatic activity in DNA construction, and uses the 5' exonuclease and 3' extension capabilities of DNA polymerase, as well as the DNA ligase activity to piece together complimentary overlapping oligonucleotide sequences generated by PCR.

With my construct newly rebuilt and all insert sequences verified for the yeast expression system, a new round of feeding experimentation was conducted with the *S. cerevisiae* cultures. This experiment followed the same protocol as previously used for the yeast cultures for both feeding, degradation and extraction. *In vivo* experimentation with the newly constructed *S. cerevisiae* SBS proved successful in producing high levels OA for all experimental groups, with no statistical significance shown with respect to amounts produced by different groups (Figure 16). Compared to the small amount of OA production in the control group of 0.428±0.047 ppb, the amount of OA produced in the experimental group was orders of magnitude higher.

The *S. cerevisiae* SBS also produced HTAL, albeit at very low quantities that were similar to that produced by the *E. coli* system. The only experimental group that accumulated a significantly higher level of HTAL compared to the control was CsOAC+CsTKS+CsAAE1. This observation, coupled with the high production of OA in this experimental group, suggests that all of these enzymes are operational within the yeast SBS.

Like in the *E. coli* system, the yeast control group produced a small amount of both OA and HTAL. However, the controls in these two systems were different. As previously discussed, the *E. coli* control consisted of IPTG-induced cells producing CsTKS and CsOAC and as such, the small quantity of OA and HTAL produced could be from these enzymes using endogenous pools of hexanoyl-CoA and malonyl-CoA. However, the yeast control consisted of cells that did not contain any of the cannabinoid biosynthetic genes, and as such it is difficult to rationalize

how these cells could have produced OA, HTAL, and PDAL given that it is very unlikely that wild-type yeast intrinsically possess these enzyme activities. An alternative explanation for these observations is the possibility of the mass spectrometer source was contaminated with residual amounts of OA, HTAL, and PDAL.[183]

It is known that OA can contaminate an electrospray ionization source, giving low-levels of OA signal regardless of the sample analyzed. Indeed, when OA was first detected in cannabis trichome extracts, the scientist who performed this analysis complained to Dr. Stout for months about the contamination of his source by the OA authentic standard. Given that, in this experiment, the calibration curve samples (up to 100 ppb) were analyzed before the control and experimental samples in both the *E. coli* and yeast SBS experiments, it is possible that the authentic standards used to produce the curves contaminated the source. If this was the case, then experimental and control samples that had low values in the 0 to 0.5 ppb range may simply be derived from this contamination. This potential source contamination could have been quantified by simply running a blank methanol injection after the samples to produce the curve. Any signal for OA, HTAL, and PDAL in a blank sample would indicate source contamination. Unfortunately, no such blank runs were performed in these experiments. Regardless, if the values are in the low 0 to 0.5 ppb range are indeed from source contamination, this would mean that a.) the *E. coli* SBS did not produce any OA, HTAL, or PDAL at all, and b.) the higher concentration of HTAL in the *E. coli* unfed control sample may be from it being run directly after the samples to produce the calibration curve (and thus may have had higher levels of contamination; Figure 13). Similarly, the levels of HTAL detected in the yeast system may have been from source contamination, although at least in the yeast system the enzymes were operating given the high amount of OA produced in this system.

In the yeast SBS, all experimental groups produced low-levels of PDAL ranging from 0.81 to 1.7 ppb (Figure 16). These observed levels could also be (partially) due to source contamination, or could represent bona fide PDAL being produced by the introduced enzymes. The truly perplexing anomaly in the yeast SBS dataset is that the control group produced higher quantities of PDAL compared to the experimental group. Indeed, the control cells should not be producing any appreciable amounts of any of the compounds. While the amount of OA and HTAL produced in the control can be potentially explained by source contamination, the control group produced on average 5.56 ppb PDAL. This may represent severe source contamination that was reduced by an order of magnitude when the next set of samples were run. An alternative explanation is that yeast endogenously produce an unrelated compound that both has the same mass as PDAL and produces the same daughter ions as PDAL when fragmented in a mass spectrometer. This possibility is reasonable considering that one of the daughter ions used to quantify PDAL is a mass of 137 (Table 4), which corresponds to a loss of an aldehyde functional group. Thus, any molecule with a mass of 181 that loses an aldehyde would produce an MRM signal in this setup. This is further exacerbated due to the fact that the quadrupoles are not set at exact masses (e.g. 181.1), but instead are usually set at ± 0.2 mass units to account for drift in the instrument. Thus, in this example, the first quadrupole would allow any molecule with a mass between 179.9 and 181.3 into the collision cell.

As part of a more thorough analysis for the latter *in vivo S. cerevisiae* SBS experimentation, LOD and LOQ values were determined for the analytical procedure carried out with all respective compounds. The LOD and LOQ for OA were calculated as 0.022 ppb and 0.074 ppb, respectively (Table 3). As these values are much lower than the reported data given

for OA concentration in the *in vivo S. cerevisiae* SBS samples, I can be certain that the concentrations given by the LC-MS/MS are accurate.

Comparison of E. coli and S. cerevisiae SBS in vivo production of OA

When looking at the amount of OA produced by each system, we can see that there is a significantly larger production in the *S. cerevisiae* system than in the *E. coli* system (Figure 17). Maximum OA production in the *E. coli* system was observed at 0.0002254 mg/L^{-1}, where in *S. cereviaise* maximum OA production was seen at 0.21974 mg/L^{-1}, almost 1000-fold higher (Figure 17). It is important to note that in these experiments, "Control" is not the same in the *E. coli* as in the *S. cerevisiae* experiments. In *E. coli*, the "control" group included expression of the *CsOAC+CsTKS*, without the addition of fed starter compounds 12 mM sodium hexanoate and 4 mM sodium malonate. In *S. cerevisiae*, the true control group is represented by INVSc1 cells, with no inclusion of plasmid cassettes or of any GOI's, that were fed the same starter compounds as the experimental groups (4 mM Malonic and 12 mM hexanoic acid). Due to this discrepancy, direct comparison between amounts produced by the control groups and experimental groups in different experiments is not possible. With that noted, production of OA in the control group of the *S. cerevisiae* experiments was higher than that of OA produced by any of the *E. coli* experiments, putting in question the validity of the OA production in said *E. coli* experiments. It can be concluded that any OA production in the *E. coli* system is minimal, and representative that the expression system does not function properly in this cellular environment.

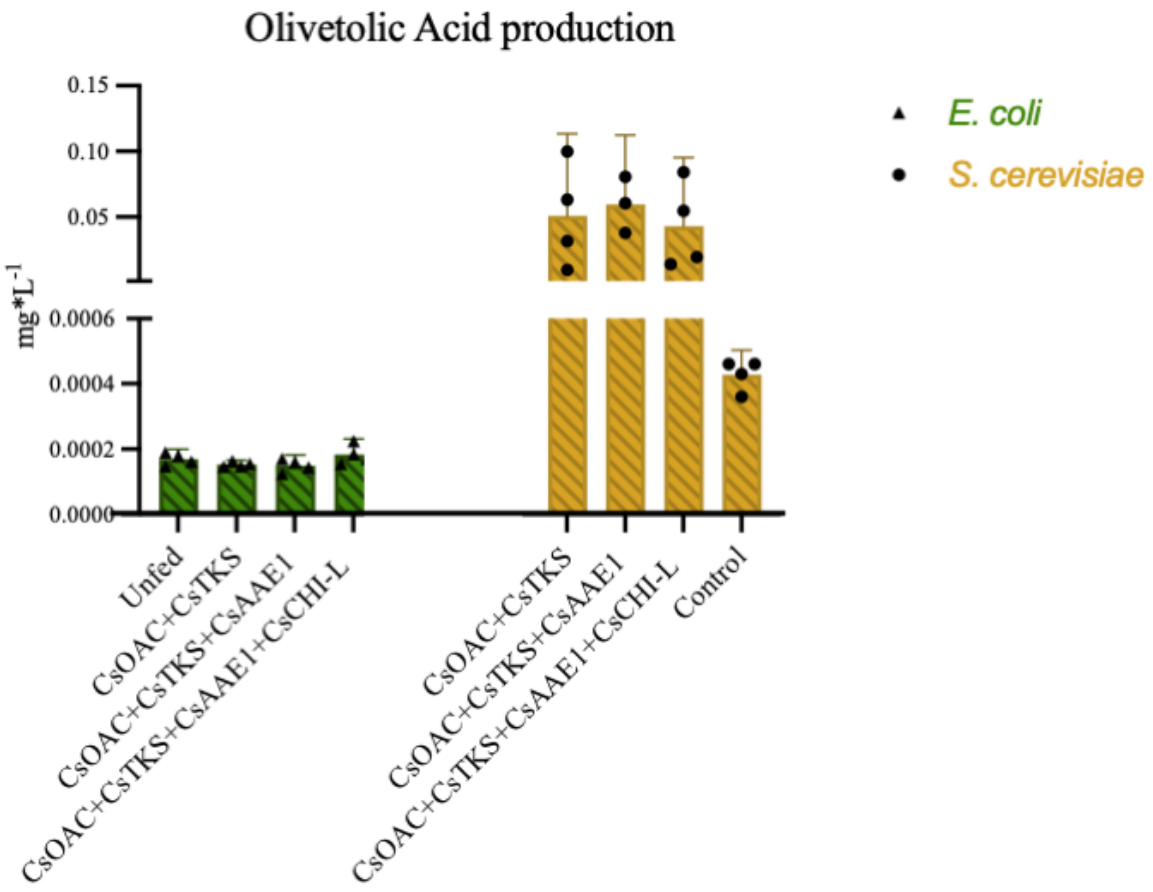


Figure 17. Comparison of Olivetolic acid production in SBS systems

Comparison of LC-MS/MS analysis OA production (expressed in concentration of $mg*^{L-1}$ from *in vivo* experimentation in *E. coli* and *S. cerevisiae* SBS. Data subjected to one-way ANOVA (Brown-Forsythe and Welch test; Dunnett's multiple comparisons test), $n = 4$, error bars: CI = 95 %, p value ≤ 0.05.

Chapter Four: Conclusions and future directions

Introduction

When considering the data presented in chapter 3, the results are indicative that the CsCHI-L protein does not function in the OA forming steps of the cannabinoid biosynthetic pathway. Although some level of OA production was observed in both SBS samples, there was no significant difference in the amount produced by any of the constructs. Additionally, higher presence of PDAL in systems containing the hypothesized protein demonstrates that inclusion of said protein does not rectify the pathway by lowering bi-product formation. However, due to the limited success and replicates of feeding experiments, further investigation in this area must be conducted to definitively assign a function (or lack thereof) to the hypothesized CsCHI-L protein.

Additional steps were then taken to further investigate the hypothesized protein, outside of the SBS. As CHI-L proteins are anticipated to be the precursor to CHI proteins and evolved to have the observed catalytic function, I performed a phylogenetic analysis of our CsCHI-L candidate alongside other known CHI and CHI-L proteins.

While phylogenetic relationships can inform of us about the evolutionary past and relationships between enzymes, I wanted to investigate possible mechanisms of action in bona fide CHI enzymes compared to our hypothesized CHI-L protein by analyzing a published crystal model of a CHI and a homology model of CsCHI-L. Sequences differentiation between the different residues involved in the active site formation are predicted to be largely responsible for the difference in action, however additional factors, including differential protein folding, could be responsible. For further protein structure investigation, I built a homology model of the CsCHI-L protein.

Materials and Methods

Phylogenetic tree construction

The CsCHI-L, Arabidopsis CHI (At3g55120), and Arabidopsis FAP1 (AT3G63170) amino acid sequences used as query sequences using BlastP to retrieve homologous sequences from the Phytozome database (https://phytozome-next.jgi.doe.gov/). Sequences below a cutoff of 1 e-10 were kept as putative homologs. Sequences that were obvious fragments were discarded, and for those with more than one variant, that with the lowest e-value was retained.

All subsequent analysis was performed using the MegaX software package. Protein sequences were aligned using the Muscle alignment algorithm. Using this alignment, the evolutionary relationships were inferred using the Maximum Likelihood method, and a final consensus tree was constructed from 1000 bootstraps of the initial tree.

Homology modeling of CsCHI-L

The CsCHI-L protein homology model was constructed using SWISS-MODEL software and the proposed AlphaFold prediction model as a template. This program predicts the 3D structure of a protein based on similarity to related proteins and a tertiary shape that results in the lowest energy conformation and is based on sequence input. This model was largely used to investigate the structure of the CsCHI-L protein when compared to other bona fide CHI proteins, and to check for similarities or differences in the CHI-fold and active site. An additional model was also built based on close homology to *Arapidopsis thaliana* FAP1.

Results

Phylogenetic relationships between CsCHI-L and other CHI and CHI-L proteins

Largely, we can note specific groupings in the tree representative of the different classes of CHI and CHI-L proteins, Types I-IV. Type III CHI-L proteins can be sub-divided into the FAP1, FAP2 and FAP3 clades. Type III CHI-L proteins are present in all organisms; from single-cell green algae through to the flowering plants. A singular out grouping of the *Chlamydomonas reinhardtii* CHI-L is noted from the Type III FAP1 proteins, indicating a diverged evolution for this algal protein. The Type III FAP3 grouping only has two representatives in this analysis, from *C. reinhardtii* and *Physcomitrella patens*. In the Type III FAP2 clade, we can see monophyletic grouping of similarly related organisms; the grasses of Poacea (*Oryza sativa*, *Zea mays*), the moss and lycophytes (*Selaginella moellendorffii*, *P. patens*) and the flowering plants (*A. thaliana*, *Medicago truncatula* and *Solanum tuberosum*). These types of groupings are also noted within the Type IV CHI's, where *P. patens* and *S. moellendorffii* cluster as a small outgroup while *S. tuberosum* and *H. lupus* group together alongside the *Z. mays* and *O. sativa* clade. When looking at Type I CHI enzymes, the groupings get even more specific where each species (*Z. mays*, *O sativa*, *A. thaliana* and *M. truncatula*) form individual groups. Distinct outgroups are observed when looking at the Type I CHI proteins, where four CHI proteins from *M. truncatula* and two CHI proteins from *S. tubersoum* partition into their own classes outside of the Type I group. The anticipated outgroup protein for the entire phylogenetic tree, *Eubacterium ramulus* CHI, is present as part of the Type III (FAP1) CHI's even though it has bona fide catalytic activity.

Our candidate protein, CsCHI-L, is grouped in the FAP1 clade of the Type III CHI-L proteins. The most direct relation to said candidate is the *H. lupus* CHI-L, HlCHIL1. These proteins share a 69% identity and both are found in members of *Cannabacea*.

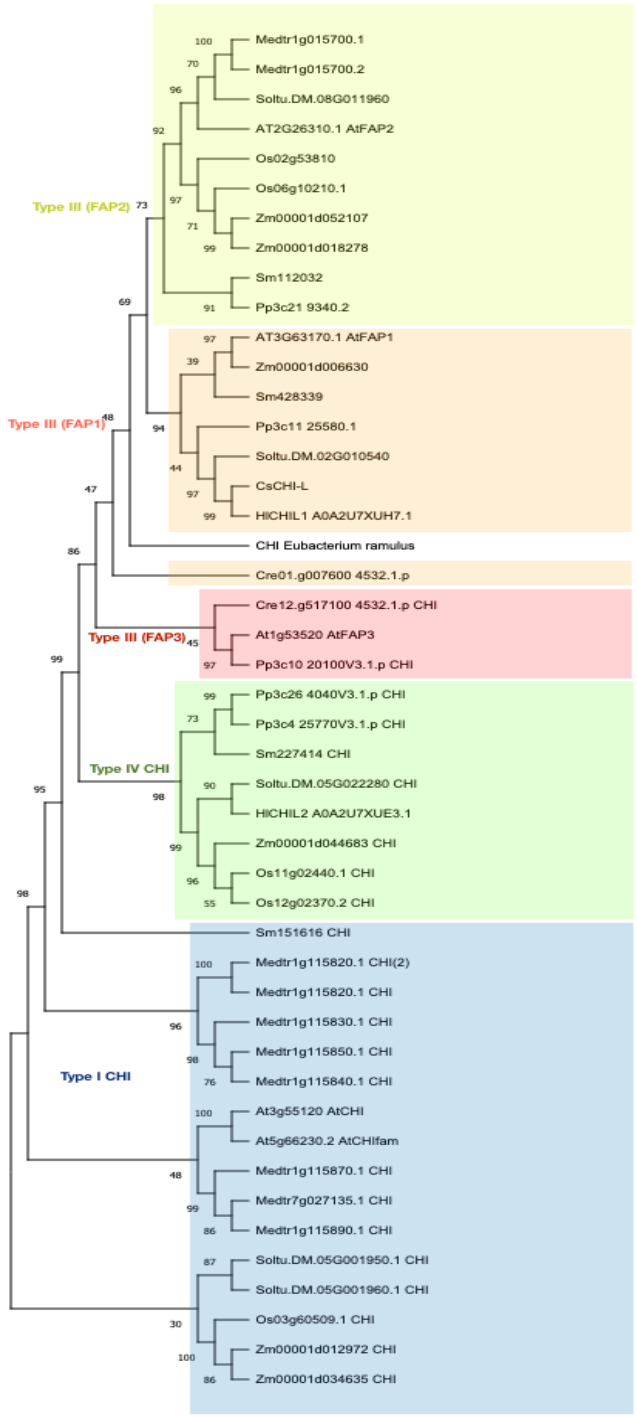


Figure 18. Phylogenetic analysis of CHI and CHI-L proteins.

A maximum likelihood phylogenetic consensus tree of 46 CHI and CHI-L proteins made using 1000 replicates.

Homology modeling of CsCHI-L

To further investigate the protein structure of CsCHI-L and check possible substitutions or mechanistic differences from other enzymes, I looked at the constructed homology model of the CsCHI-L (structure determined based on alpha1-fold protein model of the CsCHI-L). CHI proteins are typically functional monomers of approximately 220 residues, reflected in the 214 amino acid chain of our CsCHI-L candidate. From the 3D structure created by protein modelling, we can also see that the proposed CsCHI-L structure is mainly a large β-sheet, with multiple additional strands and several α-helices (Figure 19). Of the two unanimously conserved residues in bonafide CHI proteins, only one is present in the CsCHI-L – Tyr106, where Leu48 is in the place of Thr48 (Figure 19). Additionally, the highly conserved Arg36 is not present either; the CsCHI-L protein has Ile36. Comparative models of *A. thaliana* FAP and CHI proteins (AtFAP1 and AtCHI, repesticely) are included to demonstrate similarity in the folding of all three proteins and the distinctiveness of the high conserved CHI-fold.

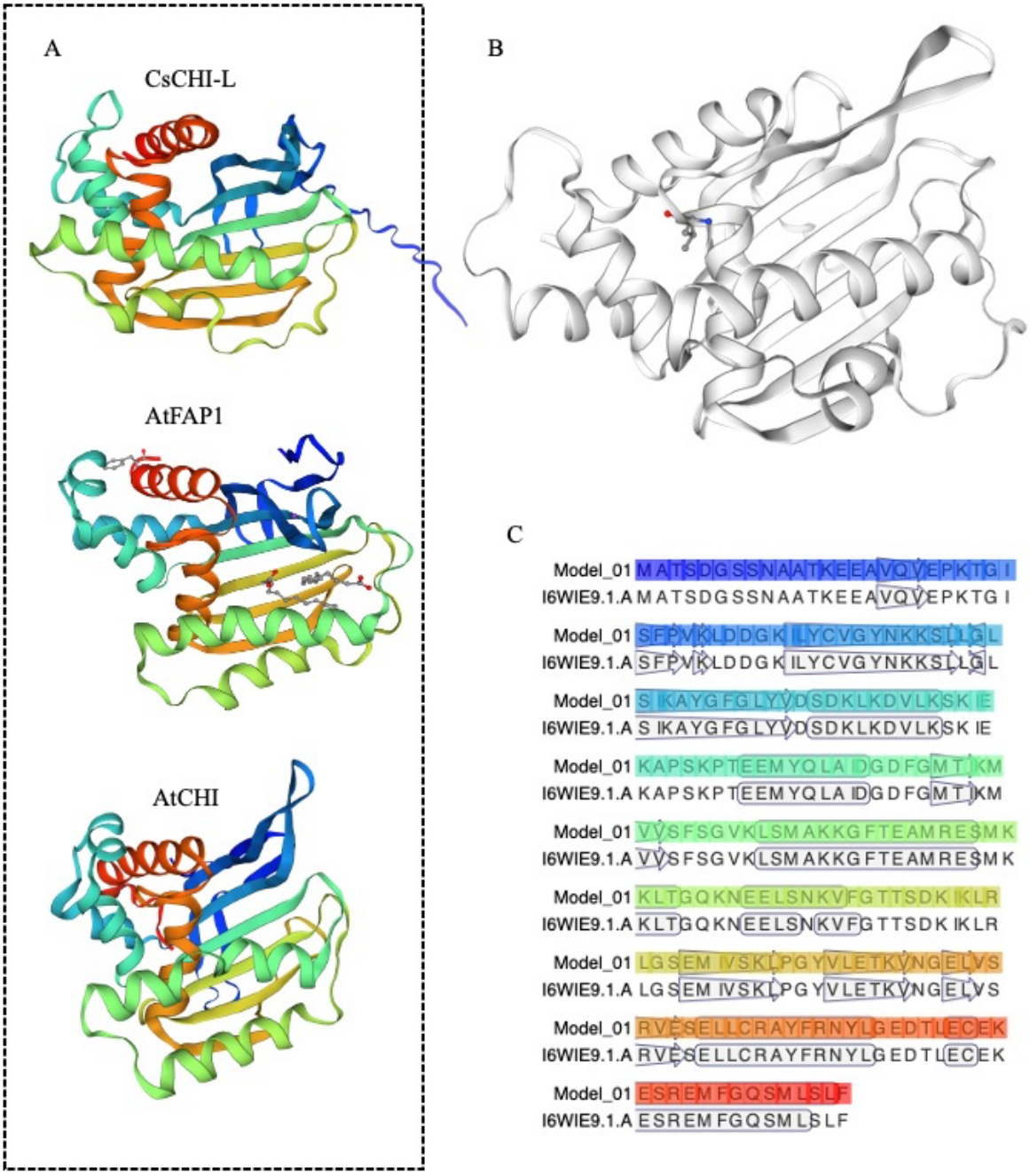


Figure 19. Protein modelling of CsCHI-L

A) Protein modelling of CsCHI-L alongside Arabidopsis proteins AtFAP1 and AtCHI demonstrating the conserved CHI-fold. B) Protein modelling of CsCHI-L highlighting the conserved Tyr109 residue. C) Sequence comparison between our model and that created by AlphaFold for the CsCHI-L protein, where arrows are indicative of regions responsible for the formation of distinct tertiary structure including the large and conserved β-sheet, β-strands and α-helices.

Discussion

Evolutionary backgrounds of flavonoids, CHS, CHI and FAP

Different classes of flavonoids arose sequentially, along with the corresponding proteins required for synthesis.[184] Prior phytochemical studies indicate that flavanones, flavonols and flavones appeared first alongside chalcones in liverwort ancestors before radiating into all niches of the plant kingdom.[185] However, more recent investigations have concluded that the production of flavonoids actually occurred earlier as flavones, isoflavones, and flavonols have been found in several different lineages of microalgae.[186] Anthocyanin production arose with the divergence of the flowering plants. Isoflavones emerged more recently, and are limited mainly to legumes. [184] Original functions of flavonoids have been postulated to be related to UV radiation protection as well as hormone regulation, however their functions have diversified and evolved with the plants.[187] Flavonoids have currently been identified as functioning in the phytoalexin pathogenic and herbivory defense, as well as acting as developmental regulators in pollen fertility and germination and contributing to co-pigmentation in flowers and seeds for the attraction of pollinators.[188–191] In humans, flavonoids show potent pharmacological activity, including anti-viral, anti-allergen, anti-inflammation and anti-carcinogenic effects. As such, they can impart significant protections against heart and cardiovascular disease, obesity and diabetes.[192] A major contributor to flavanone production, CHS, is a member of the Type III PKS superfamily, belonging to the thiolase superfamily and is found in plants, algae, fungi and bacteria.[193,194] CHS enzymes arose out of the de novo fatty acid biosynthetic pathway, and is derived from β-ketoacyl ACP synthase.[32,42] Plant Type III PKS enzymes have the same characteristic fold and the active catalytic residue triad present in the 3-ketoacyl-ACP synthase (KASIII) found in *E. coli*.[27] PKS enzymes are very prominent in plants and can be incredibly variable even within species, and this abundance is thought to facilitate the production of many novel specialized metabolites.

Non-CHS Type III PKS are still closely related, and are include enzymes such as bibenzyl synthases, pyrone synthases and acridone synthases.[195] These Type III plant PKS are also responsible for the synthesis of polyketides from fatty acyl-CoA starter molecules (as demonstrated by CsTKS action in *C. sativa*). The driving factor for the diversification of PKS enzymes is repeated gene duplication and mutation. For example, a single AA replacement changes the active site volume, chain length and mechanism of reaction for cyclization that occurs in the *Huperzia serrata PKS* (HsPKS1).[196] Enzymatic promiscuity also contributes to the diversity of compounds produced from these enzymes; where starter molecules can have a dramatic impact. Additionally, alternative incorporation of chain elongation and variable end-product cyclization all give the opportunity for diversified compound production.

The first two committed steps in the biosynthesis of flavonoids are facilitated by CHS and CHI. It is theorized that CHS evolved first, as it is responsible for the first committed step, and can proceed spontaneously without CHI (which selectively produces 2S-stereosiomers). Most enzymes arise through evolution by neofunctionalization of already existing enzyme classes through changes in substrate acceptance, catalysis mechanism, and product formation. In contrast, bona fide CHI activity arose through the gain in catalytic residues from ancestral non-catalytic CHI-L proteins. [197] The only two homologous protein families related to CHI proteins (that have currently been identified) are the FAP and CHI-L proteins mentioned previously. [42] FAP's, or fatty-acid binding proteins are a family of transport proteins for lipophilic substances, and help to modulate trafficking, metabolism and signalling. FAPs are found in protists, bacteria and algae alongside plants, while CHI and CHI-L proteins are only known to be present in plants. Both FAP and CHI/CHI-L proteins share the same characteristic CHI-fold, where CHI proteins are derived from FAP. FAP's can be classified into two branches, with FAP1 and FAP2

comprising one branch and FAP3 comprising the other (each branch evolving from common ancestors FAPa and FAPb, present in algae). The CHI-L gene family first arose in mosses, and a gene duplication event and subsequent functional diversification gave rise to the bona fide catalytic CHI proteins that are found in vascular plants.[42]

Types of CHI and CHI-L proteins

As previously noted, there are four classes of CHI and CHI-L enzymes that have been identified and classified. All proteins share a common folded structure, and this CHI-fold family also includes FAPs from fatty acid biosynthesis. Proteins are grouped together based on their functionality and sequence identity, where peptide sequence identity is >70% in the same class.[198]

Type I CHI are bona fide enzymes and ubiquitous in vascular plants. Type I CHI enzymes efficiently catalyze production of 2S-steroisomers of flavanones, specifically naringenin chalcone to 2S-narigenin (4,2',4',6'-tetrahydroxychalcone to 5-hydroxyflavanone). This reaction can occur spontaneously but increases 10^7 fold with the aid of Type I CHI enzymes.[199]

Type II CHI enzymes are also bona fide catalysts but are legume specific and primarily responsible for isoflavanone production. These enzymes have a broader substrate specificity – while they are still able to convert the original naringenin chalcone to 2S-naringenin, they can also facilitate the conversion of 6'-deoxychalcone to 5'-deoxyflavanone (2S-liquiritigenin) which leads to a subclass of legume-specific metabolites (5-deoxyflavonoids, including 5-deoxyisoflavanoids).[35] Sequence identity between type I and type II CHI enzymes is about 50%.[198] The origin of Type II CHI enzymes is currently under speculation, but was originally postulated to have arisen as part of a gene duplication event that began to produce 5-

deoxy(iso)flavonoids alongside the establishment of the Fabaceae.[198] A Type I and Type II CHI genes are present as a gene cluster on the Soybean chromosome 20, derived from tandem gene duplication.[176] However, recent investigation has identified CHI in ancient land plants such as *Marchantia* and *Selaginella* that are more similar to Type II CHI (due to a much broader substrate specificity), which is indicative that their emergence predates legumes and occurred before the divergence of liverworts.[34] Upon further investigation, it has been determined that the CHI from these basal land plants form separate groups from the Type I and II CHI's. [34,200]

Type III CHI enzymes have no chalcone cyclization but exhibit fatty acid binding properties and are localized to plastids. These enzymes are distributed in both land plants and green algae.[42,185] Type III CHIs are considered fatty acid binding proteins and participate in fatty acid metabolism and can be divided into FAP1, FAP2 and FAP3.[42] Disruption of type III CHI genes has demonstrated a significant effect on fatty acid synthesis and storage in developing Arabidopsis embryos, indicative that these enzyme play a role in fatty acid function.[42] Phylogenetic and structural analysis is indicative that these are the prototype of bona fide CHI enzymes.[42]

Type IV CHIs functionality is largely undetermined. It is propositioned that they could act an enhancers or activators in flavonoid and pigmentation production, as has been demonstrated in multiple plants.[46,201,202] More recent investigation seems to indicate that Type IV CHIs would function as pathway rectifiers, where they physically interact with CHS enzymes to promote substrate influx to the flavonoid pathway instead of the polyketide pathway.[40] This type of CHI-fold protein may have evolved earlier than initially believed (in primal Eukaryotes) as sequence analysis reveals CHI-fold proteins in fungi, mycetozoa and gammaproteobacterial, indicative that the fold occurred before divergence of the domains of life.

As previously mentioned, current phylogenetic analyses suggest that CHI-L proteins arose first, evolving from the closely related FAP's in mosses. These CHI-L proteins were then able to serve as the evolutionary ancestors of bona fide CHI proteins: Type I and Type II.[42,202] The ability of CHI-L proteins to bind CHS could have been passed on to the bona fide Type I and II CHIs, and they also possess the ability to bind CHS and could be an important evolutionary process acquired from the Type IV CHI-Ls.[203]

Phylogenetic relationships of CHI and CHI-L proteins from constructed tree

Gene duplications are the major drivers of gene family evolution. It has previously been demonstrated that gene duplication in *C. Sativa* gave rise to multiple CHI-L proteins.[204] *C. sativa* contains six copies of CHI genes (three Type III, one Type I and one Type IV), all originating from different orthologous groups. Our candidate protein, CsCHI-L, belongs to the Type III group of CHI-L proteins, specifically residing within the FAP1 clade. In this phylogenetic tree, only our candidate CsCHI-L was included from the *C. sativa* CHI/CHI-L proteins for comparison against CHI/CHI-L proteins from different phylogenetic groups. As our hypothesized candidate has no bona-fide catalytic activity, it was expected that Type I and Type II CHI enzymes will be more distantly related to our candidate than will others of the Type III class and the Type IV class.

Type I CHI were prominent and present in all vascular plants, implying conserved orthologous function. Type I CHI's evolved vertically within vascular species. The outgrouping of the Type I CHI from *S. moellendorffii* indicates that it may represent a primitive CHI. It is currently accepted that Type II CHI's are only found in members of the Fabaceae family, as they are unique to legumes; however some recent discoveries of distant Type II CHI in liverworts and mosses has challenged this belief. 7,4'-dihydroxyflavanone, which is a derivative compound of

6'-deoxychalcone, had been identified in *S. moellendorffii* (as well as in *Selaginella doederleinii*). [205,206] As these compounds were believed to be unique to legumes, their presence in such basal plants raises questions about the evolution and distribution of bona fide CHI's. The CHI proteins identified in early land plants (liverworts and *S. moellendorffii*) phylogenetically represent an ancestral outgroup of vascular plant CHI's that may have evolved before that divergence of lineages.[34] Type II CHI proteins are therefore not represented in our phylogenetic analysis, as they comprise their own specific and unique clade.

Conservation of type IV CHI's in land plants is indicative that the divergence of type IV CHI occurred before the plant diversification. Recent comprehensive phylogenetic analyses specifically focused on CHI and CHI-L proteins has revealed that FAP3 CHI's are the original CHI-L proteins from which Type IV, II and I CHI's arose. Type IV CHI can therefore be thought of as representative of an evolutionary intermediate between FAP3 (Type III CHI-L) and bona fide CHI's. Active site residues and hydrogen bond network are vital for bona fide CHI-enzymatic activity. CHI functional divergence has been affected by the amino acid substitutions in the key core residues as a result of natural selection pressure. The close grouping of bona-fide CHI enzymes (Type I) in individual species is indicative of the importance of conserved internal catalytic residues, as they are required for enzyme functionality.

As expected, our CHI-L candidate is most closely related to one of the CHI-L proteins found in hops, HlCHIL1 (Figure 18). However, the relation to the other hops CHI-L, HlCHIL2 protein, is quite distanced and indicative that the two proteins may not share a similar function as initially thought, where HlCHIL2 rectifies DMX and XN production through direct protein-protein interaction with the PKS enzyme (Figure 18). The presence of HlCHIL1 and CsCHI-L in

the same class of CHI-L proteins (Type III, FAP3) is indicative that both of these proteins contribute to fatty acid binding roles in the cell.

Homology model of CsCHI-L protein

The most supported reaction mechanism for CHI proteins involves acid-base catalysis with an enol intermediate.[207] The structure of the CHI-naringenin enzyme-substrate complex demonstrates the importance of hydrogen bonds in the binding cleft, of which Thr48, Ala49, Lys97, Tyr106 and Tyr152 largely contribute.[36,37] CHI enzymatic activity is attributed to five catalytic residues, where an arginine aligns and stabilizes the substrate and transition state and an additional four key residues participate in the hydrogen bonding mentioned above (two threonine, a tyrosine and an asparginine) to direct substrate binding and facilitate ring closure.[36,37,42] Only two residues – Thr48 and Tyr106, which contribute to hydrogen bonding at the bottom of the active site cleft, are conserved across all bona fide CHI proteins.[36,37] CHI proteins have a distinct fold, where a sizeable β-sheet is the main component, and the core is comprised of an α-helices layer with shorter β-strands opposite the sheet.[37] In comparison, AtFAP1, AtFAP2 and AtFAP3 all share the same CHI-fold with bona fide CHI proteins (including AtCHI), but have substitutions at critical catalytic positions and lack enzymatic activity.[42]

The proposed 3D structure of the CsCHI-L protein can lend insight as to how the protein may function and allows for comparison to other similar proteins in the CHI family. The homology model clearly shows the characteristic CHI-fold is present in our CsCHI-L protein, starting with short β-strands, then dominated by a large β-sheet and complimented with multiple sections of α-helices to create the "active site". The structural similarities between CsCHI-L, AtFAP3 and AtCHI are well demonstrated by comparison of their physical structure. As

mentioned previously, FAPs are homologous to CHI-fold proteins but lack catalytic activity due to substitution of several key residues in the active site. The lack of conserved residues at key positions in the CsCHI-L protein explain the lack of bona fide CHI activity. Tyr106 and Leu48 are conserved in all bona fide CHI enzymes, and only the Tyr106 is present in the CsCHI-L protein. As Arg36 is responsible for interaction with the substrate and helps to facilitate positioning into the active site, this is an important residue for the enzymatic activity of the proteins.

From the phylogenetic analysis performed here, we see that our CsCHI-L protein groups the closest to AtFAP3, indicative that it would be more related to the FAP3 branch of FAP and thus not possess catalytic activity (Figure 18). This is supported by the protein model created through comparison of the CsCHI-L protein to the AtFAP1 protein (Figure 19).

CsCHI-L may function in other sections of the cannabinoid pathway

CHI-L proteins have demonstrated their ability to bind fatty acids and intermediates, as well as rectify PKS activity to decrease by-product formation. As the cannabinoid biosynthetic pathway was noted as being halted at OA-forming steps, our major testing with the CsCHI-L candidate surrounded the tetraketide intermediate formation with the PKS enzyme (CsTKS) and possibly activity of the CsCHI-L in those steps of the pathway. In the yeast SBS the addition of CHI-L to the OA-forming steps in cannabinoid biosynthesis did not improve the yield of OA. However, the expression of this gene is quite high in cannabinoid-producing glandular trichomes, and as such it is likely that CHI-L serves some function in these specialized cells. In the cannabis sister genus Humulus (hops), CHI-L proteins function in different parts of the prenylflavanoid biosynthetic pathway. This discovery is particularly relevant to the biosynthetic production of cannabinoids, due to the close relation of hops and

cannabis and the similar mechanisms of production and sequestration of their respective natural products. One CHI-L protein (HlCHIL2) was noted to enhance the activities of CHS and prenyltransferase HlPT1L in protein-protein interactions.[39] Hops prenyltrasnferase PT1 exhibits structural similarities of the plant aromatic PT family, including a D-rich motif, as well as membrane-spanning domains and a transit peptide sequence.[208] It is a membrane-bound protein, as are all known plant prenyltransferases.[209] I attempted to incorporate the CsPT1 and CsPT4 prenyltransferase enzymes into our SBS's to test whether CsCHI-L operates at this step in the cannabinoid pathway; however cloning of *CsPT1* and *CsPT4* was not successful. Further experimentation with the CsCHI-L candidate should include a prenyltransferase representative, to monitor any protein-protein interactions and potential improvements in pathway yield that may occur. To circumvent the potential issue of expressing thylakoid membrane-bound prenytransferases in microbial hosts, a soluble prenyltransferase could be used instead. For example, NphB prenyltransferase from bacterial Streptomyces was shown the geranylate OA to CBGA in yeast.[210]

HlCHIL1 has been shown to increase the storage capacity of prenylchalcones in Hops by binding and stabilizing the open-ring conformation of narigenin chalcone and DMX.[39] It is therefore probable for the homologs of HlCHIL1 to also be responsible for binding with and stabilizing the products of Type IIII PKS where they have an aromatic ring derived from the addition of three malonyl-CoA's. As hypothesized by Ban et al., THCA and CBDA are both unstable compounds, and the CsCHI-L protein could bind to these compounds to stabilize them in the trichomes of *C. sativa*.[39] The binding action noted here could also be responsible for the high phytocannabinoid acid accumulation in trichomes, where in other organisms the presence of said compounds at such a high concentration results in cellular toxicity and death. The

stabilization imparted by binding with the CHI-L proteins would also alleviate the negative feedback inhibition seen by most biosynthetic pathways, where sequestration of free polyketides by the CHI-L would mean the pathway would still be acting to produce more compounds. The CsCHI-L protein does have some key residue substitutions from the HlCHIL1 protein and from other FAP1 CHI-L's that are important for their binding capabilities. Notably, the Arg-Tyr pair present in FAP1 proteins and responsible for the tethering of the carboxylate group is substituted in the CsCHI-L with Asn, making a Asn-Tyr pairing and in the HlCHIL1 protein, the Tyr is substituted to a Arg-Phe pair. Additionally, Val139 and Ala143 are key residues in the HlCHIL1 protein for binding of DMX and only Val139 is retained in the CsCHI-L, while Ala143 becomes Tyr143. These residue discrepancies are responsible for the differential binding ability of the protein and the preferred substrate.

Other proteins may function to stabilize OA biosynthesis in cannabis

An additional candidate protein, BetV1-L, was also identified by sequence prevalence in the Cannabis EST library. BetV1-L is in the same protein-fold family as the *Streptomyces* tetracenomycin aromatase/cyclase, which has the ability to bind linear polyketides. In this vein, a BetV1-L protein could function in the OA-forming steps of the phytocannabinoid biosynthetic pathway to facilitate transfer of the tetraketide intermediate to CsOAC. BetV1-L proteins comprise a superfamily with the ability and preference to bind large, often bulky, hydrophobic compounds.[211] This function is directly related to the large, hydrophobic cavity that opens to the exterior of the protein.[212] The second-largest family within the BetV1-L superfamily is the polyketide cyclase family, gaining its name from the polyketide cyclase and hydrase activity of BetV1-L proteins found in Streptomyces.[213] While unrelated to many other polyketide cyclases, this subset of Betv1-L proteins forms its own phylogenetic group with sequences from 449

species.[214] As is suggested by their name, polyketide cyclases are responsible to the cyclisation of polyketide molecules as part of the biosynthesis of a diverse number of compounds in prokaryotic biosynthetic pathways. The Cannabis BetV1-L protein was originally considered as a candidate during investigation of the (now elucidated) CsOAC enzyme, alongside our CsCHI-L protein.[89] Due to time constraints in the project, no testing was ever conducted with the CsBetV1-L protein.

Alternative synthetic biology system construction

In this experimentation, GOI's were introduced into the microbial host via vector cassettes, where plasmid incorporation facilitated the expression of pathway enzymes. The first plasmid editing occurred in 1973, where a gene coding for kanamycin resistance replaced a gene with tetracycline resistance and was introduced to *E. coli* by transformation.[215] Following this discovery, plasmids became popular tools for the delivery and manipulation of heterologous DNA inside a microbial host. Many molecular cloning techniques are centered around the use of plasmid vectors. General characteristics of plasmid vectors include a origin of replication (ori), and in some cases vectors suitable for expression in multiple hosts will contain multiple ori sites, where each corresponds to recognition in a respective host.[216] Additionally, plasmids must contain a selection marker, where the inclusion of the plasmid will also incorporate the inclusion of a gene that confers a selective advantage to the microbial cell and will allow for the isolation of genetically edited strains from mixed or wild cultures.[217,218] The most common types of selective markers are auxotrophic selection markers, where the plasmid corresponds to a metabolic enzyme lacking in the microbial host, or a drug resistance selection marker, where a certain gene will confer resistance to antibiotic presence. Lastly, vectors often contain a multiple cloning site (MCS), where the gene of interest can be easily cloned using restriction enzymes.[219]

For protein expression, MCS's also need to have an appropriate promoter and terminator sequences flanking them.[220] This type of SBS construction relies on the ability of the microbial cell to maintain and express genetic information from the incorporated plasmid(s). Even though advances have been achieved in plasmid design and technologies, there are still challenges to using these genetic tools. Mentioned previously in this document are copy number issues and plasmid incompatibility. Additional challenges are related to plasmid structure and stability within the host. Incorporated plasmids have been known to be spontaneously lost during cell division, related to defects in replication and inability of the plasmid to partition effectively to daughter cells.[221–223] DNA segment loss and rearrangement, resulting in non-functional GOI's, is also possible in plasmid vectors. This can be due to the toxicity of certain genes or gene products, where mutant versions of the genes now confer an advantage when compared to those highly expressing said gene.[224] Certain genes are unable to be cloned into *E. coli* due to DNA toxicity and instability, such as the Vssc1 sodium and Cch1 calcium channels.[225,226] As such, more modern approaches using genomic integration technologies are becoming more common in building SBSs as a way to circumvent the issues associated with plasmid instability.

Genomic integration is the stable incorporation of foreign DNA into an organism's chromosome, where it then becomes part of the genome. This technique can be used to introduce homologous genes for the construction of synthetic biology pathways. Genomic editing or engineering relies on the ability of the cells to repair double-stranded breaks in its DNA. The two types of repairs for DNA are homology-directed repair (HDR), which requires overlapping homologous regions of a piece of donor DNA that can be incorporated into a break site, or non-homologous end joining (NHEJ) where break ends are directly ligated and do not require a homologous template. Prokaryotic cells - such as *E. coli* bacteria - lack the NHEJ recombination

activity and double-stranded breaks lead to cell death.[227] Prokaryotic genomic editing is therefore more difficult than in eukaryotic cells and requires a more complex approach. Genomic editing in bacterial cells typically has relied on "recombineering", where bacteriophage lamba-encoded functions, referred to as Lamba Red, can integrate short segments of linear ssDNA into the chromosome.[228] Modifications have been achieved using ET-cloning, which also utilizes the homology-directed integration of double-stranded DNA.[229] Advancements have been made with respect to bacterial genomic integration, including the Multiplex Automated Genome Engineering (MAGE) system, which targets multiple chromosomal locations for modification to provide an increased level of combinatorial diversity.[230] MAGE still requires the synthesis and presence of ssDNA as well as the LambaRed recombinase expression, and editing efficiencies are usually below 20%.[230] MAGE has since been improved with the discovery of single-stranded DNA-annealing proteins and the pORTMAGE system was created.[231] Bacterial editing systems continue to be improved upon, with recent efforts such as "serial enrichment for efficient recombineering" (SEER) identifying highly active variants of single-stranded DNA annealing proteins to improve single-locus editing up to 50%.[232] A new approach for bacterial editing has recently been introduced as retron library recombineering (RLR), where targeted reverse-transcription produces ssDNA from a retron plasmid in vivo and is incorporated directly into the genome with a greater than 90% efficiency.[233] This type of editing is very promising, as it surpasses the scale and specificity of CRISPR/Cas9, and is well suited to large numbers of natural variation.[233] However, this type of approach is very much in its infancy and has yet to be widely applied as an editing technique.

In eukaryotic cells, there are many different approaches to gene editing, most of which have been developed in the past decade(s). To briefly touch on a few, there is the Cre/LoxP system,

where the prokaryotic coliphage P1 integrates genes into the host DNA through site-specific recombinase activity.[234] The mitochondrial endonuclease I-Scel has been used to target double-stranded breaks to a specific location in the genome.[235] Zinc-finger nucleases (ZFNs), are artificial endonucleases that bind to FolkI cleavage sites and allow for breaks in the DNA that can then be repaired by either NHEJ or HDR, if there is the presence of homologous donor sequences.[236] Transcription activator-like effector nuclease (TALENs) gene editing takes advantage of TALE proteins ability to bind to DNA through a binding domain consisting of monomers, which code for a specific target in the DNA.[237–240] These binding domains can be edited to target a specific site, where chimeric TALEN proteins will induce a double-stranded break and allow for genomic integration.[241,242]

Recently, the most widely used technology for genomic editing has been the CRISPR and CRISPR-associated system (Cas).[243] Using simple guide-RNA sequences, Cas9 endonuclease activity can target specific genomic loci and induce double-stranded breaks for the reincorporation of DNA or GOIs.[244] NHEJ is the dominant repair mechanism in most eukaryotes, and CRISPR/Cas9 editing can be used to induce multiple break sites in the genome without significant damage to the cell because of this repair machinery. The CRISPR/Cas9 complex can be used to create these double-stranded break sites at certain regions, and single-guide RNAs are very successful at controlling the specificity of HDR editing techniques.[245,246] Genomic editing in *S. cerevisiae* has been very successfully established using the CRISPR/Cas complex, where homologous recombination facilitates the incorporation of oligonucleotide sequences with a near 100% recombination frequency.[247]

In the experimentation I conducted, problems were encountered with respect to yeast plasmid stability. This was demonstrated when the *CsOAC* and *CsTKS* harbouring vector was

shown to no longer be present and expressing said proteins, even with rigorous use of selective media. Homologous recombination is very efficient in *S. cerevisiae* and presents a more reliable and straightforward mechanism for GOI incorporation. Genomic integration also allows for more control with respect to copy number, ensuring that all genes will be properly expressed, and the amount of proteins produced can be controlled by selecting promoter and terminator sequences with varying strengths. In future experimentation regarding the cannabinoid biosynthetic pathway, it would be ideal to use genomic integration to incorporate GOI to provide a more reliable system with stable protein expression to properly monitor the efficacy and efficiency of said proteins.

Conclusion

In this line of research, no observable action was demonstrated in the OA-forming steps of cannabinoid biosynthetic pathway by the candidate accessory protein CsCHI-L. Successful construction of the pathway in a SBS allowed for direct comparison of constructs incorporating CsCHI-L and those without, and no significant difference was noted for constructs with and without the hypothesized protein. Additionally, higher concentration of the derailment product PDAL was noted in the CsCHI-L containing than in the system(s) without. This is indicative that CsCHI-L has no beneficial action with respect to lowering by-product formation in the OA-forming steps of the cannabinoid pathway. Future experimentation is required to determine a point of action for said protein, as there is potential for CsCHI-L to work in different steps of the cannabinoid biosynthetic pathway than those explored in this experimentation. Additionally, other candidate proteins, included the previously identified CsBetV1-L protein, may work in the manner proposed for the CsCHI-L candidate, and should be investigated for their potential action in the production of cannabinoids.

References

1. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. *Proc Natl Acad Sci U S A*. 2018;115(25). doi:10.1073/pnas.1711842115
2. Morris JL, Puttick MN, Clark JW, et al. The timescale of early land plant evolution. *Proc Natl Acad Sci U S A*. 2018;115(10). doi:10.1073/pnas.1719588115
3. Ralph J, Lundquist K, Brunow G, et al. Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids. *Phytochem Rev*. 2004;3(1-2). doi:10.1023/B:PHYT.0000047809.65444.a4
4. Liu Q, Luo L, Zheng L. Lignins: Biosynthesis and biological functions in plants. *Int J Mol Sci*. 2018;19(2). doi:10.3390/ijms19020335
5. Ma YF, Xiao C. Push-pull effects of three plant secondary metabolites on oviposition of the potato tuber moth, Phthorimaea operculella. *J Insect Sci*. 2013. doi:10.1673/031.013.12801
6. Ojeda F, Midgley J, Pauw A, et al. Flower colour divergence is associated with post-fire regeneration dimorphism in the fynbos heath Erica coccinea subsp. coccinea (Ericaceae). *Evol Ecol*. 2019. doi:10.1007/s10682-019-09985-0
7. Pratt C, Pope TW, Powell G, Rossiter JT. Accumulation of glucosinolates by the cabbage aphid Brevicoryne brassicae as a defense against two coccinellid species. *J Chem Ecol*. 2008. doi:10.1007/s10886-007-9421-z
8. Venturelli S, Petersen S, Langenecker T, Weigel D, Lauer UM, Becker C. Allelochemicals of the phenoxazinone class act at physiologically relevant concentrations. *Plant Signal Behav*. 2016. doi:10.1080/15592324.2016.1176818
9. Guerriero G, Berni R, Muñoz-Sanchez JA, et al. Production of plant secondary metabolites: Examples, tips and suggestions for biotechnologists. *Genes (Basel)*.

2018;9(6). doi:10.3390/genes9060309

10. Rieseberg TP, Dadras A, Fürst-Jansen JMR, et al. Crossroads in the evolution of plant specialized metabolism. *Semin Cell Dev Biol*. 2023;134. doi:10.1016/j.semcdb.2022.03.004

11. Ashihara H, Sano H, Crozier A. Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry*. 2008;69(4). doi:10.1016/j.phytochem.2007.10.029

12. Pan L, Sinden MR, Kennedy AH, et al. Bioactive constituents of Helianthus tuberosus (Jerusalem artichoke). *Phytochem Lett*. 2009. doi:10.1016/j.phytol.2008.10.003

13. Qasim M, Abideen Z, Adnan MY, et al. Antioxidant properties, phenolic composition, bioactive compounds and nutritive value of medicinal halophytes commonly used as herbal teas. *South African J Bot*. 2017. doi:10.1016/j.sajb.2016.10.005

14. Hussain F, Rana Z, Shafique H, Malik A, Hussain Z. Phytopharmacological potential of different species of Morus alba and their bioactive phytochemicals: A review. *Asian Pac J Trop Biomed*. 2017. doi:10.1016/j.apjtb.2017.09.015

15. Sut S, Dall'Acqua S, Poloniato G, Maggi F, Malagoli M. Preliminary evaluation of quince (Cydonia oblonga Mill.) fruit as extraction source of antioxidant phytoconstituents for nutraceutical and functional food applications. *J Sci Food Agric*. 2019. doi:10.1002/jsfa.9271

16. Kallscheuer N, Classen T, Drepper T, Marienhagen J. Production of plant metabolites with applications in the food industry using engineered microorganisms. *Curr Opin Biotechnol*. 2019. doi:10.1016/j.copbio.2018.07.008

17. Rodríguez-Meizoso I, Jaime L, Santoyo S, Señoráns FJ, Cifuentes A, Ibáñez E. Subcritical

water extraction and characterization of bioactive compounds from Haematococcus pluvialis microalga. *J Pharm Biomed Anal*. 2010. doi:10.1016/j.jpba.2009.03.014

18. Garlapati D, Chandrasekaran M, Devanesan AA, Mathimani T, Pugazhendhi A. Role of cyanobacteria in agricultural and industrial sectors: an outlook on economically important byproducts. *Appl Microbiol Biotechnol*. 2019. doi:10.1007/s00253-019-09811-1
19. Roepke J, Salim V, Wu M, et al. Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proc Natl Acad Sci U S A*. 2010. doi:10.1073/pnas.0911451107
20. Uprety Y, Poudel RC, Asselin H, Boon EK, Shrestha KK. Stakeholder perspectives on use, trade, and conservation of medicinal plants in the Rasuwa district of central Nepal. *J Mt Sci*. 2011. doi:10.1007/s11629-011-1035-6
21. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol*. 1999;6(8). doi:10.1038/11553
22. Chen Y, Kelly EE, Masluk RP, Nelson CL, Cantu DC, Reilly PJ. Structural classification and properties of ketoacyl synthases. *Protein Sci*. 2011;20(10). doi:10.1002/pro.712
23. Heath RJ, Rock CO. Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in Escherichia coli. *J Biol Chem*. 1996;271(4). doi:10.1074/jbc.271.4.1833
24. Rangan VS, Joshi AK, Smith S. Mapping the functional topology of the animal fatty acid synthase by mutant complementation in vitro. *Biochemistry*. 2001;40(36). doi:10.1021/bi015535z
25. Rock CO, Cronan JE. Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim Biophys Acta - Lipids Lipid Metab*. 1996;1302(1).

doi:10.1016/0005-2760(96)00056-2

26. Hopwood DA, Sherman DH. MOLECULAR GENETICS OF POLYKETIDES AND ITS COMPARISON TO FATTY ACID BIOSYNTHESIS. *Annu Rev Genet*. 1990;24(1). doi:10.1146/annurev.ge.24.120190.000345

27. Austin MB, Noel JP. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep*. 2003;20(1):79-110. doi:10.1039/b100917f

28. Nivina A, Yuet KP, Hsu J, Khosla C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem Rev*. 2019;119(24). doi:10.1021/acs.chemrev.9b00525

29. Jez JM, Ferrer JL, Bowman ME, et al. Struc3ture and mechanism of chalcone synthase-like polyketide synthases. *J Ind Microbiol Biotechnol*. 2001;27(6). doi:10.1038/sj.jim.7000188

30. Kreuzaler F, Hahlbrock K. Enzymatic synthesis of aromatic compounds in higher plants: Formation of naringenin (5,7,4′-trihydroxyflavanone) from p-coumaroyl coenzyme A and malonyl coenzyme A. *FEBS Lett*. 1972;28(1). doi:10.1016/0014-5793(72)80679-3

31. Dixon RA, Paiva NL. Stress-induced phenylpropanoid metabolism. *Plant Cell*. 1995;7(7). doi:10.1105/tpc.7.7.1085

32. Weng JK, Noel JP. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb Symp Quant Biol*. 2012;77. doi:10.1101/sqb.2012.77.014787

33. Xu H, Zhang F, Liu B, et al. Characterization of the formation of branched short-chain fatty acid:coas for bitter acid biosynthesis in hop glandular trichomes. *Mol Plant*. 2013;6(4). doi:10.1093/mp/sst004

34. Cheng AX, Zhang X, Han XJ, et al. Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of

flavonoids. *New Phytol*. 2018;217(2). doi:10.1111/nph.14852

35. Ralston L, Subramanian S, Matsuno M, Yu O. Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. In: *Plant Physiology*. Vol 137. ; 2005. doi:10.1104/pp.104.054502

36. Jez JM, Bowman ME, Noel JP. Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. *Biochemistry*. 2002;41(16). doi:10.1021/bi0255266

37. Jez JM, Bowman ME, Dixon RA, Noel JP. Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat Struct Biol*. 2000;7(9):786-791. doi:10.1038/79025

38. Dixon RA, Richard Blyden E, Robbins MP, Van Tunen AJ, Mol JN. Comparative biochemistry of chalcone isomerases. *Phytochemistry*. 1988;27(9). doi:10.1016/0031-9422(88)80666-6

39. Ban Z, Qin H, Mitchell AJ, et al. Noncatalytic chalcone isomerase-fold proteins in Humulus lupulus are auxiliary components in prenylated flavonoid biosynthesis. *Proc Natl Acad Sci U S A*. 2018. doi:10.1073/pnas.1802223115

40. Waki T, Mameda R, Nakano T, et al. A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat Commun*. 2020;11(1):1-14. doi:10.1038/s41467-020-14558-9

41. Xu H, Lan Y, Xing J, Li Y, Liu L, Wang Y. AfCHIL, a Type IV Chalcone Isomerase, Enhances the Biosynthesis of Naringenin in Metabolic Engineering. *Front Plant Sci*. 2022;13. doi:10.3389/fpls.2022.891066

42. Ngaki MN, Louie G V., Philippe RN, et al. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature*. 2012. doi:10.1038/nature11009

43. Morita Y, Takagi K, Fukuchi-Mizutani M, et al. A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant J*. 2014;78(2). doi:10.1111/tpj.12469

44. Zhou L, Wang Y, Ren L, et al. Overexpression of Ps-CHI1, a homologue of the chalcone isomerase gene from tree peony (Paeonia suffruticosa), reduces the intensity of flower pigmentation in transgenic tobacco. *Plant Cell Tissue Organ Cult*. 2014;116(3). doi:10.1007/s11240-013-0403-2

45. Nagel J, Culley LK, Lu Y, et al. EST analysis of hop glandular trichomes identifies an O-methyltransferase that catalyzes the biosynthesis of xanthohumol. *Plant Cell*. 2008;20(1). doi:10.1105/tpc.107.055178

46. Li H, Ban Z, Qin H, Ma L, King AJ, Wang G. A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. *Plant Physiol*. 2015;167(3). doi:10.1104/pp.114.253682

47. McPartland JM, Hegman W, Long T. Cannabis in Asia: its center of origin and early cultivation, based on a synthesis of subfossil pollen and archaeobotanical studies. *Veg Hist Archaeobot*. 2019. doi:10.1007/s00334-019-00731-8

48. Touw M. The religious and medicinal uses of Cannabis in China, India and Tibet. *J Psychoactive Drugs*. 1981. doi:10.1080/02791072.1981.10471447

49. Li HL. An archaeological and historical account of cannabis in China. *Econ Bot*. 1973. doi:10.1007/BF02862859

50. Kuddus M, Ginawi IAM, Al-Hazimi A. Cannabis sativa: An ancient wild edible plant of India. *Emirates J Food Agric*. 2013. doi:10.9755/ejfa.v25i10.16400

51. Zuardi AW. History of cannabis as a medicine: A review. *Rev Bras Psiquiatr*. 2006.

doi:10.1590/S1516-44462006000200015

52. Happyana N, Agnolet S, Muntendam R, Van Dam A, Schneider B, Kayser O. Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR. *Phytochemistry*. 2013. doi:10.1016/j.phytochem.2012.11.001

53. Taura F, Morimoto S, Shoyama Y, Mechoulam R. First Direct Evidence for the Mechanism of Δ1-Tetrahydrocannabinolie Acid Biosynthesis. *J Am Chem Soc*. 1995. doi:10.1021/ja00143a024

54. Atakan Z. Cannabis, a complex plant: Different compounds and different effects on individuals. *Ther Adv Psychopharmacol*. 2012. doi:10.1177/2045125312457586

55. Hanuš LO, Meyer SM, Muñoz E, Taglialatela-Scafati O, Appendino G. Phytocannabinoids: A unified critical inventory. *Nat Prod Rep*. 2016. doi:10.1039/c6np00074f

56. Berman P, Futoran K, Lewitus GM, et al. A new ESI-LC/MS approach for comprehensive metabolic profiling of phytocannabinoids in Cannabis. *Sci Rep*. 2018;8(1). doi:10.1038/s41598-018-32651-4

57. Gachet MS, Schubert A, Calarco S, Boccard J, Gertsch J. Targeted metabolomics shows plasticity in the evolution of signaling lipids and uncovers old and new endocannabinoids in the plant kingdom. *Sci Rep*. 2017. doi:10.1038/srep41177

58. Bohlmann F, Hoffmann E. Cannabigerol-ähnliche verbindungen aus Helichrysum umbraculigerum. *Phytochemistry*. 1979. doi:10.1016/0031-9422(79)83025-3

59. Toyota M, Shimamura T, Ishii H, Renner M, Braggins J, Asakawa Y. New bibenzyl cannabinoid from the New Zealand liverwort Radula marginata. *Chem Pharm Bull*. 2002. doi:10.1248/cpb.50.1390

60. Mahlberg PG, Hammond CT, Turner JC, Hemphill JK. Structure, Development and Composition of Glandular Trichomes of Cannabis Sativa L. In: *Biology and Chemistry of Plant Trichomes*. ; 1984. doi:10.1007/978-1-4899-5355-1_2

61. Dayanandan P, Kaufman PB. TRICHOMES OF CANNABIS SATIVA L. (CANNABACEAE). *Am J Bot*. 1976;63(5):578-591. doi:10.1002/j.1537-2197.1976.tb11846.x

62. Happyana N, Kayser O. Monitoring Metabolite Profiles of Cannabis sativa LTrichomes during Flowering Period Using 1H NMR-Based Metabolomics and Real-Time PCR. *Planta Med*. 2016. doi:10.1055/s-0042-108058

63. Gaoni Y, Mechoulam R. Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish. *J Am Chem Soc*. 1964. doi:10.1021/ja01062a046

64. Mechoulam R, Gaoni Y. Hashish-IV. The isolation and structure of cannabinolic cannabidiolic and cannabigerolic acids. *Tetrahedron*. 1965. doi:10.1016/0040-4020(65)80064-3

65. Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature*. 1990. doi:10.1038/346561a0

66. Munro S, Thomas KL, Abu-Shaar M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature*. 1993. doi:10.1038/365061a0

67. Li X, Hua T, Vemuri K, et al. Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell*. 2019;176(3). doi:10.1016/j.cell.2018.12.011

68. Methodology ZJL, Validation QZ, Analysis GWH, Investigation ZJL. Crystal Structure of the Human Cannabinoid Receptor CB 1 HHS Public Access. 2016;167(3):750-762.

doi:10.1016/j.cell.2016.10.004

69. Devane WA, Hanuš L, Breuer A, et al. Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science (80-)*. 1992. doi:10.1126/science.1470919

70. Herkenham M, Lynn AB, Johnson MR, Melvin LS, De Costa BR, Rice KC. Characterization and localization of cannabinoid receptors in rat brain: A quantitative in vitro autoradiographic study. *J Neurosci*. 1991. doi:10.1523/jneurosci.11-02-00563.1991

71. Devane WA, Dysarz FA, Johnson MR, Melvin LS, Howlett AC. Determination and characterization of a cannabinoid receptor in rat brain. *Mol Pharmacol*. 1988.

72. Bosier B, Muccioli GG, Hermans E, Lambert DM. Functionally selective cannabinoid receptor signalling: Therapeutic implications and opportunities. *Biochem Pharmacol*. 2010. doi:10.1016/j.bcp.2010.02.013

73. Naderi N, Ahmad-Molaei L, Aziz Ahari F, Motamedi F. Modulation of anticonvulsant effects of cannabinoid compounds by GABA-A receptor agonist in acute pentylenetetrazole model of seizure in rat. *Neurochem Res*. 2011. doi:10.1007/s11064-011-0479-1

74. Abidi AH, Alghamdi SS, Dabbous MK, Tipton DA, Mustafa SM, Moore BM. Cannabinoid type-2 receptor agonist, inverse agonist, and anandamide regulation of inflammatory responses in IL-1β stimulated primary human periodontal ligament fibroblasts. *J Periodontal Res*. 2020. doi:10.1111/jre.12765

75. Braun A, Engel T, Aguilar-Pimentel JA, et al. Beneficial effects of cannabinoids (CB) in a murine model of allergen-induced airway inflammation: Role of CB1/CB2 receptors. *Immunobiology*. 2011. doi:10.1016/j.imbio.2010.09.004

76. Marchalant Y, Rosi S, Wenk GL. Anti-inflammatory property of the cannabinoid agonist

WIN-55212-2 in a rodent model of chronic brain inflammation. *Neuroscience*. 2007. doi:10.1016/j.neuroscience.2006.11.016

77. Cooray R, Gupta V, Suphioglu C. Current Aspects of the Endocannabinoid System and Targeted THC and CBD Phytocannabinoids as Potential Therapeutics for Parkinson's and Alzheimer's Diseases: a Review. *Mol Neurobiol*. 2020. doi:10.1007/s12035-020-02054-6

78. Burgaz S, García C, Gómez-Cañas M, Muñoz E, Fernández-Ruiz J. Development of an oral treatment with the PPAR-γ-acting cannabinoid VCE-003.2 against the inflammation-driven neuronal deterioration in experimental Parkinson's disease. *Molecules*. 2019. doi:10.3390/molecules24152702

79. Tsou K, Brown S, Sañudo-Peña MC, Mackie K, Walker JM. Immunohistochemical distribution of cannabinoid CB1 receptors in the rat central nervous system. *Neuroscience*. 1998;83(2):393-411. doi:10.1016/S0306-4522(97)00436-3

80. Nalli Y, Dar MS, Bano N, et al. Analyzing the role of cannabinoids as modulators of Wnt/β-catenin signaling pathway for their use in the management of neuropathic pain. *Bioorganic Med Chem Lett*. 2019;29(9):1043-1046. doi:10.1016/j.bmcl.2019.03.013

81. Smith FL, Cichewicz D, Martin ZL, Welch SP. The enhancement of morphine antinociception in mice by Δ9- tetrahydrocannabinol. *Pharmacol Biochem Behav*. 1998;60(2):559-566. doi:10.1016/S0091-3057(98)00012-4

82. Cichewicz DL, Martin ZL, Smith FL, Welch SP. Enhancement of μ opioid antinociception by oral Δ9- tetrahydrocannabinol: Dose-response analysis and receptor identification. In: *Journal of Pharmacology and Experimental Therapeutics*. ; 1999.

83. Van Koppen CJ, Parenti C, Soderstrom K, Soliman E, Van Dross R. Cannabinoids Modulate Neuronal Activity and Cancer by CB1 and CB2 Receptor-Independent

Mechanisms. *Front Pharmacol | www.frontiersin.org*. 2017;8. doi:10.3389/fphar.2017.00720

84. Narang S, Gibson D, Wasan AD, et al. Efficacy of Dronabinol as an Adjuvant Treatment for Chronic Pain Patients on Opioid Therapy. *J Pain*. 2008. doi:10.1016/j.jpain.2007.10.018

85. Burke CC, Wildung MR, Croteau R. Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. *Proc Natl Acad Sci U S A*. 1999. doi:10.1073/pnas.96.23.13062

86. Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. *Plant J*. 2012. doi:10.1111/j.1365-313X.2012.04949.x

87. Raharjo TJ, Chang W Te, Choi YH, Peltenburg-Looman AMG, Verpoorte R. Olivetol as product of a polyketide synthase in Cannabis sativa L. *Plant Sci*. 2004. doi:10.1016/j.plantsci.2003.09.027

88. Taura F, Tanaka S, Taguchi C, et al. Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett*. 2009. doi:10.1016/j.febslet.2009.05.024

89. Gagne SJ, Stout JM, Liu E, Boubakir Z, Clark SM, Page JE. Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proc Natl Acad Sci U S A*. 2012;109(31):12811-12816. doi:10.1073/pnas.1200330109

90. Morimoto S, Komatsu K, Taura F, Shoyama Y. Purification and characterization of cannabichromenic acid synthase from Cannabis sativa. *Phytochemistry*. 1998. doi:10.1016/S0031-9422(98)00278-7

91. Sirikantaramas S, Morimoto S, Shoyama Y, et al. The Gene Controlling Marijuana Psychoactivity. *J Biol Chem*. 2004. doi:10.1074/jbc.m403693200

92. Fellermeier M, Zenk MH. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Lett*. 1998. doi:10.1016/S0014-5793(98)00450-5

93. Fellermeier M, Eisenreich W, Bacher A, Zenk MH. *Biosynthesis of Cannabinoids Incorporation Experiments with 13 C-Labeled Glucoses*.

94. Shiponi S, Bernstein N. The Highs and Lows of P Supply in Medical Cannabis: Effects on Cannabinoids, the Ionome, and Morpho-Physiology. *Front Plant Sci*. 2021;12. doi:10.3389/fpls.2021.657323

95. Lipson Feder C, Cohen O, Shapira A, et al. Fertilization Following Pollination Predominantly Decreases Phytocannabinoids Accumulation and Alters the Accumulation of Terpenoids in Cannabis Inflorescences. *Front Plant Sci*. 2021;12. doi:10.3389/fpls.2021.753847

96. Summers HM, Sproul E, Quinn JC. The greenhouse gas emissions of indoor cannabis production in the United States. *Nat Sustain*. 2021;4(7). doi:10.1038/s41893-021-00691-w

97. Mechoulam R, Lander N, University A, Zahalka J. Synthesis of the individual, pharmacologically distinct, enantiomers of a tetrahydrocannabinol derivative. *Tetrahedron: Asymmetry*. 1990. doi:10.1016/S0957-4166(00)86322-3

98. Melvin LS, Johnson MR, Harbert CA, Milne GM, Weissman A. *A Cannabinoid Derived Prototypical Analgesic1*. Vol 27.; 1984. https://pubs.acs.org/sharingguidelines.

99. Huffman JW, Dai D, Martin BR, Compton DR. Design, Synthesis and Pharmacology of Cannabimimetic Indoles. *Bioorganic Med Chem Lett*. 1994;4(4):563-566.

doi:10.1016/S0960-894X(01)80155-4

100. Pertwee RG. The therapeutic potential of drugs that target cannabinoid receptors or modulate the tissue levels or actions of endocannabinoids. *AAPS J*. 2005. doi:10.1208/aapsj070364

101. Wiley JL, Compton DR, Dai D, et al. Structure-activity relationships of indole- and pyrrole-derived cannabinoids. *J Pharmacol Exp Ther*. 1998.

102. Hermanns-Clausen M, Kneisel S, Hutter M, Szabo B, Auwärter V. Acute intoxication by synthetic cannabinoids - Four case reports. *Drug Test Anal*. 2013. doi:10.1002/dta.1483

103. Simmons J, Cookman L, Kang C, Skinner C. Clinical Toxicology Three cases of "spice" exposure Three cases of " spice " exposure. 2011. doi:10.3109/15563650.2011.584316

104. Buser GL, Gerona RR, Horowitz BZ, et al. Acute kidney injury associated with smoking synthetic cannabinoid. *Clin Toxicol*. 2014. doi:10.3109/15563650.2014.932365

105. Frieden TR, Harold Jaffe DW, Cardo DM, et al. *Morbidity and Mortality Weekly Report Centers for Disease Control and Prevention MMWR Editorial and Production Staff MMWR Editorial Board*. Vol 15.; 2013.

106. Ozturk HM, Yetkin E, Ozturk S. Synthetic Cannabinoids and Cardiac Arrhythmia Risk: Review of the Literature. *Cardiovasc Toxicol*. 2019. doi:10.1007/s12012-019-09522-z

107. Breivogel CS, Wells JR, Jonas A, et al. Comparison of the Neurotoxic and Seizure-Inducing Effects of Synthetic and Endogenous Cannabinoids with Δ9-Tetrahydrocannabinol. *Cannabis Cannabinoid Res*. 2020. doi:10.1089/can.2019.0003

108. Schneir AB, Baumbacher T. Convulsions Associated with the Use of a Synthetic Cannabinoid Product. doi:10.1007/s13181-011-0182-2

109. Every-Palmer S. Synthetic cannabinoid JWH-018 and psychosis: An explorative study.

Drug Alcohol Depend. 2011. doi:10.1016/j.drugalcdep.2011.01.012

110. Patel M, Manning JJ, Finlay DB, et al. Signalling profiles of a structurally diverse panel of synthetic cannabinoid receptor agonists. *Biochem Pharmacol*. 2020;175:113871. doi:10.1016/j.bcp.2020.113871

111. Schachtsiek J, Warzecha H, Kayser O, Stehle F. Current Perspectives on Biotechnological Cannabinoid Production in Plants. *Planta Med*. 2018;84(4):214-220. doi:10.1055/s-0043-125087

112. Schäfer H, Wink M. Medicinally important secondary metabolites in recombinant microorganisms or plants: Progress in alkaloid biosynthesis. *Biotechnol J*. 2009. doi:10.1002/biot.200900229

113. Zhu X, Liu X, Liu T, et al. Synthetic biology of plant natural products: From pathway elucidation to engineered biosynthesis in plant cells. *Plant Commun*. 2021;2(5). doi:10.1016/j.xplc.2021.100229

114. Borodina I, Nielsen J. Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. *Biotechnol J*. 2014;9(5). doi:10.1002/biot.201300445

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

116. Chang MCY, Eachus RA, Trieu W, Ro DK, Keasling JD. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat Chem Biol*. 2007;3(5). doi:10.1038/nchembio875

117. Watts KT, Lee PC, Schmidt-Dannert C. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered Escherichia coli. *BMC Biotechnol*. 2006;6. doi:10.1186/1472-6750-6-22

118. Fossati E, Ekins A, Narcross L, et al. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in Saccharomyces cerevisiae. *Nat Commun*. 2014;5. doi:10.1038/ncomms4283

119. Eudes A, Mouille M, Robinson DS, et al. Exploiting members of the BAHD acyltransferase family to synthesize multiple hydroxycinnamate and benzoate conjugates in yeast. *Microb Cell Fact*. 2016;15(1). doi:10.1186/s12934-016-0593-5

120. Luo X, Reiter MA, d'Espaux L, et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*. 2019. doi:10.1038/s41586-019-0978-9

121. Tan Z, Clomburg JM, Gonzalez R. Synthetic Pathway for the Production of Olivetolic Acid in Escherichia coli. *ACS Synth Biol*. 2018;7(8):1886-1896. doi:10.1021/acssynbio.8b00075

122. Radwan MM, ElSohly MA, Slade D, et al. Non-cannabinoid constituents from a high potency Cannabis sativa variety. *Phytochemistry*. 2008;69(14). doi:10.1016/j.phytochem.2008.07.010

123. Jin D, Dai K, Xie Z, Chen J. Secondary Metabolites Profiled in Cannabis Inflorescences, Leaves, Stem Barks, and Roots for Medicinal Purposes. *Sci Rep*. 2020;10(1). doi:10.1038/s41598-020-60172-6

124. Pertwee RG, Rock EM, Guenther K, et al. Cannabidiolic acid methyl ester, a stable synthetic analogue of cannabidiolic acid, can produce 5-HT1A receptor-mediated suppression of nausea and anxiety in rats. *Br J Pharmacol*. 2018;175(1). doi:10.1111/bph.14073

125. Wong H, Cairns BE. Cannabidiol, cannabinol and their combinations act as peripheral analgesics in a rat model of myofascial pain. *Arch Oral Biol*. 2019;104.

doi:10.1016/j.archoralbio.2019.05.028

126. Walsh KB, McKinney AE, Holmes AE. Minor Cannabinoids: Biosynthesis, Molecular Pharmacology and Potential Therapeutic Uses. *Front Pharmacol*. 2021;12. doi:10.3389/fphar.2021.777804

127. Weydt P, Hong S, Witting A, Möller T, Stella N, Kliot M. Cannabinol delays symptom onset in SOD1 (G93A) transgenic mice without affecting survival. *Amyotroph Lateral Scler Other Mot Neuron Disord*. 2005;6(3). doi:10.1080/14660820510030149

128. Maione S, Piscitelli F, Gatta L, et al. Non-psychoactive cannabinoids modulate the descending pathway of antinociception in anaesthetized rats through several mechanisms of action. *Br J Pharmacol*. 2011;162(3). doi:10.1111/j.1476-5381.2010.01063.x

129. Izzo AA, Capasso R, Aviello G, et al. Inhibitory effect of cannabichromene, a major non-psychotropic cannabinoid extracted from Cannabis sativa, on inflammation-induced hypermotility in mice. *Br J Pharmacol*. 2012;166(4). doi:10.1111/j.1476-5381.2012.01879.x

130. DeLong GT, Wolf CE, Poklis A, Lichtman AH. Pharmacological evaluation of the natural constituent of Cannabis sativa, cannabichromene and its modulation by Δ9-tetrahydrocannabinol. *Drug Alcohol Depend*. 2010;112(1-2). doi:10.1016/j.drugalcdep.2010.05.019

131. Borrelli F, Fasolino I, Romano B, et al. Beneficial effect of the non-psychotropic plant cannabinoid cannabigerol on experimental inflammatory bowel disease. *Biochem Pharmacol*. 2013;85(9). doi:10.1016/j.bcp.2013.01.017

132. Pagano E, Iannotti FA, Piscitelli F, et al. Efficacy of combined therapy with fish oil and phytocannabinoids in murine intestinal inflammation. *Phyther Res*. 2021;35(1).

doi:10.1002/ptr.6831

133. García C, Palomo-Garo C, García-Arencibia M, Ramos JA, Pertwee RG, Fernández-Ruiz J. Symptom-relieving and neuroprotective effects of the phytocannabinoid Δ 9-THCV in animal models of Parkinson's disease. *Br J Pharmacol*. 2011;163(7). doi:10.1111/j.1476-5381.2011.01278.x

134. Trantas EA, Koffas MAG, Xu P, Ververidis F. When plants produce not enough or at all: Metabolic engineering of flavonoids in microbial hosts. *Front Plant Sci*. 2015;6(JAN). doi:10.3389/fpls.2015.00007

135. Mechoulam R, Breuer A, Pereira Kapczinski F, et al. Flourinated cannabidiol (CBD) compounds, compositions and uses thereof for preparation of antioxidant medicaments. *PCT Int Appl*. 2017;(WO2017008136A2).

136. Mechoulam R, Zuardi AW, Kapczinski FP, et al. Fluorinated cbd compounds, compositions and uses thereof. *US Pat Appl Publ US20150344403*. 2015;(Copyright (C) 2018 American Chemical Society (ACS). All Rights Reserved.).

137. Fridman E, Pichersky E. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol*. 2005;8(3 SPEC. ISS.). doi:10.1016/j.pbi.2005.03.004

138. Tolia NH, Joshua-Tor L. Strategies for protein coexpression in Escherichia coli. *Nat Methods*. 2006;3(1). doi:10.1038/nmeth0106-55

139. Kim YS, Kang SW. Steady-state kinetics of malonyl-CoA synthetase from Bradyrhizobium japonicum and evidence for malonyl-AMP formation in the reaction. *Biochem J*. 1994;297(2). doi:10.1042/bj2970327

140. Beites T, Mendes M V. Chassis optimization as a cornerstone for the application of

synthetic biology based strategies in microbial secondary metabolism. *Front Microbiol*. 2015;6(SEP). doi:10.3389/fmicb.2015.00906

141. Chen YY, Galloway KE, Smolke CD. Synthetic biology: Advancing biological frontiers by building synthetic systems. *Genome Biol*. 2012;13(2). doi:10.1186/gb-2012-13-2-240

142. Li L, Liu X, Wei K, Lu Y, Jiang W. Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnol Adv*. 2019;37(5). doi:10.1016/j.biotechadv.2019.04.002

143. Bassalo MC, Garst AD, Halweg-Edwards AL, et al. Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth Biol*. 2016;5(7). doi:10.1021/acssynbio.5b00187

144. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet*. 2008;42. doi:10.1146/annurev.genet.42.110807.091442

145. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol*. 1996;260(5). doi:10.1006/jmbi.1996.0428

146. Ikemura T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol*. 1981;146(1). doi:10.1016/0022-2836(81)90363-6

147. Kurland C, Gallant J. Errors of heterologous protein expression. *Curr Opin Biotechnol*. 1996;7(5). doi:10.1016/S0958-1669(96)80050-4

148. Welch M, Villalobos A, Gustafsson C, Minshull J. You're one in a googol: Optimizing genes for protein expression. *J R Soc Interface*. 2009;6(SUPPL. 4). doi:10.1098/rsif.2008.0520.focus

149. Al-Hawash AB, Zhang X, Ma F. Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems. *Gene Reports*. 2017;9. doi:10.1016/j.genrep.2017.08.006

150. Sriubolmas N, Panbangred W, Sriurairatana S, Meevootisom V. Localization and characterization of inclusion bodies in recombinant Escherichia coli cells overproducing penicillin G acylase. *Appl Microbiol Biotechnol*. 1997;47(4). doi:10.1007/s002530050943

151. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front Microbiol*. 2014;5(APR). doi:10.3389/fmicb.2014.00172

152. Jayadeva Bhat P, Murthy TVS. Transcriptional control of the GAL/MEL regulon of yeast Saccharomyces cerevisiae: Mechanism of galactose-mediated signal transduction. *Mol Microbiol*. 2001;40(5). doi:10.1046/j.1365-2958.2001.02421.x

153. Alipourfard I, Datukishvili N, Bakhtiyari S, et al. MIG1 glucose repression in metabolic processes of Saccharomyces cerevisiae: Genetics to metabolic engineering. *Avicenna J Med Biotechnol*. 2019;11(3).

154. Chiang HL, Schekman R, Hamamoto S. Selective uptake of cytosolic, peroxisomal, and plasma membrane proteins into the yeast lysosome for degradation. *J Biol Chem*. 1996;271(17). doi:10.1074/jbc.271.17.9934

155. Hawkins KM, Smolke CD. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. *J Biol Chem*. 2006;281(19). doi:10.1074/jbc.M512317200

156. Zenke FT, Engels R, Vollenbroich V, Meyer J, Hollenberg CP, Breunig KD. Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science (80-)*. 1996;272(5268). doi:10.1126/science.272.5268.1662

157. Chasman DI, Kornberg RD. GAL4 protein: purification, association with GAL80 protein, and conserved domain structure. *Mol Cell Biol*. 1990;10(6). doi:10.1128/mcb.10.6.2916-2923.1990

158. Bram RJ, Lue NF, Kornberg RD. A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription. *EMBO J*. 1986;5(3). doi:10.1002/j.1460-2075.1986.tb04253.x

159. Johnston M. A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiol Rev*. 1987;51(4). doi:10.1128/mmbr.51.4.458-476.1987

160. Vickers CE, Bydder SF, Zhou Y, Nielsen LK. Dual gene expression cassette vectors with antibiotic selection markers for engineering in Saccharomyces cerevisiae. *Microb Cell Fact*. 2013;12(1). doi:10.1186/1475-2859-12-96

161. Hanahan D. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*. 1983;166(4). doi:10.1016/S0022-2836(83)80284-8

162. Hanahan D, Bloom F. Mechanisms of DNA transformation. *Escherichia coli Salmonella typhimurium Cell Mol Biol*. 1987;2.

163. Hashimoto H, Morikawa H, Yamada Y, Kimura A. A novel method for transformation of intact yeast cells by electroinjection of plasmid DNA. *Appl Microbiol Biotechnol*. 1985;21(5). doi:10.1007/BF00252714

164. Suga M, Hatakeyma T. High-efficiency electroporation by freezing intact yeast cells with addition of calcium. *Curr Genet*. 2003;43(3). doi:10.1007/s00294-003-0385-4

165. Becker DM, Guarente L. High-Efficiency Transformation of Yeast by Electroporation. *Methods Enzymol*. 1991;194(C). doi:10.1016/0076-6879(91)94015-5

166. Webster TD, Dickson RC. Direct selection of Saccharomyces cerevisiae resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin-resistance gene of Tn903. *Gene*. 1983;26(2-3). doi:10.1016/0378-1119(83)90194-4

167. Kearsey LJ, Yan C, Prandi N, Toogood HS, Takano E, Scrutton NS. Biosynthesis of cannabigerol and cannabigerolic acid: The gateways to further cannabinoid production. *Synth Biol*. 2023;8(1). doi:10.1093/synbio/ysad010

168. Bellotto N, Agudo-Canalejo J, Colin R, Golestanian R, Malengo G, Sourjik V. Dependence of diffusion in Escherichia coli cytoplasm on protein size, environmental conditions, and cell growth. *Elife*. 2022;11. doi:10.7554/ELIFE.82654

169. Cayley S, Lewis BA, Guttman HJ, Record MT. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *J Mol Biol*. 1991;222(2). doi:10.1016/0022-2836(91)90212-O

170. Tabaka M, Kalwarczyk T, Szymanski J, Hou S, Holyst R. The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells. *Front Phys*. 2014;2. doi:10.3389/fphy.2014.00054

171. Klumpp S, Scott M, Pedersen S, Hwa T. Molecular crowding limits translation and cell growth. *Proc Natl Acad Sci U S A*. 2013;110(42). doi:10.1073/pnas.1310377110

172. Tabaka M, Sun L, Kalwarczyk T, Hołyst R. Implications of macromolecular crowding for protein-protein association kinetics in the cytoplasm of living cells. *Soft Matter*. 2013;9(17). doi:10.1039/c3sm00013c

173. Rai M, Padh H. Expression systems for production of heterologous proteins. *Curr Sci*. 2001;80(9).

174. Marsden RL, Orengo CA. Target selection for structural genomics: an overview. *Methods*

Mol Biol. 2008;426. doi:10.1007/978-1-60327-058-8_1

175. Tokmakov AA, Kurotani A, Takagi T, et al. Multiple post-translational modifications affect heterologous protein synthesis. *J Biol Chem*. 2012;287(32). doi:10.1074/jbc.M112.366351

176. Dastmalchi M, Dhaubhadel S. Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta*. 2015;241(2). doi:10.1007/s00425-014-2200-5

177. Calderone TL, Stevens RD, Oas TG. High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in Escherichia coli. *J Mol Biol*. 1996;262(4). doi:10.1006/jmbi.1996.0524

178. Hartland RP, Vermeulen CA, Sietsma JH, Wessels JGH, Klis FM. The linkage of (1–3)-β-glucan to chitin during cell wall assembly in Saccharomyces cerevisiae. *Yeast*. 1994;10(12). doi:10.1002/yea.320101208

179. Smith AE, Zhang Z, Thomas CR, Moxham KE, Middelberg APJ. The mechanical properties of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*. 2000;97(18). doi:10.1073/pnas.97.18.9871

180. Kitamura K, Kaneko T, Yamamoto Y. Lysis of viable yeast cells by enzymes of Arthrobacter luteus. *Arch Biochem Biophys*. 1971;145(1). doi:10.1016/0003-9861(71)90053-1

181. Leighton TG, Apfel RE. The Acoustic Bubble . *J Acoust Soc Am*. 1994;96(4). doi:10.1121/1.410082

182. Doulah MS. Mechanism of disintegration of biological cells in ultrasonic cavitation. *Biotechnol Bioeng*. 1977;19(5). doi:10.1002/bit.260190504

183. Kang Y, Schneider BB, Covey TR. On the Nature of Mass Spectrometer Analyzer Contamination. *J Am Soc Mass Spectrom*. 2017;28(11). doi:10.1007/s13361-017-1747-3

184. Koes RE, Quattrocchio F, Mol JNM. The flavonoid biosynthetic pathway in plants: Function and evolution. *BioEssays*. 1994;16(2). doi:10.1002/bies.950160209

185. Yonekura-Sakakibara K, Higashi Y, Nakabayashi R. The Origin and Evolution of Plant Flavonoid Metabolism. *Front Plant Sci*. 2019;10. doi:10.3389/fpls.2019.00943

186. Goiris K, Muylaert K, Voorspoels S, et al. Detection of flavonoids in microalgae from different evolutionary lineages. *J Phycol*. 2014;50(3). doi:10.1111/jpy.12180

187. Rausher MD. The evolution of flavonoids and their genes. In: *The Science of Flavonoids*. ; 2006. doi:10.1007/978-0-387-28822-2_7

188. Taylor LP, Grotewold E. Flavonoids as developmental regulators. *Curr Opin Plant Biol*. 2005;8(3 SPEC. ISS.). doi:10.1016/j.pbi.2005.03.005

189. Tian B, Pei Y, Huang W, Ding J, Siemann E. Increasing flavonoid concentrations in root exudates enhance associations between arbuscular mycorrhizal fungi and an invasive plant. *ISME J*. 2021;15(7). doi:10.1038/s41396-021-00894-1

190. Iwashina T. Contribution to flower colors of flavonoids including anthocyanins: A review. *Nat Prod Commun*. 2015;10(3). doi:10.1177/1934578x1501000335

191. Hassan S, Mathesius U. The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *J Exp Bot*. 2012;63(9). doi:10.1093/jxb/err430

192. Kumar S, Pandey AK. Chemistry and biological activities of flavonoids: An overview. *Sci World J*. 2013;2013. doi:10.1155/2013/162750

193. Jiang C, Kim SY, Suh DY. Divergent evolution of the thiolase superfamily and chalcone

synthase family. *Mol Phylogenet Evol*. 2008;49(3). doi:10.1016/j.ympev.2008.09.002

194. Shimizu Y, Ogata H, Goto S. Type III Polyketide Synthases: Functional Classification and Phylogenomics. *ChemBioChem*. 2017;18(1). doi:10.1002/cbic.201600522

195. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol*. 2001;126(2):485-493. doi:10.1104/pp.126.2.485

196. Wanibuchi K, Zhang P, Abe T, et al. An acridone-producing novel multifunctional type III polyketide synthase from Huperzia serrata. *FEBS J*. 2007;274(4). doi:10.1111/j.1742-4658.2007.05656.x

197. Adrain C, Freeman M. New lives for old: Evolution of pseudoenzyme function illustrated by iRhoms. *Nat Rev Mol Cell Biol*. 2012;13(8). doi:10.1038/nrm3392

198. Shimada N, Aoki T, Sato S, Nakamura Y, Tabata S, Ayabe SI. A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in Lotus japonicus. *Plant Physiol*. 2003;131(3). doi:10.1104/pp.004820

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

200. Kaltenbach M, Burke JR, Dindo M, et al. Evolution of chalcone isomerase from a noncatalytic ancestor article. *Nat Chem Biol*. 2018;14(6). doi:10.1038/s41589-018-0042-3

201. Abe I, Morita H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat Prod Rep*. 2010;27(6). doi:10.1039/b909988n

202. Jiang W, Yin Q, Wu R, et al. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in Arabidopsis thaliana. *J Exp Bot*. 2015;66(22). doi:10.1093/jxb/erv413

203. Jørgensen K, Rasmussen AV, Morant M, et al. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr Opin Plant Biol*. 2005;8(3 SPEC. ISS.). doi:10.1016/j.pbi.2005.03.014

204. Wang J, Jiang Y, Sun T, Zhang C, Liu X, Li Y. Genome-Wide Classification and Evolutionary Analysis Reveal Diverged Patterns of Chalcone Isomerase in Plants. *Biomolecules*. 2022;12(7). doi:10.3390/biom12070961

205. Weng JK, Noel JP. Chemodiversity in Selaginella: A reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci*. 2013;4(MAY). doi:10.3389/fpls.2013.00119

206. Wang G, Ma X, Li D, Jiang Y, Liang Q, Hui C. Chemical composition of essential volatile oils of Selaginella spp. and their antibacterial activity. *Bangladesh J Bot*. 2018;47(3).

207. Boland MJ, Wong E. Mechanism of action of chalcone isomerase. *Bioorg Chem*. 1979;8(1). doi:10.1016/0045-2068(79)90030-0

208. Hong K, Wang L, Johnpaul A, Lv C, Ma C. Key enzymes involved in the synthesis of hops phytochemical compounds: From structure, functions to applications. *Int J Mol Sci*. 2021;22(17). doi:10.3390/ijms22179373

209. de Bruijn WJC, Levisson M, Beekwilder J, van Berkel WJH, Vincken JP. Plant Aromatic Prenyltransferases: Tools for Microbial Cell Factories. *Trends Biotechnol*. 2020;38(8). doi:10.1016/j.tibtech.2020.02.006

210. Zirpel B, Degenhardt F, Martin C, Kayser O, Stehle F. Engineering yeasts as platform organisms for cannabinoid biosynthesis. *J Biotechnol*. 2017;259:204-212. doi:10.1016/j.jbiotec.2017.07.008

211. Neudecker P, Schweimer K, Nerkamp J, et al. Allergic cross-reactivity made visible. Solution structure of the major cherry allergen Pru av 1. *J Biol Chem*. 2001;276(25). doi:10.1074/jbc.M101657200

212. Gajhede M, Osmark P, Poulsen FM, et al. X-ray and NMR structure of bet v 1, the origin of birch pollen allergy. *Nat Struct Biol*. 1996;3(12). doi:10.1038/nsb1296-1040

213. Bergh S, Uhlén M. Analysis of a polyketide synthesis-encoding gene cluster of Streptomyces curacoi. *Gene*. 1992;117(1). doi:10.1016/0378-1119(92)90501-F

214. Radauer C, Lackner P, Breiteneder H. The Bet v 1 fold: An ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol Biol*. 2008;8(1). doi:10.1186/1471-2148-8-286

215. Cohen SN, Chang ACY, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. *Proc Natl Acad Sci U S A*. 1973;70(11). doi:10.1073/pnas.70.11.3240

216. Struhl K, Stinchcomb DT, Scherer S, Davis RW. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc Natl Acad Sci U S A*. 1979;76(3). doi:10.1073/pnas.76.3.1035

217. Gnügge R, Rudolf F. Saccharomyces cerevisiae Shuttle vectors. *Yeast*. 2017;34(5). doi:10.1002/yea.3228

218. Bishé B, Taton A, Golden JW. Modification of RSF1010-Based Broad-Host-Range Plasmids for Improved Conjugation and Cyanobacterial Bioprospecting. *iScience*. 2019;20. doi:10.1016/j.isci.2019.09.002

219. Terpe K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl Microbiol*

Biotechnol. 2006;72(2). doi:10.1007/s00253-006-0465-8

220. Stanley KK, Luzio JP. Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. *EMBO J*. 1984;3(6). doi:10.1002/j.1460-2075.1984.tb01988.x

221. McDermott PF, White DG, Podglajen I, Alekshun MN, Levy SB. Multidrug resistance following expression of the Escherichia coli marA gene in Mycobacterium smegmatis. *J Bacteriol*. 1998;180(11). doi:10.1128/jb.180.11.2995-2998.1998

222. Mourand G, Touzain F, Jouy E, Fleury MA, Denamur E, Kempf I. Rare spontaneous loss of multiresistance gene carrying IncI/ST12 plasmid in Escherichia coli in pig microbiota. *Antimicrob Agents Chemother*. 2016;60(10). doi:10.1128/AAC.00864-16

223. Nordstrom K, Austin SJ. Mechanisms that contribute to the stable segregation of plasmids. *Annu Rev Genet*. 1989;23. doi:10.1146/annurev.ge.23.120189.000345

224. Ehrlich SD, Bruand C, Sozhamannan S, et al. Plasmid replication and structural stability in Bacillus subtilis. *Res Microbiol*. 1991;142(7-8). doi:10.1016/0923-2508(91)90067-K

225. Lee SH, Soderlund DM. Cloning and maintenance of the housefly sodium channel gene using low copy number vector and two sequential host strains. *J Asia Pac Entomol*. 2009;12(1). doi:10.1016/j.aspen.2008.10.005

226. Vu K, Bautos J, Hong MP, Gelli A. The functional expression of toxic genes: Lessons learned from molecular cloning of CCH1, a high-affinity Ca2+ channel. *Anal Biochem*. 2009;393(2). doi:10.1016/j.ab.2009.06.039

227. Cui L, Bikard D. Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic Acids Res*. 2016;44(9). doi:10.1093/nar/gkw223

228. Costantino N, Court DL. Enhanced levels of lambda Red-mediated recombinants in

mismatch repair mutants. *Proc Natl Acad Sci USA*. 2003;100(26).

229. Muyrers JPP, Zhang Y, Testa G, Stewart AF. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res*. 1999;27(6). doi:10.1093/nar/27.6.1555

230. Wang HH, Isaacs FJ, Carr PA, et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. 2009;460(7257). doi:10.1038/nature08187

231. Nyerges Á, Csörgo B, Nagy I, et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proc Natl Acad Sci U S A*. 2016;113(9). doi:10.1073/pnas.1520040113

232. Wannier TM, Nyerges A, Kuchwara HM, et al. Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci U S A*. 2020;117(24). doi:10.1073/pnas.2001588117

233. Schubert MG, Goodman DB, Wannier TM, et al. High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proc Natl Acad Sci U S A*. 2021;118(18). doi:10.1073/pnas.2018181118

234. Sauer B. Functional Expression of the cre-lox Site-Specific Recombination System in the Yeast Saccharomyces cerevisiae . *Mol Cell Biol*. 1987;7(6). doi:10.1128/mcb.7.6.2087-2096.1987

235. Bellaiche Y, Mogila V, Perrimon N. I-SceI endonuclease, a new tool for studying DNA double strand break repair mechanisms in Drosophila. *Genetics*. 1999;152(3). doi:10.1093/genetics/152.3.1037

236. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*. 2010;11(9). doi:10.1038/nrg2842

237. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science (80-)*. 2009;326(5959). doi:10.1126/science.1178817

238. Boch J, Scholze H, Schornack S, et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (80-)*. 2009;326(5959). doi:10.1126/science.1178811

239. Schornack S, Meyer A, Römer P, Jordan T, Lahaye T. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *J Plant Physiol*. 2006;163(3). doi:10.1016/j.jplph.2005.12.001

240. Römer P, Hahn S, Jordan T, Strauß T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science (80-)*. 2007;318(5850). doi:10.1126/science.1144958

241. Sun N, Zhao H. Transcription activator-like effector nucleases (TALENs): a highly efficient and versatile tool for genome editing. *Biotechnol Bioeng*. 2013;110(7). doi:10.1002/bit.24890

242. Nemudryi AA, Valetdinova KR, Medvedev SP, Zakian SM. TALEN and CRISPR/Cas genome editing systems: Tools of discovery. *Acta Naturae*. 2014;6(22). doi:10.32607/20758251-2014-6-3-19-40

243. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-)*. 2014;346(6213). doi:10.1126/science.1258096

244. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014;32(4). doi:10.1038/nbt.2842

245. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*. 2014;3. doi:10.7554/eLife.04766

246. Mali P, Aach J, Stranges PB, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*. 2013;31(9). doi:10.1038/nbt.2675

247. Dicarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res*. 2013;41(7). doi:10.1093/nar/gkt135

Supplementary Figures

Sample	HTAL *(E. coli)*		
	Resp.	Final Conc.	S/N
CC 0.1ppb	26.3207	0.1550	33.8891
CC 1ppb	5640.4120	0.5542	3446.9391
CC 5ppb	35240.0369	4.5094	19625.3217
CC 10ppb	73876.6539	9.8295	48149.5885
CC 50ppb	447431.0105	35.8539	13099.7571
CC 100ppb	857137.3900	101.0519	48815.8357
Control A	12.7957	0.1549	2.9079
Control B	8.2175	0.1550	12.3322
Control C	22.1147	0.1550	3.0035
Control D	20.3618	0.1550	10.8476
CsOAC+CsTKS A	36.4264	0.1556	5.4084
CsOAC+CsTKS B	11.6660	0.1553	5.8301
CsOAC+CsTKS C	4.7454	0.1547	1.1611
CsOAC+CsTKS D	10.2045	0.1550	1.1387
CsOAC+CsTKS+CsAAE1 A	118.7467	0.1588	38.4561
CsOAC+CsTKS+CsAAE1 B	91.0197	0.1566	38.4951
CsOAC+CsTKS+CsAAE1 C	15.2727	0.1552	5.5294
CsOAC+CsTKS+CsAAE1 D	11.6652	0.1549	15.4144
CsOAC+CsTKS+CsAAE1+CsCHI-L A	14.7468	0.1554	3.4804
CsOAC+CsTKS+CsAAE1+CsCHI-L B	26.4669	0.1561	3.7053
CsOAC+CsTKS+CsAAE1+CsCHI-L C	13.3262	0.1551	4.9180
CsOAC+CsTKS+CsAAE1+CsCHI-L D	59.3714	0.1579	81.5777

Data tables from LC-MS/MS and MassHunter analysis of HTAL in *E. coli* SBS samples
Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control and experimental samples from *E. coli* in vivo feeding experiments fed 12mM sodium malonate and 4mM sodium hexanoate and allowed to react for 48 hours.

Sample	PDAL *(E. coli)*		
	Resp.	Final Conc.	S/N
CC 0.1ppb	747.0951	0.1974	16.4591

CC 1ppb	8135.0417	0.6132	290.1713
CC 5ppb	34421.0566	3.3358	291.9823
CC 10ppb	69869.7647	6.9612	1748.7248
CC 50ppb	534313.2987	31.7509	1378.0369
CC 100ppb	1202508.4556	104.9924	10694.4003
Control A	18328.0935	0.5894	525.7651
Control B	9115.3909	0.5667	100.7601
Control C	14330.2934	0.4145	350.8589
Control D	31225.8682	0.7226	522.8341
CsOAC+CsTKS A	3033.9772	0.2521	79.2338
CsOAC+CsTKS B	1917.3971	0.2762	63.8653
CsOAC+CsTKS C	1340.6808	0.2267	49.7476
CsOAC+CsTKS D	1006.8270	0.2201	27.9826
CsOAC+CsTKS+CsAAE1 A	1277.7521	0.2203	15.9976
CsOAC+CsTKS+CsAAE1 B	951.2735	0.2026	11.9580
CsOAC+CsTKS+CsAAE1 C	732.2982	0.2095	19.1572
CsOAC+CsTKS+CsAAE1 D	584.6814	0.2006	9.2086
CsOAC+CsTKS+CsAAE1+CsCHI-L A	2051.1567	0.2801	11.4230
CsOAC+CsTKS+CsAAE1+CsCHI-L B	1315.2957	0.2460	12.9567
CsOAC+CsTKS+CsAAE1+CsCHI-L C	966.8593	0.2164	15.2465
CsOAC+CsTKS+CsAAE1+CsCHI-L D	1707.6236	0.2578	23.4521

Data tables from LC-MS/MS and MassHunter analysis of PDAL in *E. coli* SBS samples

Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control and experimental samples from *E. coli* in vivo feeding experiments fed 12mM sodium malonate and 4mM sodium hexanoate and allowed to react for 48 hours.

	OA *(E. coli)*		
Sample	Resp.	Final Conc.	S/N
CC 0.1ppb	193.5076	0.1216	4.0983
CC 1ppb	15699.8330	0.6530	218.7711
CC 5ppb	85108.0411	5.1610	482.0447
CC 10ppb	174526.4988	11.0752	860.7039
CC 50ppb	951430.7042	36.5061	6836.8556
CC 100ppb	1754062.7369	99.0891	4239.4849
Control A	4244.1751	0.1802	24.2695
Control B	2580.0425	0.1895	52.1000

Control C	4006.3331	0.1611	23.4867
Control D	2503.6473	0.1476	25.2437
CsOAC+CsTKS A	1850.8776	0.1457	21.2936
CsOAC+CsTKS B	1167.7444	0.1552	13.8029
CsOAC+CsTKS C	2227.4401	0.1632	21.3655
CsOAC+CsTKS D	1257.0599	0.1470	47.2796
CsOAC+CsTKS+CsAAE1 A	2372.0020	0.1606	24.9852
CsOAC+CsTKS+CsAAE1 B	434.9631	0.1246	6.2513
CsOAC+CsTKS+CsAAE1 C	2477.5050	0.1704	13.4460
CsOAC+CsTKS+CsAAE1 D	1533.9858	0.1439	19.8264
CsOAC+CsTKS+CsAAE1+CsCHI-L A	3570.7079	0.2254	46.0000
CsOAC+CsTKS+CsAAE1+CsCHI-L B	2210.7198	0.1847	11.9720
CsOAC+CsTKS+CsAAE1+CsCHI-L C	1755.8024	0.1551	27.2076
CsOAC+CsTKS+CsAAE1+CsCHI-L D	1860.4855	0.1702	4.6263

Data tables from LC-MS/MS and MassHunter analysis of Olivetolic Acid in *E. coli* SBS samples

Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control and experimental samples from *E. coli* in vivo feeding experiments fed 12mM sodium malonate and 4mM sodium hexanoate and allowed to react for 48 hours.

	OA *(S. cerevisiae)*		
Sample	Resp	Final Conc.	S/N
CC 0.1ppb	3272.5269	0.0463	51.5059
CC 1ppb	31189.5197	1.3738	595.3190
CC 5ppb	135632.7463	5.5677	1602.0048
CC 10ppb	321610.2879	14.8703	6215.9773
CC 50ppb	1402508.8521	55.8544	5018.3190
CC 100ppb	2419715.9391	93.2577	20976.9289
Control A	9330.5045	0.4594	57.6135
Control B	7825.2037	0.4274	91.7744
Control C	7698.9329	0.3606	64.7617
Control D	12076.9874	0.4560	135.0061
CsOAC+CsTKS A	709590.5599	99.7061	4363.9673
CsOAC+CsTKS B	100682.9192	31.5989	∞
CsOAC+CsTKS C	201859.1991	63.0644	∞
CsOAC+CsTKS D	66320.0932	9.5658	1829.9156

CsOAC+CsTKS+CsAAE1 A	220415.9105	37.9624	3724.3637
CsOAC+CsTKS+CsAAE1 B	445656.6734	60.4091	∞
CsOAC+CsTKS+CsAAE1 C	1366894.9153	219.7375	11620.2598
CsOAC+CsTKS+CsAAE1 D	506821.9874	80.4529	12366.6419
CsOAC+CsTKS+CsAAE1+CsCHI-L A	536259.6821	54.6715	3967.0953
CsOAC+CsTKS+CsAAE1+CsCHI-L B	670384.3463	84.1840	∞
CsOAC+CsTKS+CsAAE1+CsCHI-L C	77099.9983	13.7809	∞
CsOAC+CsTKS+CsAAE1+CsCHI-L D	150603.2500	19.4340	∞
LOD1.d	34187.3164	1.3435	580.8674
LOD2.d	34144.0609	1.3013	496.8591
LOD3.d	33221.5740	1.3447	743.2428
LOD4.d	34560.8382	1.2749	407.7405
LOD5.d	34532.5416	1.3190	490.8349
LOD6.d	33857.3725	1.3065	526.4011
LOD7.d	34534.7240	1.3207	413.6557
LOD8.d	34206.8858	1.3218	680.4927
LOD9.d	34111.7878	1.3541	378.7411
LOD10.d	33988.8665	1.3288	505.9439

Data tables from LC-MS/MS and MassHunter analysis of Olivetolic Acid in *S. cerevisiae* SBS samples

Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control, experimental samples from *S. cerevisiae* in vivo feeding experiments fed 12mM malonic and 4mM hexanoic acid and allowed to react for 48 hours as well as injections to determine the limit of detection (LOD) and limit of quantification (LOQ).

	HTAL *(S. cerevisiae)*		
Sample	Resp	Final Conc.	S/N
CC 0.1ppb	678.3354	0.1445	82.4032
CC 1ppb	9110.8602	0.8069	1634.0634
CC 5ppb	48923.5331	3.4576	6637.7879
CC 10ppb	130737.9595	10.0854	15223.6133
CC 50ppb	811689.0473	53.2051	34320.4646
CC 100ppb	1553878.4038	98.4004	31272.2965
Control A	145.6950	0.0945	38.6046
Control B	394.1042	0.1264	48.0697
Control C	327.6762	0.1144	84.6128
Control D	226.4847	0.0975	∞

CsOAC+CsTKS A	538.0743	0.2031	20.8528
CsOAC+CsTKS B	187.1372	0.1757	50.7977
CsOAC+CsTKS C	169.8952	0.1662	13.9207
CsOAC+CsTKS D	391.8067	0.1730	28.6931
CsOAC+CsTKS+CsAAE1 A	241.1007	0.1473	17.6963
CsOAC+CsTKS+CsAAE1 B	248.6286	0.1344	20.7270
CsOAC+CsTKS+CsAAE1 C	321.5916	0.1638	13.7838
CsOAC+CsTKS+CsAAE1 D	250.2598	0.1442	32.9295
CsOAC+CsTKS+CsAAE1+CsCHI-L A	131.7484	0.1011	20.1131
CsOAC+CsTKS+CsAAE1+CsCHI-L B	494.3257	0.1809	55.8065
CsOAC+CsTKS+CsAAE1+CsCHI-L C	508.2829	0.2295	34.2869
CsOAC+CsTKS+CsAAE1+CsCHI-L D	1154.7734	0.3251	194.1130
LOD1.d	9785.1998	0.7780	708.8333
LOD2.d	9718.9395	0.7545	639.0194
LOD3.d	9569.8292	0.7831	492.3889
LOD4.d	9675.9750	0.7313	837.4402
LOD5.d	9987.1708	0.7737	1059.8816
LOD6.d	9463.9009	0.7447	552.3278
LOD7.d	9534.6699	0.7430	885.0769
LOD8.d	9584.2767	0.7533	1509.9585
LOD9.d	9608.3954	0.7718	861.1858
LOD10.d	9401.6013	0.7479	586.3641

Data tables from LC-MS/MS and MassHunter analysis of HTAL in *S. cerevisiae* SBS samples

Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control, experimental samples from *S. cerevisiae* in vivo feeding experiments fed 12mM malonic and 4mM hexanoic acid and allowed to react for 48 hours as well as injections to determine the limit of detection (LOD) and limit of quantification (LOQ).

	PDAL *(S. cerevisiae)*		
Sample	Resp	Final Conc.	S/N
CC 0.1ppb	708.9909	0.1637	43.2650
CC 1ppb	13980.4532	0.7948	329.2597
CC 5ppb	73302.3448	3.1692	1216.4071
CC 10ppb	191867.0867	8.9609	1899.3128

CC 50ppb	1287170.8075	50.8274	13707.4574
CC 100ppb	2680003.8152	102.1840	30628.9656
Control A	51347.5827	3.4120	3231.4779
Control B	125990.9695	9.2382	13260.4664
Control C	49001.6380	3.3063	1424.3808
Control D	127365.9115	6.3903	3627.5850
CsOAC+CsTKS A	1339.0995	0.3084	19.1679
CsOAC+CsTKS B	4690.2404	1.5815	17.0053
CsOAC+CsTKS C	5886.6735	1.9411	33.3692
CsOAC+CsTKS D	6385.4895	1.0451	189.0340
CsOAC+CsTKS+CsAAE1 A	2406.1063	0.5329	78.4172
CsOAC+CsTKS+CsAAE1 B	3122.9225	0.5411	17.4697
CsOAC+CsTKS+CsAAE1 C	9691.5771	1.6606	340.8248
CsOAC+CsTKS+CsAAE1 D	2414.0397	0.5012	45.5693
CsOAC+CsTKS+CsAAE1+CsCHI-L A	4751.6853	0.6017	142.4290
CsOAC+CsTKS+CsAAE1+CsCHI-L B	32250.6616	4.1249	822.4890
CsOAC+CsTKS+CsAAE1+CsCHI-L C	4054.9873	0.8451	78.0415
CsOAC+CsTKS+CsAAE1+CsCHI-L D	7124.4472	1.0363	26.7801
LOD1.d	16240.2779	0.8208	335.6563
LOD2.d	16499.8526	0.8127	343.8470
LOD3.d	15851.3455	0.8244	447.5779
LOD4.d	16016.2482	0.7723	248.6056
LOD5.d	15702.7422	0.7799	202.7509
LOD6.d	15679.3806	0.7863	273.1095
LOD7.d	15800.5624	0.7847	373.1272
LOD8.d	15761.0446	0.7899	247.8725
LOD9.d	15581.1491	0.7987	243.5849
LOD10.d	15941.6338	0.8051	133.0358

Data tables from LC-MS/MS and MassHunter analysis of PDAL in *S. cerevisiae* SBS samples

Values reported from MassHunter software analysis following 10uL sample injections on LC-MS/MS. Values measured include instrument response (Resp.), calculated concentration of the compound (Final conc.) and signal to noise ratios of each response (S/N). Concentration given in ppb (parts per billion). Samples include calibration curve (CC), control, experimental samples from *S. cerevisiae* in vivo feeding experiments fed 12mM malonic and 4mM hexanoic acid and allowed to react for 48 hours as well as injections to determine the limit of detection (LOD) and limit of quantification (LOQ).