Characterization of acyl-lipid thioesterase (ALT) enzymes from plants

By

Rebecca (Kris) Kalinger

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy:

Biology

Carleton University

Ottawa, Ontario

© 2023, Rebecca Kalinger
Abstract

Medium-chain (6-14 carbon) fatty acids and their oleochemical derivatives are high-value compounds used in the fuel, food, pharmaceutical, and plastics industries, among others. Sourcing these chemicals sustainably while meeting market demands is a global challenge. Plants can naturally produce medium-chain fatty acids through the action of specialized acyl-acyl carrier protein (acyl-ACP) thioesterase enzymes, which prematurely terminate fatty acid biosynthesis. Nearly all plant species possess medium-chain acyl-ACP thioesterases belonging to the ACYL-LIPID THIOESTERASE (ALT) family. Studies of ALTs from wild tomato and Arabidopsis thaliana demonstrated that they have the unusual ability to act on intermediates of fatty acid biosynthesis, giving rise to both fully reduced and β-keto fatty acids that likely occupy several roles in secondary metabolism. However, despite being highly conserved across the plant kingdom, no ALTs from additional species had been characterized since, and the biochemical basis of their atypical activity was not understood. In this thesis, I characterize the biochemical functions of ALTs from diverse plant taxa, elucidate structure-function relationships in these thioesterases, and explore their biotechnological potential as sustainable sources of medium-chain fatty acids and the precursors of valuable methylketones. By profiling the substrate specificities of ALTs from several plant species through heterologous expression in E. coli bacteria, I demonstrate that a wide range of catalytic diversity has evolved within the ALT family, and that most plants possess multiple ALTs with different chain-length and oxidation state preferences. I provide evidence that ALTs also produce 3-hydroxy fatty acids, which have distinct applications as components of surfactants and lubricants. By overexpressing ALT enzymes in Arabidopsis thaliana seeds and Nicotiana benthamiana leaves, I demonstrate that, if optimized for greater catalytic efficiency, ALTs can be used as tools to increase medium-chain
fatty acid content in commercially important species such as oilseed crops. Through targeted mutagenesis experiments guided by homology modelling, I also identify and characterize amino acid motifs that dictate ALT substrate specificity. The findings presented here can guide further applied research to engineer ALT enzymes for use in the sustainable, large-scale production of medium-chain oleochemicals, and inform future experiments to characterize the biological functions of ALTs, which are mostly unknown.
Acknowledgements

I don’t know how I’ll fit a decade’s worth of thank-yous in two pages, but I’ll try! First and foremost, I would like to sincerely thank my undergraduate and graduate thesis supervisor, Dr. Owen Rowland. Through your mentorship and support over the past ten years, I’ve grown into a capable scientific researcher. You have taught me to be a clear communicator, thorough experimenter, insightful thinker, and a more confident person overall! Dr. Rowland, thank you so much for taking a chance on a transfer student who cold e-mailed you out of nowhere. I am so grateful for all the amazing opportunities that you have provided me during my time in university, from neat side projects, to travelling to conferences, to teaching. I’m also very thankful for all of the freedom that you have allowed me in the lab, even letting me bring some of my more out-there experiment ideas to life!

I would also like to thank my advisory committee members, Dr. Allyson MacLean and Dr. Kyle Biggar, for their constructive feedback and kind words of encouragement over the course of my studies. Thank you for all your help with experimental design, and for keeping me on my toes with my genetics knowledge and terminology! Dr. MacLean, I am so excited to join the lab as the next step in my scientific journey. Thank you for having me!

I’m also extremely grateful to Dr. Ian Pulsifer for taking the time to teach me so many experimental techniques, from how to clone a gene, to how to do a Western Blot, to how to fix the gas chromatograph when it’s on bad behaviour.

I hope that in the future, my career will lead me to a workplace that is just as wonderful as Carleton’s Biology Department. From the minute I walked into the Nesbitt Building and was introduced to everyone, I felt like I was being welcomed into a large, happy family. I have never seen so much kindness in one place. To my wonderful labmates past and present, thank you for
your support, camaraderie, and endless supply of cat photos to sustain me on the darkest days. I’ll really miss all the chats we’d have while waiting the requisite two hours to go to the plant growth rooms. Thank you especially to the Rowland Lab ALT Squad members, Alicia Morewood, Alexandra King, and Andrew Culhane, for your hard work! Your findings answered some really important questions that were a huge help in determining the direction of my thesis.

I also want to thank my former students for being a joy to teach, and for pushing me to become a better science communicator. You asked the most difficult and thought-provoking questions of anyone, and your enthusiasm kept me going. Seeing you grow as scientists has been so wonderful, and I wish you all the best in your future endeavours!

I would also like to thank my family for believing in me even when I didn’t believe in myself. Marie, my beloved, thank you for helping me build my confidence in all areas of my life, reminding me that it’s okay to take breaks, and for hyping up my research like it was the most interesting th. Just think, now we’ll be able to run errands without me asking if we can stop by the lab quickly on the way home! Mom, thank you for refusing to let me talk myself out of chasing my dreams. I now understand that you saw I had the potential to succeed long before I did. Dad, I’m grateful for all of your sage advice about life, the universe, and everything, for the dinner invitations, and for the practical engineering tips. Last, but most certainly not least, this thesis is dedicated in loving memory of my grandfather Paul Strigner, gentleman and chemist, who introduced me to the wonderful world of science as a child. I will always fondly remember the kitchen chemistry we did, taking a look at bugs and flowers under the microscope together, and listening to your lessons on natural history. You started me down this path, and I finally made it, Grandpa! I hope I’ve done you proud.
Preface

This thesis follows the integrated article thesis format. Data and text in Chapters 1 – 4 are taken from peer-reviewed journal articles for which I am the primary author. The publications from which Chapters 2 – 4 are derived have been reproduced in their entirety for this thesis. Some modifications to the text and figures have been made where necessary to reflect more recent findings, or to minimize redundancy and improve cohesiveness of the thesis. This has been done in order to present “a coherent account of a unified research project”, as mandated by Carleton University’s “Criteria for Integrated Article Thesis” under the Integrated Thesis Policy (Section 12.4A of Graduate Calendar Regulations).

Chapter 1 includes text and figures that originally appeared in the review paper:


[https://doi.org/10.1002/lipd.12226](https://doi.org/10.1002/lipd.12226)

*Statement of contribution for this article:*

The subject of this review article and the list of topics discussed therein was conceptualized by O. Rowland, S. Hepworth, I. Pulsifer, and myself. I prepared the manuscript text, with guidance from all other authors. Tables and figures in this article were generated by me, with guidance from I. Pulsifer. All text and figures from this article are reproduced with permission of John Wiley & Sons publishers.
**Chapter 2 is derived from the research paper:**


[https://doi.org/10.1016/j.plaphy.2018.03.013](https://doi.org/10.1016/j.plaphy.2018.03.013)

*Statement of contribution for this article:*

The experiments described in this chapter were designed by O. Rowland, I. Pulsifer, and myself. I conducted all experiments and data analysis, and prepared the manuscript text with guidance from O. Rowland and I. Pulsifer. All text and figures from this article are reproduced with permission of Elsevier publishers.

*Other authors’ contributions:*

- I. Pulsifer generated plasmid constructs for expressing *A. thaliana* ALTs in *E. coli* bacteria.

*Additional contributions by others:*

- Karl Wasslen (Carleton Mass Spectrometry Centre) ran gas chromatography-mass spectrometry (GC-MS) on select samples.

**Chapter 3 is derived from the research paper:**


[https://doi.org/10.1002/lipd.12299](https://doi.org/10.1002/lipd.12299)

*Statement of contribution for this article:*

The experiments described in this chapter were designed by O. Rowland, I. Pulsifer, and


myself. All data presented in this article originates from experiments that I conducted myself. I prepared the manuscript text and figures, with guidance from all other authors. Text and figures from this article are reproduced with permission of John Wiley & Sons publishers.

**Experimental work conducted by me:**

- Fluorescent screening of seeds from segregating *Arabidopsis thaliana* plant lines to identify pure-breeding transgenic lines overexpressing *Arabidopsis ALTI – 4*, and propagation of these pure-breeding lines

- Screening the seed oil composition of multiple pure-breeding *Arabidopsis* lines via gas chromatography – flame ionization detection (GC-FID) to identify three representative, independently transformed lines for each ALT

- Quantitative analysis of the seed oil composition of representative transgenic *Arabidopsis* and *Camelina sativa* plant lines overexpressing ALTs in seeds via GC-FID and GC-MS

- Generation of plasmid constructs for transient overexpression of *Arabidopsis ALTI – 4* in *Nicotiana benthamiana* leaves

- All experimental work and data analysis related to transient overexpression of *Arabidopsis* ALTs in *Nicotiana benthamiana* leaves

- Imaging of transgenic *Arabidopsis* seeds, and of *N. benthamiana* leaves post-injection
Experimental work conducted by other authors:

- I. Pulsifer and D. Williams transformed *Arabidopsis thaliana* plant lines with plasmid constructs for seed-specific overexpression of *Arabidopsis ALT1 – 4*, screened for positive transformants, and propagated these lines.

- D. Williams performed initial screening of seed oil composition in segregating transgenic *Arabidopsis* plant lines via GC-FID analysis (this preliminary dataset was not included in the article)

- A. Ahmadi Pirshahid generated transgenic *Camelina sativa* lines overexpressing *Arabidopsis ALT1* and *ALT4* in seeds, and performed initial screening of seed oil composition in segregating lines via GC-FID analysis (this preliminary dataset was not included in the article.)

Additional contributions by others:

- Cara Beatty (Carleton University) generated plasmid constructs for overexpression of *Solanum hirsutum* subsp. *glabratum MKS1* and *MKS2* in *Nicotiana benthamiana*.

- A plasmid construct for the overexpression of *Cuphea palustris* FatB2 in *Nicotiana benthamiana* leaves was provided by Professor Per Hofvander (Swedish University of Agricultural Sciences).

- Karl Wasslen (Carleton Mass Spectrometry Centre) ran gas chromatography-mass spectrometry (GC-MS) on select samples.
Chapter 4 is derived from the research paper:


https://doi.org/10.1186/s12870-022-04003-y

Statement of contribution for this article:

The experiments described in this chapter were designed by O. Rowland and myself. I conducted all experiments and data analysis. All text and figures from this article are reproduced with permission of Springer Nature publishers.

Additional contributions by others:

- Karl Wasslen (Carleton Mass Spectrometry Centre) ran gas chromatography-mass spectrometry (GC-MS) on select samples.
Chapter 1: Fatty acyl thioesterases in plant lipid metabolism: diverse functions and biotechnological applications

1.1 – Introduction

1.2 - Overview of plant acyl lipid metabolism

1.2.1 Fatty acids are metabolically active as acyl-ACP or acyl-CoA thioesters.

1.2.2 De novo fatty acid biosynthesis in the plant cell

1.2.3 Further metabolism of fatty acids

1.3 – Introduction to fatty acyl thioesterases

1.3.1 Plant fatty acyl thioesterases possess “hot-dog fold” domain structure

1.3.2 Acyl-CoA thioesterases from plants currently elude functional characterization

1.3.3 FatA- and FatB-type acyl-ACP thioesterases release acyl chains from the de novo fatty acid biosynthesis cycle

1.3.4 Fat-type acyl-ACP thioesterases in specialized metabolism

1.3.5 ACYL-LIPID THIOESTERASE (ALT) enzymes are a novel class of acyl-ACP thioesterase that produce medium-chain fatty acids as specialized metabolites

1.4 – Biotechnological applications of plant acyl thioesterases
1.4.1 Acyl-ACP thioesterases have biotechnological potential as sources of valuable medium-chain oleochemicals .................................................... 27

1.4.2 Recombinant acyl-ACP thioesterases must be optimized for efficient production of medium-chain fatty acids .............................................. 30

1.5 – Thesis Objectives........................................................................................................ 33

Chapter 2: Elucidating the substrate specificities of acyl-lipid thioesterases from diverse plant taxa ........................................................................ 35

2.1 – Abstract ....................................................................................................................... 35

2.2 – Introduction .................................................................................................................. 36

2.3 – Materials and Methods............................................................................................... 39

  2.3.1 Phylogenetic analysis of ALT protein sequences ..................................................... 39

  2.3.2 Generating DNA constructs for expression of ALT proteins in K27(DE3) 
     E. coli ................................................................................................................................ 39

  2.3.3 Profiling and quantification of lipids by GC-FID and GC-MS ................................. 41

  2.3.4 Esterification of free fatty acids in bacterial culture media ................................. 42

  2.3.5 Silylation of fatty acids produced by ALT-expressing E. coli to assist in the 
     identification of 3-hydroxy fatty acids ................................................................. 42

  2.3.6 Detection of ALT protein expression via Western blotting .............................. 43

  2.3.7 Homology modelling and structure-based phylogenetic analysis of ALT 
     proteins .................................................................................................................... 44

2.4 – Results .......................................................................................................................... 44

  2.4.1 Identification and selection of ALT enzymes in diverse plant taxa ................. 44

  2.4.2 Medium-chain 3-hydroxy fatty acids are observed as novel products of ALT 
     enzymes in E. coli ................................................................................................. 48

  2.4.3 ALT enzymes can be grouped according to their catalytic preferences in 
     E. coli .................................................................................................................... 53

  2.4.4 ALT proteins are expressed at different levels in E. coli ............................ 61
2.4.5 ALT enzymes with exclusive preference for 6-10 carbon substrates appear to differ significantly from other ALTs in their tertiary structure.............. 63

2.5 – Discussion........................................................................................................................................66
2.5.1 ALTs have highly diverse substrate specificities......................... 66
2.5.2 ALT enzymes can act on 3-hydroxyacyl-ACP substrates.................. 68
2.5.3 Evolutionary origins of ALTs................................................................. 69
2.5.4 Potential biological roles of ALT enzymes........................................ 70
2.5.5 E. coli is a useful platform for ALT protein expression.................... 72
2.5.6 ALTs generate products of biotechnological significance................. 74

2.6 – Conclusions......................................................................................................................................75

Chapter 3: Production of C6-C14 medium-chain fatty acids in seeds and leaves via overexpression of ALT-type acyl-ACP thioesterases..................................................77

3.1 – Abstract..............................................................................................................................................77
3.2 – Introduction........................................................................................................................................78
3.3 – Materials and Methods.....................................................................................................................82
3.3.1 Plant growth conditions...................................................................................................................82
3.3.2 Seed-specific expression of AtALT1-4 in Arabidopsis and Camelina..... 83
3.3.3 Fatty acid composition analysis of Arabidopsis and Camelina seed oils...84
3.3.4 Genetic construct assembly for transient ALT gene expression in N. benthamiana.................................................................85
3.3.5 N. benthamiana transient assay.....................................................................................................86
3.3.6 Total lipid extractions from N. benthamiana leaves and GC analysis of fatty acids....................................................................................87
3.3.7 Statistical analysis of data.............................................................................................................88

3.4 – Results...............................................................................................................................................89
3.4.1 Seed-specific overexpression of ALT-type thioesterases increases medium-chain fatty acid content in seed triacylglycerol.........................89
3.4.2 Medium-chain fatty acid and methylketone production in N. benthamiana leaves as a result of overexpression of ALT-type thioesterases.................102

3.5 – Discussion ...........................................................................................................108

3.5.1 ALT enzymes show diverse chain-length and oxidation state preferences in planta .................................................................108

3.5.2 Recombinant ALT-like thioesterases could be co-expressed with medium-chain acyltransferases to increase MCFA content in seed triacylglycerol .................................................................111

3.5.3 AtALT4 overexpression leads to significant restructuring of seed triacylglycerol composition and morphological defects in Arabidopsis and Camelina seeds .................................................................114

3.5.4 Production of large amounts of methylketones in N. benthamiana by ShMKS1 / ShMKS2 co-expression, but not ShMKS1 / AtALT co-expression, suggests a physical interaction between ShMKS1 and ShMKS2...........117

3.6 – Conclusions .......................................................................................................118

Chapter 4: Structural determinants of substrate specificity in ALT-type acyl-ACP thioesterases.................................119

4.1 – Abstract ..............................................................................................................119

4.2 – Introduction .......................................................................................................120

4.3 – Materials and Methods ....................................................................................123

4.3.1 Sequence data collection ..............................................................................123

4.3.2 Assembly of DNA constructs for the expression of chimeric ALT proteins in K27(DE3) E. coli .................................................................124

4.3.3 Extraction of fatty acids produced by ALT-expressing E. coli ..................125

4.3.4 Identification and quantification of fatty acids and methylketones produced by ALT-expressing E. coli via GC-FID and GC-MS .......................125

4.3.5 Statistical analysis of GC-FID data .................................................................126

4.3.6 Confirmation of ALT-like protein expression via Tris-Tricine SDS-PAGE and immunoblotting ..........................................................127
4.3.7 Structural modelling of ALT proteins ................................................................. 128

4.3.8 In silico docking of E. coli acyl-ACPs with AtALT4 ........................................ 130

4.4 – Results .................................................................................................................. 131

4.4.1 Two short amino acid motifs are consistently found in ALTs that have preference for β-ketoacyl-ACP and 3-hydroxyacyl-ACP substrates .................................. 131

4.4.2 Disruption of the 31-CQH[G/C]RH-36 motif on the central alpha-helix of the ALT thioesterase domain significantly alters chain-length and oxidation state preference ........................................................................................................ 133

4.4.3 Exchanging residues 31-36 and 108-111 of the thioesterase domain between ALT enzymes that do not possess the 31-CQH[G/C]RH-36 and 108-KXXA-111 motifs also results in changes to chain-length and oxidation state specificity .................................................................................................................. 140

4.4.4 Residues 78-93 of the ALT hot-dog fold domain influence substrate chain-length specificity ......................................................................................................................... 145

4.4.5 The 108-KXXA-111 motif and compatible sequence at aa 78-93 of the ALT hot-dog fold domain are required to accommodate the 31-CQH[G/C]RH motif .................................................................................................................. 146

4.4.6 Residues 94-96 and 64-67 of the ALT catalytic domain also contribute to oxidation state specificity ...................................................................................................................... 148

4.4.7 Additional AtALT3/4 chimeras reveal more information regarding the influence of identified regions of the ALT catalytic domain on substrate specificity ........................................................................................................... 151

4.4.8 Computational modelling of ALT variants and in silico docking with E. coli ACPs illustrates how residues that influence substrate specificity contribute to acyl binding cavity structure and interactions with ACP partners ................................................................................................................................. 152

4.5 – Discussion ............................................................................................................. 157

4.5.1 Chimeric enzymes inform our understanding of ALT substrate specificity ................................................................................................................................. 157
4.5.2 Knowledge gained from the study of ALT-like thioesterases is applicable to other bacterial single hot-dog fold thioesterases that are widely used in metabolic engineering strategies

4.5.3 ALT variants have unique potential as metabolic engineering tools in plant and microbial hosts

4.5.4 Decoding individual-residue contributions to ALT substrate specificity will allow for rational design of recombinant ALTs for distinct biotechnological applications

4.6 – Conclusions

Chapter 5: Summary of findings and future directions

5.1 – Summary of research findings

5.1.1 ALTs are a highly diverse family of unusual plant acyl-ACP thioesterases that likely occupy multiple, highly specialized roles in plant defence

5.1.2 ALT enzymes maintain their catalytic diversity in planta, and can be used to increase medium-chain fatty acid content in plant tissues

5.1.3 Three short interacting amino acid motifs within the ALT hot dog-fold domain are core determinants of thioesterase chain-length and oxidation state specificity

5.1.4 Significance of findings and future research directions

5.2 – Improving the specificity and productivity of ALT enzymes for biotechnological applications

5.2.1 Developing screening techniques to isolate ALT variants with improved activity towards desirable substrates

5.2.2 Examining protein-protein interactions (PPIs) involving ALTs to optimize the production of valuable compounds

i. ALTs with acyl carrier proteins

ii. ALTs with one another, and other single hot-dog fold proteins
iii. ALTs with other potential partner enzymes......................178

5.2.3 Using ALT enzymes to engineer transgenic plants with novel seed oil composition.................................................................179

5.3 – Understanding the biological roles of ALT enzymes.......................181

5.3.1 Tracing how the products of ALTs are metabolized in planta through lipidomics, reverse genetics, and protein-protein interaction mapping .... 181

5.3.2 Chemical profiling of volatile compounds emitted from plants deficient in ALT activity..............................................................183

5.3.3 Exploring the potential roles of ALT genes in plant defence and other plant-environment interactions.......................................................186

   i. Investigating the association of root-expressed ALTs from the model legume Medicago truncatula with pathogen defence and arbuscular mycorrhizal symbiosis..............................................187

   ii. Examining the effects of ALT activity on root microbiome composition ..............................................................189

   iii. Determining whether a loss of ALT gene function influences plant susceptibility to insect herbivory .........................190

Concluding Remarks ..................................................................................................................192

Appendices ..................................................................................................................................195

References ....................................................................................................................................219
List of Figures

Figure 1.1 – Structure of acyl-CoA and acyl-ACP thioesters ................................................................. 5
Figure 1.2 – Framework of acyl lipid metabolism in plant cells ............................................................... 8
Figure 1.3 – ALT-type thioesterases are structurally and functionally distinct from Fat-type thioesterases .......................................................................................................................... 20
Figure 1.4 – Phylogenetic tree of ALT, FatA, and FatB thioesterases from various plant species ............................................................................................................................... 26

Figure 2.1 – Phylogenetic tree of 102 ALTs from 71 plant species ............................................................... 46
Figure 2.2 – GC-FID traces of lipids secreted into the media of K27(DE3) E. coli expressing M. truncatula ALT2 or A. thaliana ALT3, as compared to bacteria harbouring an empty pET-28a vector .......................................................... 50
Figure 2.3 – Identification of novel 3-hydroxy fatty acids in lipid extracts from K27(DE3) E. coli expressing ALT enzymes ........................................................................................................ 52
Figure 2.4 – Quantification of fully reduced and 3-hydroxy fatty acids secreted into the media of K27(DE3) E. coli expressing ALT enzymes ........................................................................... 58
Figure 2.5 – Quantification of methylketones secreted into the media of K27(DE3) E. coli expressing ALT enzymes .................................................................................................................. 60
Figure 2.6 – Immunoblots of crude cell lysates of K27(DE3) E. coli expressing ALT enzymes .. 62
Figure 2.7 – Unrooted structure-based phylogenetic tree and example homology models of ALT proteins ....................................................................................................................................... 65

Figure 3.1 – Box-and-whisker plots depicting medium-chain fatty acid content in seed TAG of Arabidopsis expressing AtALT1-4 in seeds of a Col-0 wild-type background, versus an aae15 aae16 double mutant background ................................................................. 92
Figure 3.2 – Seed triacylglycerol composition from aae15 aae16 Arabidopsis overexpressing AtALT1-4 under control of the seed-specific soybean glycinin promoter .................................................. 94
Figure 3.3 – Representative GC-FID chromatograms of transmethylated seed triacylglycerol from aae15 aae 16 Arabidopsis plants overexpressing AtALT1-4 under control of the seed-specific soybean glycinin promoter ........................................................................... 96
Figure 3.4 – AtALT4-expressing Arabidopsis and Camelina sativa seeds exhibit morphological deformities as compared to untransformed seeds ................................................................. 99
Figure 3.5 – Fatty acid composition of seed triacylglycerol from Camelina sativa (L.) Crantz cultivar ‘Suneson’ overexpressing AtALT1 or AtALT4 ................................................................. 101
Figure 3.6 – Phenotypes of *N. benthamiana* leaves transiently expressing the FAT-type thioesterase *Cp*FatB2 or the ALT-like thioesterases AtALT1-4 and ShMKS2, either alone or in combination with the ShMKS1 decarboxylase...104

Figure 3.7 – Quantification of medium-chain fatty acids and methylketones detected in *N. benthamiana* leaves transiently expressing AtALT1-4, ShMKS2, or *Cp*FatB2 either alone or in combination with the ShMKS1 decarboxylase...106

Figure 4.1 – Two amino acid motifs occur in ALTs that prefer β-keto and 3-hydroxy acyl-ACP substrates. .................................134

Figure 4.2 – Predicted catalytic and acyl-binding cavity residues in modelled ALTs align with those of *E. coli* YbgC and *Umbellularia californica* FatB..................................135

Figure 4.3 – GC-FID chromatograms of secreted lipids from K27(DE3) *E. coli* strains expressing wild-type or chimeric ALTs...............................................................139

Figure 4.4 – Sequence fragments exchanged between ALT pairs to create chimeric thioesterases and their contributions to substrate specificity .................................................142

Figure 4.5 – Structure, substrate specificity, and productivity of wild-type and mutant ALTs in *E. coli*.................................................................................................143

Figure 4.6 – Recombinant ALT-like thioesterases are expressed successfully in K27 (DE3) *E. coli*......................................................................................................144

Figure 4.7 – GC-FID chromatograms of secreted lipids from K27(DE3) *E. coli* strains expressing AtALT4 variants, or AtALT3.................................................................150

Figure 4.8 – Predicted acyl binding cavity structures of modelled wild-type and mutant ALTs..153

Figure 4.9 – Residues within regions that influence ALT substrate specificity may contribute to interactions with ACPs...............................................................156

Figure 5.1 – Measuring the productivity of ALT enzymes in *E. coli* with a Nile Red fluorometric screen..............................................................174

List of Tables

*Table 2.1* – Relative expression levels of ALT enzymes with respect to total lipid production in K27(DE3) *E. coli*............................................................63

*Table 3.1* – Summary of the products of AtALT1-4 when expressed in *E. coli, Arabidopsis* seeds and *Nicotiana benthamiana* leaves..................................91
List of Appendices

**Appendix 1** – Protein sequence alignment of ALT-type thioesterases characterized in this thesis and other literature to date .................................................................195

**Appendix 2** – PCR primers used to assemble genetic constructs described in this thesis……197

*Table S2.A* – PCR primers used to assemble genetic constructs for overexpression of ALT-like thioesterases in Arabidopsis thaliana seeds and Nicotiana benthamiana leaves.................................................................197

*Table S2.B* – PCR primers used to assemble genetic constructs encoding chimeric ALT enzymes for heterologous expression in *E. coli* bacteria..............197

**Appendix 3** – Insert sequences of genetic constructs for heterologous expression of ALT-type thioesterases in *E. coli*.................................................................199

**Appendix 4** – Electron impact (EI) mass spectra used to identify the products of ALT-type thioesterases.................................................................207

*Figure S4.A* – EI mass spectra of fatty acids, 3-hydroxy fatty acids, and methylketones identified in spent media from K27(DE3) *E. coli* cultures expressing ALT-type thioesterases.................................207

*Figure S4.B* – EI mass spectra of fatty acid methyl esters identified in Arabidopsis thaliana and Camelina sativa seed oil samples, and of medium-chain fatty acid methyl esters and methylketones in lipid extracts from *N. benthamiana* leaves overexpressing ALT-type thioesterases.........213
List of Abbreviations

3-OH FA – 3-hydroxy fatty acid
AAE – acyl-activating enzyme
ACP – acyl carrier protein
ACOT – acyl-CoA thioesterase
ALT – acyl-lipid thioesterase
BSA – bovine serum albumin
CoA – coenzyme A
DGAT – diacylglycerol acyltransferase
EI – electron impact
ER – endoplasmic reticulum
FA – fatty acid
FAME – fatty acid methyl ester
GC-FID – gas chromatography-flame ionization detection
GC-MS – gas chromatography-mass spectrometry
LACS – long-chain acyl-CoA synthetase
LPAAT – lysophosphatidic acid acyltransferase
MCFA – medium-chain fatty acid
MK – methylketone
MKS – methylketone synthase
PCR – polymerase chain reaction
PPI – protein-protein interaction
SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TAG – triacylglycerol
TE – thioesterase
TMS – trimethylsilyl group

* Please note that shorthand nomenclature is also used to refer to fatty acid, 3-hydroxy fatty acid, and methylketone species. In this shorthand, fatty acids are referred to by their number of carbon atoms, followed by the number of double-bonds they contain after a colon. For example, “16:0”
“FA” or “C16:0” refers to a 16-carbon fatty acid with no double bonds, while “14:1 FA” or “C14:1” refers to a 14-carbon fatty acid containing one double bond.

The position of the double-bond(s) is sometimes specified as shown in the following examples:

**C20:1 d11** – refers to a 20-carbon fatty acid with one double bond between carbon numbers 11 and 12, counting from the carboxyl (α) to the methyl (ω) end.

**C18:3 d9,12,15** – refers to an 18-carbon fatty acid with three double bonds between carbon numbers 9-10, 12-13, and 15-16, respectively, counting from the carboxyl (α) to the methyl (ω) end.
Chapter 1: Fatty acyl thioesterases in plant lipid metabolism: diverse functions and biotechnological applications

1.1 - INTRODUCTION

Plants use fatty acids to synthesize acyl lipids for many different cellular, physiological, and defensive roles. Acyl lipids are the major components of cellular membranes, seed oils, and surface lipids (reviewed in Pollard et al., 2008; Dobritsa et al., 2009). Various fatty acid-derived metabolites are also used by plants for signaling and defence (reviewed in Lim et al., 2017; Hou et al., 2016; Kachroo & Kachroo, 2009). Since plants do not acquire lipids from their environment, nor transport acyl lipids between cells or tissues, every plant cell must synthesize and metabolize its own acyl lipids. This chapter will begin with an overview of plant acyl lipid metabolism, from de novo fatty acid synthesis to acyl lipid breakdown. Production of acyl lipids occurs primarily at two sites, in the plastid (e.g. chloroplast) or at the endoplasmic reticulum (ER), while breakdown occurs mainly in the peroxisome.

Before they can be used in metabolic processes, fatty acids must first be activated by a thioester linkage to either acyl carrier protein (ACP) or coenzyme A (CoA) (reviewed in Ohlrogge & Browse, 1995). Acyl synthetases and acyl thioesterases are critical enzymes in plant lipid metabolism that are responsible for activating and deactivating fatty acids by adding or removing these cofactors, respectively. In the 1970s-80s, biochemical characterization of fatty acyl synthetases and thioesterases took place in various plant species, including spinach, avocado, safflower, and pea (Andrews & Keegstra, 1983; Block et al., 1983; Joyard & Stumpf, 1980; McKeon & Stumpf, 1982; Ohlrogge et al., 1978; Roughan & Slack, 1977). Since the adoption of Arabidopsis thaliana as a genetic model for plants in the 1980s, most research on
fatty acyl synthetases has since been conducted in this species (Somerville & Koornneef, 2002). Specialized fatty acyl thioesterases, however, continue to be characterized in non-model plants, due to their diverse substrate specificities across plant taxa. Given their ability to produce a variety of fatty acid products in a tissue-specific manner, acyl thioesterases are responsible for much of the fatty acid chain-length diversity observed in particular plant tissue types. The major focus of this chapter will be on the structure and biochemistry of ubiquitous and specialized plant fatty acyl thioesterases, and the many roles that these enzymes occupy in the larger context of plant acyl lipid metabolism. While some specialized acyl thioesterases have known roles in seed oil biosynthesis or plant-environment interactions, most have yet to be characterized both in terms of their substrate preferences and their biological functions.

A portion of this chapter will also be devoted to the biotechnological applications of plant acyl thioesterases, since these enzymes are often used to engineer the acyl lipid profile of oilseed plants and microbes. Fatty acids and their oleochemical derivatives are used in the manufacture of a wide range of commercial products, including fuels, flavourings, cosmetics, pharmaceuticals, and plastic polymers, among others, and the market demand for these compounds is projected to reach $50 billion by the year 2030 (reviewed in Sarria et al., 2017; McWilliams, 2017). An ongoing research current in biotechnology is the sustainable, petroleum-free production of medium-chain (6-14 carbon) fatty acids and their derivatives for industrial use, and many studies describe the manipulation of acyl thioesterase expression to increase medium chain-length content in microbes or plant seed oils (Voelker et al., 1996; Dehesh et al., 1996; Kim et al., 2015; Tjellström et al., 2013; Lennen & Pfleger, 2012; Yan et al., 2022; Jindra et al., 2023). Specialized plant acyl thioesterases, including those characterized in this work, show great promise in the production of industrially valuable
oleochemicals due to the natural variation in their substrate specificities, and advancements made to date in this area will be presented.

1.2 - SUMMARY OF PLANT ACYL LIPID METABOLISM

1.2.1 - Fatty acids are metabolically active as acyl-ACP or acyl-CoA thioesters

Free fatty acids have low solubility in water, and are not very reactive due to the poor leaving group ability of the carboxyl -OH. Within the cellular environment, these properties would prevent free fatty acids from interacting readily with the active sites of many enzymes involved in their metabolism, impede their transport, and render essential biochemical reactions, such as the esterification of fatty acids in the formation of storage and membrane lipids, energetically unfavourable. To overcome these obstacles, free fatty acids are metabolically activated through the formation of thioesters with either Coenzyme A (CoA) or acyl carrier protein (ACP) (Fig. 1). Thioesterification of fatty acids to a hydrophilic CoA or ACP activating group allows them to participate in enzyme-catalyzed reactions by increasing their solubility, and making these reactions thermodynamically favourable. Thioester bonds are high-energy bonds; formation of a fatty acyl-CoA thioester, for instance, requires the equivalent of two ATP units. Hydrolysis of the thioester bond can therefore provide the energy required to drive forward reactions that would otherwise be energetically disfavoured (Mogelson et al., 1984).

CoA is a metabolite found in all organisms and is used in many biochemical reactions involving the activation or transfer of acyl groups. ACP is also found in all organisms, although it does not always exist as a separate protein as it does in plants. In animal and fungal cells, ACP is a subdomain of a larger, multifunctional protein (Jenni et al., 2007; Tsukamoto et al., 1983), while in plants, ACP is encoded by multigene families (Bonaventure & Ohlrogge, 2002).
For example, the *Arabidopsis thaliana* genome encodes eight ACP isoforms, five of which are predicted to be localized to plastids, and three to mitochondria (Fu *et al.*, 2020; Mekhedov *et al.*, 2000). The expression patterns of individual ACP isoforms also vary in different plant tissues (Bonaventure & Ohlrogge, 2002; Ohlrogge & Kuo, 1985). The *Arabidopsis* acyl carrier protein 1 (ACP1) isoform has recently been characterized as a negative regulator of pattern-triggered immunity to microbial pathogens (Zhao *et al.*, 2022), but it is not yet known what distinct roles other ACP isoforms might serve in *Arabidopsis* or other plants.

In plants, ACP is the activating group for acyl chains undergoing *de novo* fatty acid synthesis in plastids or in mitochondria, whereas CoA is the activating group in metabolic pathways occurring outside these compartments (Gueguen *et al.*, 2000; Ohlrogge *et al.*, 1979). The key structural feature of both ACP and CoA is their 4-phosphopantethine prosthetic group (Fig. 1.1). In CoA, this group is joined to an adenine nucleotide, while in ACP, the 4-phosphopantetheine cofactor is attached to a serine residue of the protein within a conserved Asp-Ser-Leu-Asp motif (Majerus *et al.*, 1965; Fu *et al.*, 2020). A free thiol group at the end of this cofactor forms a thioester bond with a free fatty acid. The long 4-phosphopantetheine arm acts as a flexible tether, allowing the acyl group to interact with enzyme active sites while remaining covalently bound to either CoA or ACP (Fig. 1). This is especially important for cyclic pathways like *de novo* fatty acid synthesis or oxidative degradation of fatty acids, where each protein in a complex acts on a given substrate molecule repeatedly (Kindl, 1993; Slabas & Fawcett, 1992). Metabolic activation or deactivation of fatty acids by addition or removal of a thioester group is a core checkpoint in the overall scheme of plant lipid metabolism (Fig. 2), which is discussed in the following sections.
Figure 1.1. Fatty acids are metabolically active as thioesters. (a) Structures of Coenzyme A (CoA) and acyl carrier protein (ACP), both of which carry a 4-phosphopantetheine group. The 4-phosphopantetheine group is attached either to an adenine nucleotide in CoA, or to a conserved serine residue in ACP. (b) Fatty acids are activated to acyl thioesters by acyl synthetase enzymes, and deactivated to free fatty acids by thioesterase enzymes. R = CoA or ACP. (c) NMR-determined backbone structure of a chloroplastic decanoyl-ACP thioester (10:0–ACP) from Spinacea oleracea (PDB Structure ID: 2FVF). The 4-phosphopantetheine cofactor is carried by Ser38. The fatty acyl chain is normally buried between protein’s α-helices (“docked” conformation), and is rotated about the 4-phosphopantetheine tether (“open” conformation) to interact with the active site of a suitable enzyme.

1.2.2 – De novo fatty acid biosynthesis in the plant cell

In all plant cells, acyl lipid metabolism generally begins with de novo synthesis of fatty acids in the plastid. This is a cyclic pathway, in which a fatty acyl chain linked to the 4-phosphopantetheine cofactor of ACP by a thioester bond is repeatedly elongated by iterative additions of two-carbon groups (reviewed in Ohlrogge & Jaworski, 1997. The first committed
step in fatty acid synthesis is the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (ACC). The malonyl group is then linked to ACP by a transacylation reaction. A sequence of four enzymatic reactions follows: condensation of acetyl-CoA with malonyl-ACP to form β-ketoacyl-ACP, an NADPH-dependent reduction to 3-hydroxyacyl-ACP, dehydration to trans-β-enoylacyl-ACP, and finally, a second NADPH-dependent reduction to a fully reduced fatty acyl-ACP, which has has been extended by two carbons (Shimakata & Stumpf, 1982a; reviewed in Ohlrogge & Jaworski, 1997) (Fig. 2). The four-reaction cycle then repeats, beginning with the condensation of the growing acyl-ACP thioester with malonyl-CoA, and the acyl chain is lengthened by two carbons with each full turn of the cycle (Fig. 2). This typically continues until the acyl chain reaches 16 or 18 carbons.

The reactions of the de novo fatty acid biosynthesis cycle in plants are catalyzed by individual enzymes that associate to form a Type II fatty acid synthase (FAS) multiprotein complex resembling that of Escherichia coli and other bacteria (reviewed in White et al., 2005 and Ohlrogge & Jaworski, 1997. The enzymes that carry out reduction and dehydration reactions participate in every turn of the cycle. However, like in E. coli, β-ketoacyl-ACP synthesis is catalyzed by multiple isoforms of 3-ketoacyl-acyl carrier protein synthase (KAS) that have specificity towards different acyl chain lengths. The initial condensation of acetyl-CoA with malonyl-ACP is catalyzed by the KASIII isoform, while the KASI isoform participates in subsequent turns of the fatty acid synthesis cycle until the acyl chain reaches 16 carbons in length (Clough et al., 1992; Shimakata & Stumpf, 1983). At this point, a portion of fatty acids are released from the cycle, while others are elongated from 16:0-ACP to 18:0-ACP in a final turn of the cycle that involves KASII (Shimakata & Stumpf, 1982b). Nearly all 16-carbon fatty acids remain saturated after synthesis, while most 18-carbon fatty acids become
monounsaturated by the action of a soluble stearoyl-ACP desaturase that introduces a \textit{cis} double-bond between C9-10 (Jacobson \textit{et al.}, 1974; McKeon \& Stumpf, 1981; Kachroo \textit{et al.}, 2007). The major end products of \textit{de novo} fatty acid synthesis in the plastid are therefore 16:0-ACP and 18:1-ACP, with 18:0-ACP as a minor product (Browse \& Sommerville, 1991). The ACP group is then removed from the acyl chain by one of two methods: transacylation of the acyl chain onto a glycerol-3-phosphate backbone catalyzed by an acyltransferase, or hydrolysis to a free fatty acid by an acyl-ACP thioesterase (reviewed in Benning, 2009). The relative activities of plastidial acyltransferases and acyl-ACP thioesterases are key for regulating fatty acid flux through downstream pathways, as this determines the subcellular location at which the acyl chain will undergo further metabolism (El Tahchy \textit{et al.}, 2017; Löhden \& Frentzen, 1988).

Mitochondria conduct a limited amount of fatty acid biosynthesis through a similar reaction cycle that primarily yields 8:0-ACP, 16:0-ACP, and 18:0-ACP (Wada \textit{et al.}, 1997; Gueguen \textit{et al.}, 2000). The main purpose of mitochondrial fatty acid biosynthesis in plants seems to be to generate the octanoyl-ACP precursor of lipoic acid, a sulfolipid cofactor used by several mitochondrial enzymes (Wada \textit{et al.}, 1997; Guan \textit{et al.}, 2017). A proportion of mitochondrial long-chain (C16-C18) fatty acids are also incorporated into mitochondrial membranes (reviewed in Michaud \textit{et al.}, 2017). Fatty acids generated in the mitochondria are thought to be directed to downstream pathways by an acyltransferase; no fatty acyl thioesterase activity has been observed in plant mitochondria (reviewed in Li-Beisson \textit{et al.}, 2013).
Figure 1.2. Framework of acyl lipid metabolism in plant cells. Although pathway steps are depicted within organelle lumens to allow for full pathway visualization, the enzymatic reactions depicted in this figure generally occur at the organelle membranes. Fatty acid synthesis occurs mainly in the plastid, where acyl-ACP synthetase and acyl-ACP thioesterase enzyme families are involved in production of
medium- and long-chain fatty acyl groups. Some of these groups are further metabolized in the plastid, generating galactolipids and sulfolipids in all cells, or oxylipins in specialized cell types. Free fatty acids exported from the plastid are rapidly activated to CoA thioesters in the cytosol by acyl-CoA synthetases. Long-chain acyl-CoA groups at the ER may be further elongated to very long-chain acyl-CoA groups. Collectively, these products can enter primary metabolic pathways to form membrane lipids (phospholipids and sphingolipids) or storage lipids (triacylglycerols). Alternatively, products may enter specialized metabolic pathways to form cutin monomers, suberin monomers, and cuticular waxes for incorporation into surface lipid barriers, or specialized triacylglycerols for incorporation into unusual seed oils. Catabolism of long-chain fatty acyl groups occurs primarily at the peroxisome, where chains are broken down by successive removal of two-carbon units by ß-oxidation.

1.2.3 – Further metabolism of fatty acids

A portion of newly synthesized fatty acyl chains are directed to intraplastidial lipid metabolism by plastid-localized acyltransferase enzymes (Benning, 2009; Hölzl & Dörmann, 2019; Li-Beisson et al., 2013). Meanwhile, acyl chains that are released from ACP by acyl-ACP thioesterases are unavailable to plastid acyltransferases, so these instead exit the plastid, and are eventually metabolized further at the endoplasmic reticulum (ER) (Fig. 2) (Dörmann et al., 1995; Jones et al., 1995). In Arabidopsis, the transmembrane proteins fatty acyl export 1-4 (FAX1-4) mediate fatty acid export from the plastid, but the precise mechanism by which this transport occurs is unknown (Li et al., 2015; Tian et al., 2019; Bugaeva et al., 2023). Traditional biological models hold that that free fatty acids are rapidly re-activated to acyl-CoA thioesters at the plastid outer envelope by an acyl-CoA synthetase, and are then mostly transported to the ER.

Once at the ER membrane, a portion of long-chain fatty acyl-CoAs are elongated into very-long-chain (C20+) acyl-CoAs through the fatty acid elongation cycle. This pathway resembles de novo fatty acid biosynthesis in that the acyl chain is sequentially elongated by two carbons at a time through the same four-reaction series of condensation, reduction, dehydration, and a
final reduction (reviewed in Li-Beisson et al., 2013). However, acyl-CoA thioesters are used in place of acyl-ACPs, and little is known about the termination of this cycle. The acyl chains of fatty acyl-CoAs are incorporated into ubiquitous lipid types such as polar membrane phospholipids and sphingolipids, and neutral triacylglycerols, which primarily serve as energy stores. In certain cell types, these acyl chains are also incorporated into extracellularly deposited lipid-based polymers, including cutin in the epidermis of aerial plant parts, suberin in root endodermis, exodermis and periderm, and sporopollenin on the outer layer (exine) of pollen grains (Pollard et al., 2008; Bernards, 2002; Kim & Douglas, 2013). These cross-linked hydrophobic lipid barriers protect against uncontrolled water loss, wounding, herbivory, pathogen infection, and other environmental stresses.

Triacylglycerols are compact molecules for energy and carbon storage, since their highly hydrophobic nature allows them to pack tightly together. Plants therefore mainly store their reserve lipids as triacylglycerols in ER-derived lipid droplets. These are most abundant in seeds, though some triacylglycerol stores are also maintained in pollen, fruit pericarp, and leaves (reviewed in Li-Beisson et al., 2013). In oilseed plants such as Brassica and Cuphea spp., triacylglycerols form the bulk of the seed endosperm and are the primary energy source for emerging seedlings before they can rely on photosynthesis. The beginning steps of ER-localized triacylglycerol biosynthesis are identical to those of membrane phospholipid biosynthesis. Acyl chains are transferred from CoA thioesters to the sn-1 and sn-2 positions of glycerol-3-phosphate, and usually desaturated at various positions. However, instead of the addition of a polar head group to the sn-3 position, as in membrane phospholipid biosynthesis, the phosphate group is removed, and a third acyl chain is transacylated to the sn-3 position in its place. Typically, acyl chain-lengths in plant triacylglycerols range from 16-24 carbons, however,
certain plants incorporate fatty acids with longer or shorter chain-lengths, and/or with unusual functional groups in their seed oils (Graham, 1999; Bafor et al., 1991; Li et al., 2018).

When triacylglycerols are broken down in plant cells for energy, acyl chains are hydrolyzed from the glycerol backbone by lipases, releasing them as free fatty acids (Eastmond, 2006). These are then transported to the peroxisomes, which are the main site of fatty acid breakdown, or β-oxidation, in plant cells (Kindl, 1993). Here, the free fatty acids are re-activated to CoA thioesters by long-chain acyl-CoA synthetases, and broken down by successive removal of two-carbon units in a cycle of four enzymatic reactions that is the reverse of the fatty acid synthesis pathway. This cycle is repeated until the acyl chain is fully oxidized to acetyl-CoA, and the acetyl-CoA generated from this process enters the citric acid cycle.

1.3 – OVERVIEW OF FATTY ACYL THIOESTERASES FROM PLANTS

1.3.1 - Plant fatty acyl thioesterases possess a “hot dog-fold” domain structure

All fatty acyl thioesterase enzymes from plants that have been identified to date belong to the “hot dog-fold” protein superfamily, which is widespread across bacteria, archaea, and eukaryotes (Dillon & Bateman, 2004). These proteins are named for a characteristic structural domain consisting of a central hydrophobic five-turn α-helical “hot dog” surrounded by a 5 – 7 strand antiparallel β-sheet “bun” (Dillon & Bateman, 2004; Leesong et al., 1996) (Fig. 1.3). Proteins with hot dog-fold domains possess thioesterase or dehydratase activity towards a wide array of substrates, and are involved in many different biological processes including hydrolysis of acyl thioesters in lipid metabolism, breakdown of phenylacetic acid to form citric acid cycle intermediates, degradation of environmental toxins, and dehydration of intermediates in the biosynthesis of bacterial virulence factors (Dillon & Bateman, 2004; Caswell et al., 2022).
dog-fold proteins also include one non-catalytic subfamily, the FapR transcription factors, which regulate Type II fatty acid biosynthesis and phospholipid biosynthesis in gram-positive bacteria (Dillon & Bateman, 2004). While the fold structure itself is very well-conserved among even distantly related hot dog-fold proteins, an unusually low degree of sequence conservation at the amino acid level likely accounts for their highly diverse functions and substrate specificities. The lack of a conserved set of catalytic residues within this protein family creates challenges in structural modelling and function prediction of uncharacterized hot dog-fold proteins (Zhuang et al., 2008).

Enzymes with a hot-dog fold domain structure probably cannot function as monomers, and are known to form homo-and hetero-dimers and higher order oligomers as active units in vivo (Dillon & Bateman, 2004; Kunishima et al., 2005; Willis et al., 2008; Pidugu et al., 2009). Monomers can associate through interactions between their β-sheets or between their α-helices, and often assort in hexameric “trimer-of-dimers” or tetrameric “dimer-of-dimers” assemblies (Willis et al., 2008; Pidugu et al., 2009; Gonzalez et al., 2012). It is common for two functional hot-dog fold domains to be encoded within a single polypeptide (Dillon & Bateman, 2004).

1.3.2 – Acyl-CoA thioesterases from plants currently elude functional characterization

In animals and fungi, thioesterases that catalyze the hydrolysis of acyl-CoA thioesters are known to function in fatty acid elongation, and as auxiliary enzymes in peroxisomal β-oxidation (reviewed in Hunt et al., 2012). These thioesterases are encoded by members of the ACOT (acyl-CoA thioesterase) multigene family, and can be classified into two major types: those belonging to the α/β hydrolase fold protein superfamily, which occur only in prokaryotes and animals, and those belonging to the hot dog-fold superfamily, which are found in all
eukaryotes (Hunt et al., 2006). The Arabidopsis genome encodes two peroxisome-targeted homologues of mammalian peroxisomal ACOT8, a hot dog-fold acyl-CoA thioesterase that generates saturated, unsaturated, branched-chain, and dicarboxylic fatty acids for export to the cytosol or mitochondria (Hunt et al., 2012). One of these homologues, ACH2, hydrolyzes 12–20 carbon saturated and unsaturated acyl-CoA thioesters in vitro (Tilton et al., 2004). Despite its activity, ACH2 is not thought to function in peroxisomal fatty acid oxidation based on its enriched expression in mature tissues, rather than in germinating seedlings where fatty acid turnover by β-oxidation occurs at the most rapid rate (Tilton et al., 2004). To date, no acyl-CoA thioesterases have been proven to be directly involved in fatty acid β-oxidation in plants (Hunt et al., 2012). Despite decades of research in plant lipid metabolism, the physiological functions of acyl-CoA thioesterase genes in plants are yet to be determined.

1.3.3 – FatA- and FatB-type acyl-ACP thioesterases release acyl chains from the de novo fatty acid biosynthesis cycle

Considerably more is known about thioesterase enzymes in plants that catalyze hydrolysis of acyl-ACP thioesters. The most widely recognized function of acyl-ACP thioesterases is in partitioning newly synthesized acyl chains between the plastid and the ER for further metabolism. This function is performed by plastid-localized acyl-ACP thioesterases belonging to two closely related subfamilies: the FatA and FatB-type thioesterases. These are part of a large clade within the hot dog-fold superfamily consisting of proteins with two functional hot dog-fold domains that are formed by a single polypeptide (Mayer & Shanklin, 2005; Cantu et al., 2010; Caswell et al., 2022). Each plays a distinct role in catalyzing the hydrolysis of acyl-ACP thioesters: residues of the N-terminal hot-dog fold domain contain a hydrophobic substrate binding pocket that accommodates the fatty acyl chain tethered to ACP, while the catalytic
residues are located on the C-terminal hot dog-fold domain (Feng et al., 2017; Jing et al., 2018). Positively charged surface patches that mediate electrostatic interactions with ACP and facilitate extrusion of the tethered fatty acyl chain from the ACP core into the acyl binding pocket are present in both domains (Feng et al., 2017; Jing et al., 2018; Ziesack et al., 2018).

Isoforms of FatA and FatB are ubiquitously expressed, and acyl-ACP thioesterase activity is found in most, if not all, plant tissues (Bonaventure et al., 2003). The main differences between the FatA and FatB enzyme families are their substrate specificities in terms of fatty acyl chain-length and saturation level, and their precise localization within the plastid. Most FatA-type thioesterases, which are soluble within the plastid stroma, act almost exclusively on 18:1-ACP (Salas & Ohlrogge, 2002; Sánchez-García et al., 2010; Moreno-Pérez et al., 2011), although there is at least one example of a FatA enzyme with appreciable activity toward 18:0-ACP in vitro (Hawkins & Kridl, 1998). FatB enzymes, which are bound to the inner plastid membrane, show preference for substrates with saturated acyl chains, particularly 16:0-ACP, but seem to be less specific in terms of chain length (Jones et al., 1995; Aznar-Moreno et al., 2018). While ubiquitously expressed FatB enzymes show preference toward 16:0 substrates, specialized FatB thioesterases with activity toward medium-chain (8:0–14:0) acyl substrates are expressed in the developing seeds of some oilseed species, including *Umbellularia californica* (California bay) and *Cuphea* spp. (Dehesh et al., 1996; Dörmann et al., 2000; Kim et al., 2015; Voelker et al., 1992).

As described in Section 1.2, plants express multiple isoforms of ACP, and the substrate specificities of FatA- and FatB-type thioesterases are influenced by which ACP isoform delivers the fatty acyl chain. Spinach FatA is known to display strong preference for oleoyl-ACP (18:1-ACP) linked to one plastid-expressed ACP isoform over another (Ohlrogge et
Seed-expressed FatA- and FatB-type thioesterases in *Helianthus annuus* (sunflower) also show subtle differences in fatty acyl chain length (C16 vs C18) and saturation level preference, depending on which of three plastidial ACP isoforms they interact with (Aznar-Moreno *et al.*, 2016). Differential expression of ACP isoforms alongside acyl-ACP thioesterases may therefore serve to fine-tune lipid biosynthesis in plants. Recent research on the interactions of acyl-ACP thioesters with enzymes of the *E. coli* fatty acid synthase complex has demonstrated that the ACP relays information to a potential partner enzyme about the length and saturation level of the acyl chain it carries, through a series of conformational changes that occur upon binding to the enzyme (Sztain *et al.*, 2021). These conformational changes thereby allosterically regulate what is termed the “chain-flipping” event, in which the tethered fatty acyl chain, initially buried within the acyl carrier protein core, is translocated to the active site of a suitable enzyme by rotation about the 4-phosphopantetheine arm (Cronan, 2014; Sztain *et al.*, 2021) (Fig. 1). Variation in the substrate specificities of plant acyl-ACP thioesterases when they interact with different ACP isoforms is likely mediated by this mechanism.

1.3.4 – Fat-type acyl-ACP thioesterases in specialized metabolism

In addition to their roles in primary metabolism, acyl-ACP thioesterases are involved in the metabolism of various specialized lipids. One well-established role is the production of 8–14 carbon medium-chain fatty acids for incorporation into seed oils. All plants synthesize triacylglycerols typically composed of 16 to 24-carbon long-chain fatty acyl groups. Certain oilseed species also produce triacylglycerols containing appreciable amounts of medium-chain saturated acyl groups (Davies, 1993; Graham, 1989; reviewed in Dyer *et al.*, 2008). These medium-chain acyl groups are generated by specialized, seed-specific FatB variants that interrupt fatty acid synthesis by hydrolyzing acyl-ACP to produce 8:0–14:0 fatty acids (Davies,
1993; Jones et al., 1995; Voelker & Davies, 1994). For instance, FatB2 from Cuphea hookeriana produces 8:0 and 10:0 fatty acids (Dehesh et al., 1996), FatB1 from Cuphea palustris primarily generates 14:0 fatty acid (Tjellström et al., 2013), and FatB2 from Umbellularia californica generates 12:0 fatty acid in seeds (Eccleston et al., 1996; Eccleston & Ohlrogge, 1998; Voelker & Davies, 1994). Without the ACP thioester, further elongation of medium-chain fatty acids is blocked and products are exported to the ER, where they are incorporated into triacylglycerols. Although many specialized FatB enzymes have been characterized in terms of their enzymatic products, the physiological benefits of medium-chain fatty acids in seed oils remain to be determined. It has been suggested that certain oilseeds accumulate medium-chain fatty acids as these are more easily mobilized, and are more stable at higher temperatures (reviewed in Voelker & Kinney, 2001).

Specialized FatB enzymes share strong sequence and structural similarity to those involved in primary metabolism. They consist of the same characteristic double hot-dog fold domain structure, where the N-terminal hot-dog fold domain houses the hydrophobic acyl binding cavity, and the C-terminal domain contains catalytic residues (Mayer & Shanklin, 2005; Feng et al., 2017; Jing et al., 2018) (Fig. 1.3). Analysis of the Umbellularia californica FatB1 crystal structure, and targeted mutagenesis and homology modelling of the 8:0-specific and 14:0-specific CvFatB1 and CvFatB2 thioesterases from Cuphea viscosissima has revealed which residues contribute to acyl binding cavity structure and depth in medium-chain FatB thioesterases, and thereby dictate chain-length preference (Feng et al., 2018; Jing et al., 2018; Ziesack et al., 2018). The FatB-type enzyme class has been categorized into three groups based on chain-length specificity. Subclass I FatB enzymes, which act on 14:0-16:0 chain-lengths and include variants involved in both primary and specialized lipid metabolism, are thought to
represent the ancestral form of the thioesterase (Jones et al., 1995; Jing et al., 2018). Enzymes in subclass II display broader chain-length specificities ranging from 8:0-16:0, while subclass III is comprised of specialized FatB enzymes with preference for 8:0 acyl-ACP (Jing et al., 2011). A recently characterized FatB isoform expressed in seeds of the deciduous tree Koelruteria paniculata suggests the existence of a potential fourth subclass (Martins-Noguerol, 2020). Although phylogenetic analysis places this thioesterase in the FatB family, it behaves more similarly to FatA enzymes, showing preference for monounsaturated 18:1-ACP and 18:0-ACP over 16:0-ACP when expressed heterologously in E. coli bacteria and in leaves of the model plant Nicotiana benthamiana (Martins-Noguerol et al., 2020).

Plants capable of forming symbiotic interactions with arbuscular mycorrhizal fungi possess a unique, additional Fat-type thioesterase gene, called FatM. This gene encodes a double hot-dog fold thioesterase belonging to the same structural clade as FatB enzymes, that is responsible for mobilizing 16:0 fatty acid from the chloroplast for delivery to the root-colonizing fungus (Bravo et al., 2016, 2017). The Type I fatty acid synthase complex responsible for de novo fatty acid synthesis in most fungi is absent from arbuscular mycorrhizal fungi, which are therefore dependent on fatty acids produced by the host plant to survive (Wewer et al., 2014). In terms of its catalytic function, FatM is redundant to FatB. However, while FatB is constitutively expressed in all plant tissues, FatM is specifically expressed during fungal symbiosis, with transcript levels increasing more than 100-fold in root cells at the time of colonization by fungal arbuscules (Bravo et al., 2017). Plants lacking FatM expression display severe impairments in arbuscule branching in mycorrhizal roots, and FatB expression cannot complement the loss of FatM function (Bravo et al. 2016, 2017; Brands et al., 2018).
1.3.5 – ACYL-LIPID THIOESTERASE (ALT) enzymes are a novel class of acyl-ACP thioesterase that produce medium-chain fatty acids as specialized metabolites

In addition to the Fat-type thioesterases, members of a separate acyl-ACP thioesterase gene family are almost universally present within available plant genomes, and a few of these genes had been characterized prior to the work presented in this thesis. These **ACYL-LIPID THIOESTERASE (ALT)** genes, as they will be referred to henceforward, encode plastid-localized proteins that consist only of a single hot dog-fold domain, and are more closely related to prokaryotic single hot-dog fold thioesterases than plant Fat-type thioesterases (Ben-Israel *et al.*, 2009; Pulsifer *et al.*, 2014; Caswell *et al.*, 2022). While ALT-type thioesterases are capable of interrupting fatty acid synthesis prematurely, akin to specialized FatB enzymes, they appear to have more diverse secondary metabolic functions (Yu *et al.*, 2010; Pulsifer *et al.*, 2014).

The first ALT-encoding gene to be discovered and characterized was METHYLKETONE SYNTHASE 2 (*ShMKS2*), which is expressed within the glandular leaf trichomes of wild tomato (*Solanum habrochaites* subsp. *glabratum*) (Ben-Israel *et al.*, 2009; Yu *et al.*, 2010). While specialized FatB enzymes release fully reduced medium-chain fatty acids from the *de novo* fatty acid synthesis cycle, the thioesterase encoded by *ShMKS2* mainly acts on the 12:0–14:0 β-ketoacyl-ACP intermediates of this cycle, cleaving these thioesters to release medium-chain β-keto fatty acids (Fig. 1.3). The medium-chain β-keto fatty acids produced by *ShMKS2* are then decarboxylated by *ShMKS1*, a companion plastid-localized decarboxylase, to yield 11:0 and 13:0 methylketones (MKs), which accumulate in the trichomes and function as insecticides (Fig. 1.3) (Antonious *et al*. 2003; Fridman *et al.*, 2005; Yu *et al.*, 2010).
A family of four paralogous single hot dog-fold acyl-ACP thioesterases from the model plant *Arabidopsis thaliana* were later characterized (Pulsifer *et al.*, 2014). These plastid-localized proteins, named ACYL-LIPID THIOESTERASE 1-4, (*AtALT1*-4), are orthologues of *S. hirsutum* MKS2, yet they exhibit distinct activity profiles from the wild tomato thioesterase. Heterologous expression of *AtALT1*-4 in *Escherichia coli* bacteria demonstrated that although they share over 75% sequence identity, they each display unique substrate specificity (Pulsifer *et al.*, 2014). When expressed in *E. coli*, *AtALT1* primarily produces 12-14 carbon fatty acids (fully reduced), *AtALT2* generates 8:0 and 10:0 β-keto fatty acids, which are the chemical precursors of 7:0 and 9:0 methylketones, and *AtALT4* generates volatile 6:0 and 8:0 fatty acids (Pulsifer *et al.*, 2014). *AtALT3* was shown to have broader substrate specificity than the other *Arabidopsis* ALTs, but mainly generated the 12-16 carbon β-keto fatty acid precursors of 11-15 carbon methylketones (Pulsifer *et al.*, 2014).
Fig 1.3. ALT-type thioesterases are structurally and functionally distinct from specialized FatB-type thioesterases. (a) Termination of de novo fatty acid biosynthesis by ALT-type and specialized FatB-type thioesterases. ALT-type thioesterases can act on both fatty acyl-ACPs and the β-ketoacyl-ACP intermediates of fatty acid biosynthesis. Medium-chain β-keto fatty acids generated by ALT enzymes can be chemically or enzymatically decarboxylated to methylketones. (b) Three-dimensional ribbon models of ALT4 from Arabidopsis thaliana, YbgC from E. coli (Watanabe et al., 2006), and 12:0-specific FatB2 from Umbellularia californica (Xue & Feng, 2017). The N-terminal plastid targeting peptides of AtALT4 and UcFatB2 have been omitted. ALT-type thioesterases consist of a single hot-dog fold domain, similar to bacterial YbgC acyl-CoA thioesterases, while FatB-type thioesterases comprise two hot-dog fold domains within a single polypeptide. AtALT4 structure was generated with AlphaFold 2.0 (Jumper et al., 2021).

There was little information to explain how the sequences and structures of AtALT1-4 would give rise to their widespread activity profiles in E. coli, since few studies have addressed the mechanisms of substrate specificity in single hot dog-fold thioesterases. These proteins exist in all kingdoms of life and act on substrates with extremely varied aliphatic and aromatic acyl group structures, yet only a small fraction of hot dog-fold thioesterases are well-studied. According to the ThYme (Thioester-active enzYme) database, which catalogues thioesterases according to their structure and phylogenetic origins, ALT enzymes belong to the same clade as bacterial YbgC-like long-, medium-, and short-chain fatty acyl-CoA thioesterases (Cantu et
Although the physiological functions and enzymatic activities of YbgC proteins from several bacteria have now been characterized, little research has been done on the biochemical basis of acyl chain selectivity in these enzymes (Zhuang et al., 2002; Angelini et al., 2008; Caswell et al., 2022). Bacterial YbgC-like and plant ALT-like thioesterases share a conserved amino acid motif containing what is now known to be the YbgC active site (Zhuang et al., 2002; Angelini et al., 2008). Therefore, ALT-type thioesterases are predicted to have a Tyr-Asp-His catalytic triad like their distant YbgC relatives, with the conserved aspartate residue acting as the catalytic nucleophile (Zhuang et al., 2002; Dillon & Bateman, 2004; Caswell et al., 2022) (Fig. 1.3). Mutation of this conserved aspartate to alanine was shown to inactivate AtALT1 in *E. coli*, confirming that it is indeed a key catalytic residue in ALT enzymes (Pulsifer et al., 2014). However, this work did not identify relationships between AtALT1-4 sequence and acyl chain-length and oxidation state specificity.

While the biological functions of *AtALT1*-4 are unknown, their gene expression patterns *in planta* and product profiles in *E. coli* point towards involvement in plant-environment interactions, particularly in defence against pests and pathogens (Pulsifer et al., 2014). Based on their gene expression patterns, *AtALT1* and *AtALT2* have been implicated in the biosynthesis of components of cuticle and suberin, respectively (Pulsifer et al., 2014). The plant cuticle, which is comprised of cutin, a cross-linked lipid polymer, and non-polymerized cuticular waxes, is deposited at the outer epidermal cell walls of the aerial parts of all land plants (reviewed in Bernard and Joubès, 2013). Suberin, which consists of alternating polylphatic and polyaromatic domains derived from fatty acid and phenylpropanoid pathways, is deposited in the cell walls of root endodermal and periderm cells, in the seed coat, and at wound healing sites (reviewed in Bernards, 2002). *AtALT1* is mainly expressed in stem epidermal cells and in
flower petals, and its expression pattern correlates strongly with several genes involved in the deposition of cuticular waxes by epidermal cells (Pulsifer et al., 2014). *AtALT2* is mainly localized to root endodermis and peridermis, which are sites of suberin deposition, and transcriptomic data shows that *AtALT2* is co-expressed with several genes that participate in suberin biosynthesis (Pulsifer et al., 2014).

While the expression patterns of *AtALT1* and *AtALT2* suggest they may participate in biosynthesis of the wax component of plant cuticle and of polyaliphatic suberin, respectively, their demonstrated plastidial localization and activity as medium-chain acyl-ACP thioesterases contradict these roles (Pulsifer et al., 2014). Cuticular waxes and suberin monomers are both derived from long-chain fatty acyl-CoA thioesters, synthesized at the endoplasmic reticulum. It is possible that *AtALT1* and *AtALT2* produce medium-chain compounds that accumulate at sites of cuticle and suberin deposition to provide additional fortification against pest and pathogen attack in these areas. The medium-chain fatty acids generated by *AtALT1* possess antimicrobial properties, and the products of *AtALT2* are the direct chemical precursors of volatile C7:0 and C9:0 methylketones, which display notable inhibitory activity against several fungal phytopathogens such as *Colletotrichum*, *Botrytis*, and *Fusarium* species (Pohl et al., 2011; Huang et al., 2011; Zheng et al., 2013; Neri et al., 2007; Yuan et al., 2012). These methylketones, which also act as attractant, repellent, or alarm signals for various insects, have been detected in volatile emissions from the fruits and flowers of diverse plant species including hops, coffee (*Coffea canephora*), rice, cloves, red clover, and *Arum maculatum* (reviewed in Forney & Markovetz, 1971; Buttery et al., 1984; Prokopy et al., 1998; Lu et al., 2014; Sun et al., 2014; Kite et al., 1995).
In line with its broader substrate specificity, *AtALT3* also has a broad gene expression pattern. The *AtALT3* transcript is present in all aerial plant parts and in roots, with transcriptomic data indicating that it is most abundant in rosette leaves and flowers (Pulsifer *et al.*, 2014; Schmid *et al.*, 2005). Since its localization and activity profile (at least, in *E. coli*) overlap somewhat with other *Arabidopsis* ALT enzymes, *AtALT3* may exhibit some functional redundancy, but it likely has some unique function due to its ability to generate 12-16 carbon β-keto fatty acids (Pulsifer *et al.*, 2014). It is possible that these β-keto fatty acids undergo a similar fate as the products of *ShMKS2*, where they are decarboxylated to 11-15 carbon methylketones that are stored in plant tissues for defence against insect herbivores. However, endogenous decarboxylase activity would be required to efficiently convert the β-keto fatty acid products of *AtALT3* and *AtALT2* to methylketones, and no candidate decarboxylase has been identified in *Arabidopsis*. Close orthologues of *ShMKS1* only occur in members of the Solanaceae that are very closely related to *S. habrochaites* subsp. *glabratum*, and it has been hypothesized that MKS1-like decarboxylase activity may have evolved very recently within the Solanaceae (Yu *et al.*, 2010; Auldridge *et al.*, 2011). Nonetheless, medium-chain methylketones ranging from 7-15 carbons in chain-length have been detected in numerous non-solanaceous plant species that do not possess *MKS1*-like decarboxylases, indicating that these compounds are generated by some other means. In addition to the 7:0-9:0 MK-producing plants mentioned previously, several species also accumulate 11-15 carbon methylketones. Several species in the Rutaceae family (i.e. *Zanthoxylum*, *Ruta*, and *Pilocarpus* spp.), accumulate large quantities of 11-15 carbon methylketones in their leaves, 2-undecanone is major constituent of resin from *Commiphora rostrata*, and *Allium cepa* (field onion) accumulates 2-undecanone and 2-tridecanone (reviewed in Forney & Markovetz, 1971; McDowell *et al.*, 1988; Burdock, 2010;
Craveiro et al., 1979; Romero et al., 2006; Setzer et al. 2005; Antonious, 2013; Kumar et al., 2016). These products could result from the activity of ALTs with specificity for β-ketoacyl-ACP substrates working in tandem with other decarboxylases that do not resemble ShMKS1.

AtALT4 has the most specific expression pattern of the Arabidopsis ALTs, being localized exclusively to floral anthers (Pulsifer et al., 2014). Volatile fatty acids are common components in floral scent mixtures, where they often serve to attract pollinators and beneficial insect bodyguards, and the volatile 6:0 and 8:0 fatty acids produced by AtALT4 may serve this purpose (Knudsen et al., 1993). While Arabidopsis self-pollinates efficiently, visitation of flowers by pollinating insects has been observed in wild populations (Tan et al., 2005; Chen et al., 2003). Short- and medium-chain fatty acids emitted by flowers have also been shown to mediate ethylene sensitivity, and consequently, influence floral development, in species such as carnation and tobacco, and the products of AtALT4 could also share this function (Whitehead & Vasiljevic, 1993).

While the above described studies were focused on ALT-type thioesterases from two plant species, the ALT enzyme family is in fact highly conserved across the plant kingdom. Hundreds of ALT-encoding genes are present in sequenced plant and green microalgalae genomes and have not been lost through genetic drift, indicating that these genes fulfil important biological functions. Unlike FatA- and FatB-type thioesterases, which are grouped according to function in a phylogenetic tree, ALTs are grouped by occurrence in plant species or families, and an individual species can possess one or more paralogous ALT genes (Fig. 1.4). Analysis of publicly available transcriptomic data repositories indicates that ALT genes belonging to many species are up-regulated in response to a variety of biotic and abiotic stressors, further implicating this gene family in defensive processes (Waese et al., 2017; Barrett et al., 2012).
Given the diversity of substrate specificities observed among a small group of paralogous ALT-type thioesterases, ALT enzymes from other plant species could have unique, wide-ranging product profiles. The in-depth functional characterization of ALT enzymes from diverse plant taxa is necessary to understand the biological roles and evolutionary origins of this ubiquitous, yet poorly studied enzyme family. ALT-type thioesterase activity could be responsible for producing fatty acid-derived metabolites with biosynthetic origins that are presently unclear, or for generating compounds in certain plants that were not previously thought to be part of these species’ metabolomes.
Figure 1.4. Phylogenetic tree of FatA-, FatB-, and ALT-type thioesterases from various plant species. This tree was generated with 90 complete ALT sequences, and 10 complete FatA and FatB sequences. Subtrees consisting of multiple ALT proteins from the same plant family were collapsed. Protein sequences were aligned using MUSCLE (Edgar et al., 2004), and the phylogenetic tree was generated using the maximum likelihood model. The tree was rooted using the single hot-dog fold acyl-CoA thioesterase YgbC from E. coli as the outgroup. The Jones-Taylor-Thomton model (Jones et al., 1992) was chosen for amino acid substitutions, and gaps were subject to partial deletion (below 90% coverage). ACM – Ananas cosmosus, AMTR – Amborella trichopoda, Ao – Asparagus officinalis, At – Arabidopsis thaliana, Bn – Brassica napus, Ch – Cuphea hookeriana, Cr – Chlamydomonas
1.4 - BIOTECHNOLOGICAL APPLICATIONS OF PLANT ACYL THIOESTERASES

1.4.1 - Acyl-ACP thioesterases have biotechnological potential as sources of industrially valuable medium-chain oleochemicals

Acyl-ACP thioesterases are of biotechnological interest due to the many industrial uses of medium-chain fatty acids and their oleochemical derivatives (i.e. methylketones, fatty alcohols, esters). These compounds have anticorrosive, antimicrobial, and lubricant properties (Kuznetsov & Ibatullin, 2002; Korlipara et al., 2011) and can be used for various industrial purposes including pharmaceuticals (Molly & Bruggemann, 2003), surfactants, cosmetics (reviewed in Hayes, 2004), bioplastics (Srivastava & Tripathi, 2013; Yan et al. 2022), flavourings (Hu et al., 2018) and jet fuel (reviewed in Knothe, 2008; Lennen et al., 2010). Methylketones, specifically, are also used or have the potential to be used for various other commercial applications, including fragrances (Goh et al., 2012), insect repellants (Roe, 2002), anaesthetics (Papachristoforou et al., 2012), and biodiesel (Goh et al., 2012). However, the production of medium-chain fatty acids and their derivatives for industry currently depends on non-renewable sources or unsustainable practices. Medium-chain fatty acids are mainly sourced from the cultivation of oil palm as an oleaginous crop, which has a very negative environmental impact. Palm oil processing is associated with massive greenhouse gas emissions, and deforestation to clear land for oil palm cultivation is considered the primary cause of biodiversity loss in tropical regions of Southeast Asia (Meijaard et al., 2020). While palm oil is
somewhat enriched in medium-chain content compared to other crop species, this represents only 5% of total oil composition, and certain chain-lengths, such as 8:0 and 10:0, are not very abundant (Rupilius & Ahmad, 2007). Isolating and purifying desirable fatty acid chain-lengths, and converting these to other chemical derivatives, also requires additional effort and resources. Therefore, certain medium-chain fatty acids, and derivatives that are difficult to source naturally like methylketones, are mainly obtained from chemical modification of petrochemical feedstocks (McKeon, 2016; Park et al., 2012; Phippen et al., 2006). Developing sustainable replacements for palm oil and fossil hydrocarbons as sources of oleochemicals for industry is a global challenge.

By manipulating the pathways that plants and microbes naturally use to synthesize acyl lipids to improve their efficiency and yield, biochemical engineers aim to address the issue of sustainable production of oleochemicals for industrial applications. Metabolic engineering strategies for overproduction of medium-chain fatty acids and their derivatives are typically based on one of two approaches: the overexpression of thiolase enzymes to drive β-reduction, in other words, to reverse the fatty acid β-oxidation pathway, or the overexpression of acyl-ACP thioesterase enzymes to interrupt de novo fatty acid biosynthesis at the desired point and direct intermediates away from this cycle (Dellomonaco et al., 2011; Lennen et al., 2012). While β-reduction strategies have resulted in very high total medium-chain fatty acid and fatty alcohol yields in bacteria, a major disadvantage of this process is a lack of controlled chain-length selectivity (Dellomonaco et al., 2011; Mehrer et al., 2018; Wu et al., 2017). In contrast, heterologous overexpression of an acyl-ACP thioesterase with known chain-length specificity ensures that a pool of desirable acyl chain lengths will accumulate. Along with thioesterase overexpression, the acyl-CoA synthetase that directs fatty acids towards β-oxidation is also
usually inactivated, in order to prevent newly hydrolyzed acyl chains from entering a futile cycle. Ultimately, generating only products of suitable chain-lengths for certain industrial applications is more cost-effective than the energy savings that come from reversing β-oxidation (Yan et al., 2022). Additional recombinant enzymes can be expressed to convert the fatty acids released by acyl-ACP thioesterases to alcohols, methylketones, and esters.

Much effort has therefore been applied toward generating medium-chain fatty acids and their derivatives in microbes via heterologous expression of plant acyl-ACP thioesterases, with the end goal of developing a microbial strain capable of efficiently producing these valuable compounds on an industrial scale (Cao et al., 2014; Fan et al., 2013; Goh et al., 2012; 2014; Lennen et al., 2011; Lin et al., 2018; Steen et al., 2010; Wu & Sun, 2014; Hernández Lozada et al., 2020; Yan et al., 2022). Additional focus has also been placed on optimizing other parts of the fatty acid biosynthesis cycle for increased medium-chain fatty acid yield, and improving the utilization of abundant carbon sources by these engineered microbes. Overexpression of enzymes that catalyze rate-limiting steps of de novo fatty acid biosynthesis, and replacement of native ketoacyl-ACP synthase isoforms with ones incapable of elongating fatty acyl-ACPs beyond a certain chain-length, has been shown to dramatically increase medium-chain fatty acid titers in E. coli (Yu et al., 2011; Torella et al., 2013). Bacterial strains where free fatty acid accumulation is greatly improved through the inactivation of the fadD acyl-CoA synthetase, which catalyzes the first step of fatty acid β-oxidation, are also routinely used to characterize the enzymatic activities of various plant acyl-ACP thioesterases (Jha et al., 2010; Jing et al., 2011; Jones et al., 1995; Leonard et al., 1997; Pulsifer et al., 2014; Ghosh et al., 2007).

Another major area of research interest surrounds the development of transgenic crops that readily incorporate high-value medium-chain fatty acids into triacylglycerols. Most of these
efforts involve the seed-specific heterologous expression of specialized FatB-type thioesterases in oilseed plants, in order to increase medium-chain-length content in seed triacylglycerols. For example, the overexpression of FatB1 from *Umbellularia californica*, FatB2 from *Cuphea hookeriana*, or FatB2 from *Cuphea palustris* has been used to dramatically increase the proportion of C12:0, C10:0, or C14:0 acyl chains in the seed oil of *Camelina sativa* and *Brassica napus* (Dehesh *et al*., 1996; Eccleston *et al*., 1996; Voelker *et al*., 1996; Tjellström *et al*., 2013; Kim *et al*., 2015; Hu *et al*., 2017; Bansal *et al*., 2018). Other strategies aim to increase the accumulation of triacylglycerols in vegetative tissues of high-biomass crop plants, such as tobacco, *Sorghum bicolor*, and sugarcane (*Saccharum* spp.), through the combined overexpression of medium-chain-specific thioesterases and acyltransferases (Reynolds *et al*., 2017; Vanhercke *et al*., 2019; Park *et al*., 2021). Acyltransferases that efficiently incorporate medium acyl chains into triacylglycerol channel the fatty acids released by the thioesterase towards oil production, rather than membrane lipid biosynthesis (Reynolds *et al*., 2017).

1.4.2 – *Recombinant acyl-ACP thioesterases must be optimized for efficient production of medium-chain fatty acids*

While acyl-ACP thioesterases from plants have significant biotechnological potential as sustainable sources of medium-chain fatty acids for industrial applications, some challenges must be addressed before they can be used effectively as such. Preference for undesirable acyl chain lengths that complicates downstream product purification, low activity of thioesterase enzymes in heterologous hosts, or the inability of such hosts to efficiently incorporate certain chain lengths generated by acyl-ACP thioesterases into larger lipid molecules, are all major obstacles to the biological production of valuable lipids containing medium-chain acyl groups. The manipulation of endogenous enzymes involved in fatty acid biosynthesis or acyl chain
trafficking, co-expression of other recombinant enzymes that promote the accumulation of desirable medium-chain fatty acids, and optimization of thioesterase specificity and activity are likely all required to overcome these problems.

In microbial systems, where de novo fatty acid biosynthesis is not compartmentalized, medium-chain fatty acid buildup must be mitigated, as the accumulation of these compounds in the cell leads to membrane toxicity (Lennen et al., 2011; Lin et al., 2018; Liu et al., 2013; Royce et al., 2013). Additionally, the overexpression of medium-chain specific acyl-ACP thioesterases often results in a concomitant decrease in the activity of acyl synthetases that activate fatty acids to acyl-CoA thioesters for further metabolism (Lin et al., 2018). The size of available fatty acyl-ACP and fatty acyl-CoA pools must be carefully balanced to promote the biosynthesis of desirable lipids, and the development of microbial systems for renewable lipid production usually also involves altering expression of endogenous acyl synthetases, or heterologously expressing an acyl synthetase from another organism. For example, overexpression of endogenous long-chain acyl-CoA synthetase (LACS) genes in E. coli has been used to boost the production of C12-C18 fatty acid ethyl esters and the hexadecyl wax esters of C14-C18 fatty acids, which are used in the manufacture of biodiesel, soaps, and surfactants (Steen et al., 2010). Overexpressing LACS genes from Arabidopsis and Brassica napus improves storage lipid accumulation in yeast (Pulsifer et al., 2012; Tan et al., 2014).

Meanwhile, in transgenic oilseed crops, incorporation of medium-chain fatty acids into storage lipids is often inefficient. In species that do not naturally incorporate medium-chain fatty acids into their triacylglycerols, endogenous acyltransferase enzymes generally have poor affinity for medium-chain acyl-CoA thioesters. As a result, these thioesters are directed preferentially to the β-oxidation pathway rather than to triacylglycerol biosynthesis (Bansal et al., 2010).
al., 2018; Larson et al., 2002). Silencing or knocking out endogenous acyltransferases, and in their place, expressing medium-chain specific acyltransferases from plants such as Cuphea spp. and coconut can increase medium-chain fatty acid content in the seed or leaf triacylglycerols of transgenic crops (Reynolds et al., 2017; Bansal et al., 2018). Tjellström et al. (2013) demonstrated that increased medium-chain fatty acid production can also be attained through the used of transgenic plant lines deficient in the activity of acyl-activating (AAE) enzymes, which normally serve to re-activate prematurely hydrolyzed (i.e. short- and medium-chain) fatty acids for further elongation.

In addition to manipulating the expression of other enzymes alongside plant acyl-ACP thioesterases to optimize production of industrially valuable lipids, many recent studies have centered on engineering Fat-type thioesterase enzymes themselves for increased activity toward their native substrates, or for altered substrate specificity towards a desired acyl-ACP chain-length. The experimental pipelines used to accomplish this generally require a rational design element; genetic selection strategies, iterative “domain-swapping” approaches, and structure-guided mutagenesis based on computational redesign have all been used to increase the activity of plant thioesterases or direct their substrate preferences toward acyl-ACP chain lengths that suit specific industrial purposes (Feng et al., 2017; Grisewood et al., 2017; Hernández Lozada et al., 2018; Jing et al., 2018; Ziesack et al., 2018). The success of targeted mutagenesis approaches in manipulating the acyl chain-length specificity of medium-chain-specific FatB thioesterases inspired the work described in Chapter 4 of this thesis.

Given their natural substrate specificities described in Section 1.3.4, ALT-type thioesterases are very promising candidates for protein engineering as potential renewable sources of medium-chain fatty acids and their derivatives. The unique ability of ALTs to
produce functionalized β-keto fatty acids directly is particularly advantageous, since it means that engineered ALTs could be very efficient sources of valuable medium-chain methylketones. Beta-keto fatty acids produced by overexpressed ALT enzymes could be rapidly decarboxylated to methylketones through exposure to low heat and dilute acid, or through the activity of a simultaneously expressed recombinant ketoacid decarboxylase (Yu et al., 2010). Through mutagenesis and screening, highly active ALT variants could be developed, and the chain length and oxidation state preferences of ALT enzymes could be tuned towards the β-keto fatty acid precursors of either the 7-11 carbon saturated methyl ketones that are widely used as flavourings, or the 11–13 carbon monounsaturated methyl ketones useful in the manufacture of biofuels (Goh et al., 2012; reviewed in Hagedorn and Kaphammer, 1994).

1.5 – THESIS OBJECTIVES

In contrast to other plant acyl-ACP thioesterase enzymes, the biochemistry and functions of ALT-type thioesterases are poorly understood overall. This must be investigated before the biotechnological applications of ALT enzymes can be explored. In this thesis, I present work that provides new insights into the natural catalytic diversity found within the ALT enzyme family, their behaviour in planta as well as in bacterial systems, and the factors that dictate their wide-ranging substrate preferences. In Chapter 2, I describe the functional characterization of ALT-type thioesterases from a diverse set of plant species through heterologous expression in <i>Escherichia coli</i>, and demonstrate that ALTs exhibit much more variation in their substrate specificities than was previously understood. I provide evidence that certain ALT enzymes are capable of acting on the 3-hydroxyacyl-ACP intermediates of fatty acid biosynthesis, giving rise to medium-chain 3-hydroxy fatty acid products in addition to β-keto- and fully reduced fatty acids. In Chapter 3, I investigate the behaviour of ALT enzymes in planta, and present
findings which demonstrate that, if optimized for greater catalytic efficiency, ALT-type thioesterases could be used as tools to modify the lipid profile of plant species of biotechnological interest, such as oilseed crops. These results show that the substrate specificities of ALT enzymes \textit{in planta} and in \textit{E. coli} are in general agreement, and that the seed-specific overexpression of ALT-type thioesterases leads to increased medium-chain content in seed triacylglycerols. In Chapter 4, I describe the identification and characterization of several amino acid motifs that dictate chain-length and oxidation state specificity in ALT enzymes, through targeted mutagenesis experiments guided by computational modelling. To conclude this thesis, in Chapter 5, I propose future experiments to improve the activity and selectivity of ALT variants towards desirable acyl-ACP substrates, as well as experiments to elucidate the physiological functions of ALT enzymes in non-solanaceous plants, which have proven difficult to characterize. Understanding the biological significance of ALT-type thioesterases will improve our understanding of the roles of medium-chain fatty acids and their derivatives in plant secondary metabolism, which are not as well-studied as other classes of lipid-based secondary metabolites.
Chapter 2: Elucidating the substrate specificities of acyl-lipid thioesterases from diverse plant taxa

2.1 - ABSTRACT

Acyl-acyl carrier protein (ACP) thioesterase enzymes, which cleave fatty acyl thioester bonds to release free fatty acids, contribute to much of the fatty acid diversity in plants. In *Arabidopsis thaliana*, a family of four single hot dog-fold domain, plastid-localized acyl-lipid thioesterases (*At*ALT1-4) generate medium-chain (C6-C14) fatty acids and β–keto fatty acids as secondary metabolites. These volatile products may serve to attract insect pollinators, or deter predatory insects and pathogens. Homologs of *At*ALT1-4 are present in all plant taxa, but were nearly all uncharacterized prior to this study. Despite high sequence identity, *At*ALT1-4 generate different lipid products, suggesting that ALT homologs in other plants also have highly varied activities. The catalytic diversity of ALTs was investigated by screening the substrate specificities of 15 ALT homologs from monocots, eudicots, a lycophyte, a green microalga, and the ancient gymnosperm *Ginkgo biloba*, via expression in *Escherichia coli*. Overall, these enzymes had highly varied substrate preferences compared to one another and *At*ALT1-4, and could be classified into four catalytic groups comprising members from diverse taxa. Group 1 ALTs primarily generated 14:1 β-keto fatty acid and novel 14-carbon 3-hydroxy fatty acids, Group 2 ALTs produced 6-10 carbon fatty/β-keto fatty acids, Group 3 ALTs predominantly produced 12-14 carbon fatty acids, and Group 4 ALTs mainly generated 16 carbon fatty acids. Enzymes in each group differed significantly in the quantities of lipids and types of minor products they generated in *E. coli*. Medium-chain fatty acids are used to manufacture insecticides, pharmaceuticals, and biofuels, and ALT proteins are ideal candidates for metabolic
engineering to produce medium-chain fatty acids and their derivatives in significant quantities.

2.2 - INTRODUCTION

To participate in lipid metabolism, fatty acids must first be activated via a thioester linkage to an acyl carrier protein (ACP) or coenzyme A (CoA) (reviewed in Li-Beisson et al., 2013). In plant cells, de novo fatty acid synthesis mainly occurs in the plastid, through a cyclic pathway in which an ACP-linked fatty acid substrate is sequentially elongated by two-carbon units (Ohlrogge & Jaworski, 1997). Fatty acids are then directed to either the plastid-localized ‘prokaryotic’ lipid biosynthesis pathway by acyltransferases, or the endoplasmic reticulum (ER)-localized ‘eukaryotic’ lipid biosynthesis pathway by acyl-ACP thioesterases (reviewed in Benning, 2009). Acyl-ACP thioesterases cleave the thioester bond that links the growing fatty acyl chain to the phosphopantetheine cofactor carried by the ACP, releasing free, metabolically inactive fatty acids (Ohlrogge & Jaworski, 1997; Benning, 2009). These are then exported from the plastid to the ER, where they are incorporated into glycerophospholipids, triacylglycerols, and other polar and neutral lipids (Jones et al., 1995; Browse & Sommerville, 1991).

As 16:0 and 18:1 fatty acids are the major components of plant lipids, all plants possess primary metabolic plastid-localized acyl-ACP thioesterases that act on 16- or 18-carbon substrates. These are classified as FatA- and FatB-type thioesterases, which produce unsaturated and saturated fatty acids, respectively (Jones et al., 1995). In some oilseed plants, such as Cuphea spp. and Umbellularia californica, specialized FATB-type thioesterases that yield medium-chain (6-14 carbon) fatty acids are expressed in seeds, where their products are incorporated into seed oil triacylglycerols (Dehesh et al., 1996; Eccleston & Ohlrogge, 1998; Tjellström et al., 2013).
Plant acyl-ACP thioesterases characterized to date belong to the “hot dog-fold” protein superfamily, named for their thioesterase domain structure consisting of a 5-7 strand antiparallel β-sheet surrounding a central α-helix (Dillon & Bateman, 2004). The hot dog-fold domain dimerizes or forms higher-order oligomers in vivo (Pidugu et al., 2009). While FatA- and FatB-type thioesterases possess a double hot dog-fold domain, single hot dog-fold acyl-ACP thioesterases that generate medium-chain fatty acids have also been characterized in wild tomato (Solanum habrochaites subsp. glabratum) and Arabidopsis thaliana. The wild tomato METHYL KETONE SYNTHASE 2 (ShMKS2) was the first plant acyl-ACP thioesterase found to be involved in specialized metabolic processes other than seed oil production (Ben-Israel et al., 2009). In trichome cell chloroplasts, ShMKS2 hydrolyzes 12-16 carbon β-ketoacyl-ACPs to release β-keto fatty acids, which are oxidized to methylketones by a decarboxylase, ShMKS1 (Yu et al., 2010). These volatile methylketones act as insecticides (Fridman et al., 2005; Antonious, 2003).

Following the characterization of ShMKS2, a family of four single hotdog-fold, plastid-localized thioesterases, named ACYL LIPID THIOESTERASE 1-4 (AtALT1-4) was discovered in Arabidopsis thaliana (Pulsifer et al., 2014). Heterologous expression of AtALT1-4 in Escherichia coli revealed that despite sharing over 75% amino acid sequence identity, each of these thioesterases has unique substrate specificity in terms of chain length, saturation level, and oxidation state (Pulsifer et al., 2014). When expressed in E. coli, AtALT1 primarily produces 12:0 and 14:1 fatty acids, AtALT2 mainly produces 8:0 and 10:0 β-keto fatty acids, and AtALT4 mostly generates 6:0 and 8:0 fatty acids. AtALT3 has broader specificity than the other Arabidopsis ALTs when expressed in E. coli, but primarily produces 14:1 β-keto fatty acids, with minor β-keto fatty acid products ranging from 8-16 carbons in length (Pulsifer et
Although the precise biological roles of \textit{AtALT1-4} are unknown, their medium-chain fatty acid and \(\beta\)-keto fatty acid products are likely specialized metabolites that function in plant defence. \textit{AtALT1} and \textit{AtALT2} co-regulate with cuticular wax and suberin biosynthetic genes, respectively, indicating that their volatile products may accumulate in the cuticle and suberized tissues to provide protection against pathogens and insect pests (Pulsifer \textit{et al.}, 2014). \textit{AtALT4} is expressed in anthers, which suggests that the volatile 6:0 and 8:0 fatty acids it generates may serve to attract insect pollinators or bodyguards (Pulsifer \textit{et al.}, 2014). In keeping with its broad substrate specificity, \textit{AtALT3} is expressed in all plant tissues. The \(\beta\)-keto fatty acids it produces may be converted to methylketones for plant-wide insect defence.

Homologs of \textit{AtALT1-4} exist in all plant taxa, indicating that their metabolites are widely important (Pulsifer \textit{et al.}, 2014). However, apart from \textit{ShMKS2} and \textit{AtALT1-4}, the ALT enzyme family remains largely uncharacterized. Given the widespread activity profiles of \textit{AtALT1-4}, we hypothesized that the few previously characterized ALTs likely did not represent the full range of catalytic diversity that exists among these thioesterases. In this study, the natural diversity of ALT enzymes was investigated by screening the substrate specificities of 15 ALTs from representative species in a wide range of plant taxa, including monocots, eudicots, lycophytes, green microalgae, and the ancient gymnosperm order Ginkgophyllales, of which \textit{Ginkgo biloba} is the only extant member.

\textbf{2.3 – MATERIALS AND METHODS}

\textit{2.3.1 - Phylogenetic analysis of ALT protein sequences}

Basic local alignment searches of the NCBI GenBank, EnsemblPlants, Cannabis Genome
Browser, and the annotated transcriptome of *Ginkgo biloba* (GigaDB accession: PRJNA307642) were conducted using the protein sequences of *AtALT1-4* as queries (Van Bakel *et al.*, 2015; Guan *et al.*, 2016). A multiple sequence alignment was generated from these sequences using the MUSCLE algorithm with default parameters (Edgar, 2004). A phylogenetic tree was constructed from this alignment in MEGA X software using the maximum likelihood model (Kumar *et al.*, 2018; Guindon & Gascuel, 2003). The Jones-Taylor-Thornton model was chosen for amino acid substitutions, and gaps were subject to partial deletion below 90% coverage (Jones *et al.*, 1992). The tree was rooted using a distant homologue, the single hot dog-fold acyl-CoA thioesterase YbgC from *E. coli* (UNIPROT ID: P0A8Z3) as the outgroup. The phylogeny was verified by a bootstrap test consisting of 1000 replicates.

2.3.2 - Generating DNA constructs for expression of ALT proteins in K27(DE3) *E. coli*

The coding sequences of *Z. mays ALT1* (XM_008670822.4), *Z. mays ALT2* (NM_001326549.1), *Z. mays ALT3* (XM_008647798.1), *Z. mays ALT4* (XM_008647802.3), *G. max ALT1* (NM_001377384.1), *V. vinifera ALT1* (XM_002283509.4), *O. sativa* subsp. *japonica ALT1* (XM_473441.1), *B. distachyon ALT1* (XM_003563133.4), *B. distachyon ALT2* (XM_003581459.4), *M. truncatula ALT1* (XM_024783401.2), *M. truncatula ALT2* (XM_003614577.4), and *C. reinhardtii ALT1* (XM_043061639.1) were obtained from the NCBI GenBank. The coding sequence of *C. sativa ALT1* (PK15701.1) was retrieved from the Cannabis Genome Browser Gateway database, and that of *S. moellendorffii ALT1* (SELMODRAFT_128895) was retrieved from EnsemblPlants (Van Bakel *et al.*, 2015). The *G. biloba ALT1* coding sequence was obtained from publicly available Ginkgo leaf transcriptomic data (Guan *et al.*, 2016).
Double restriction digests of pET-28a plasmid DNA using restriction endonucleases BamHI and HindIII, or BamHI and EcoRI, were prepared. Digested plasmid DNA was run on a 0.8% agarose gel, and the linearized plasmid was excised and purified using a NucleoSpin Gel and PCR Purification Kit (Macherey-Nagel). The coding sequences of the ALT genes listed above (excluding the predicted N-terminal plastid targeting sequence) were chemically synthesized with 16-bp extensions at their 5’ and 3’ ends, such that the synthesized DNA could be ligated into the digested pET-28a vector using an In-Fusion Snap Assembly Kit (Takara-Clontech). The 16-bp extensions were complementary to the pET-28a vector sequence at the site of linearization, and introduced a BamHI restriction site immediately before the start codon of each gene, and, in most cases, a HindIII restriction site after the stop codon. For SmALT1, CrALT1, and GbALT1, an EcoRI restriction site was introduced after the stop codon because these genes contain HindIII restriction sites within the coding region. ALT-encoding insert sequences used to assemble these constructs are listed in Appendix 3.

The pET-28a vector also introduced a 6xHis tag at the N-terminus of each ALT protein. Isolated plasmid DNA was subsequently transformed into the K27(DE3) strain of E. coli (Overath et al., 1969). The presence of the DE3 lysogen allows for isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible expression of T7 RNA polymerase, which subsequently drives expression of the transgene, in this case, the ALT gene, present on the pET-28a plasmid (Lü et al., 2009).

2.3.3 - Profiling and quantification of lipids by GC-FID and GC-MS

K27(DE3) E. coli transformed with pET-28a (empty vector) or pET-28a containing ALT coding sequences were grown in 50 mL of kanamycin-containing Luria-Bertani (LB) media at
37 °C with 200 rpm shaking, to an OD$_{600}$ of 0.6-0.7. IPTG was added to a final concentration of 0.5 mM to induce ALT protein expression, and cells were grown for 18 h at 18 °C with 200 rpm shaking. Cells from 5 mL of each induced bacterial culture were pelleted by centrifugation in glass tubes. One mL of culture supernatant was mixed with 1 mL of 20 mM H$_2$SO$_4$ and incubated at 75 °C for 30 min. Samples were allowed to cool to room temperature, after which 10 μL of a 0.1 μg / μL solution of tetracosane (24:0 alkane) dissolved in toluene was added to each sample as an internal standard, for a total of 1 μg tetracosane / sample. Lipids were then extracted into 250 μL hexane by vortexing for 30 s. Sample vials were centrifuged at 4,500 x g following extraction to separate phases. One μL of the upper hexane layer was used for gas chromatographic analysis. Lipids were profiled and quantified using a Varian 450-GC gas chromatograph equipped with an HP-5MS column (30 m length, 0.25 mm inner diameter, 0.25 μm film thickness) and a flame ionization detector (FID). Samples were injected in splitless mode, with an injector temperature of 250 °C. The carrier gas was helium at a constant flow rate of 2 mL / min. The column oven temperature was initially held at 50 °C for 8 min, then increased at a rate of 15 °C / min up to 325 °C, which was held for 4 min.

An Agilent 7890A gas chromatograph equipped with an HP-5MS column (30 m length, 0.25 mm inner diameter, 0.25 μm film thickness) and an Agilent 5977E mass spectrometric detector (70 eV ionization energy, mass-to-charge ratio 40-500) was used to obtain electron ionization (EI) mass spectra for compound identification in lipid extracts. Sample injection, carrier gas flow, and oven temperature parameters were the same as described above. All fatty acids and saturated methylketones were identified via their mass spectra in comparison to standard spectra in the National Institute of Standards and Technology NIST 17 Spectral Library, and/or their retention time in comparison to the GC-FID spectra of authentic standards.
Monounsaturated methylketones were identified by comparison to spectra provided by Pulsifer et al. (2014) and Goh et al. (2012). All sample and standard spectra used to determine compound identities are shown in Appendix 4.

2.3.4 - Esterification of free fatty acids in bacterial culture media

ALT protein expression was induced in K27(DE3) E. coli as described previously. Cells from 5 mL of culture were pelleted via centrifugation at 4,500 x g. One mL of culture supernatant was mixed with 2 mL of 1:1 chloroform/methanol and 44 μL of glacial acetic acid. Lipids were extracted by vortexing for 10 s, and phases were separated by centrifugation at 4,500 x g. The lower phase was transferred to a clean glass tube, and evaporated to dryness under a N₂ stream. Samples were resuspended in 1 mL of 2% H₂SO₄ in methanol and incubated at 90 °C for 1 hr, after which 1 mL of 0.9% NaCl was added to stop the esterification reaction. Fatty acid methyl esters were then extracted into 250 μL hexane by vortexing for 30 s, and phases were separated by centrifugation. Fatty acid methyl esters were identified and quantified via GC-FID and GC-MS as described in Section 2.3.3.

2.3.5 - Silylation of fatty acids produced by ALT-expressing E coli to assist identification of 3-hydroxy fatty acids

ALT expression was induced in E. coli and secreted lipids were extracted as described above. Hexane extracts were evaporated to near dryness under a gentle stream of nitrogen gas. Dried samples were then resuspended in 100 μL N,O-Bis(trimethylsilyl)trifluoroacetamide (BSTFA) with 100 μL pyridine as a catalyst. Samples were incubated at 115°C for 15 minutes to silylate carboxyl and secondary alcohol groups, after which they were evaporated again to near dryness under nitrogen gas. Samples were resuspended in 200 μL hexane, and 1 μL of
each resuspended sample was used for GC-MS analysis as described in Section 2.3.3. The 2-trimethylsilyl (TMS) derivatives of 3-hydroxy fatty acids were identified by comparison with standard spectra in the NIST 17 Mass Spectral Library, or by the presence of the molecular ion (M⁺) and characteristic fragment ions (m/z = 73, m/z = 147, m/z = 233, M⁺ – 31) when no standard spectrum was available. These mass spectra are shown in Fig. 2.3, and Appendix 4.

2.3.6 - Detection of ALT protein expression via Western blotting

ALT protein expression was induced in K27(DE3) E. coli as described previously. After 18 h of incubation at 18 °C, the OD₆₀₀ of each bacterial culture was measured. Aliquots of each culture corresponding to 200 OD₆₀₀ units were centrifuged at 4,500 x g for 5 min to harvest cells. Cell pellets were resuspended in 200 μL lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM PMSF, 10 μL prepared E. coli protease inhibitor cocktail (Sigma-Aldrich). Cells were placed on ice and lysed by 3 x 15 s sonication cycles, with 30 s wait time between cycles. Lysates were clarified by centrifugation at 13,000 x g for 2 min. The total protein concentrations of clarified lysates were determined using a Bio-Rad protein assay kit (modified Bradford dye-binding method). Total protein concentration was the average of triplicates of each sample. Aliquots of lysate corresponding to 20 μg of total protein were mixed with 5 μL of loading buffer (120 mM Tris-HCl, pH 6.8, 20% glycerol, 4% SDS, 0.02% w/v bromophenol blue, 50 mM dithiothreitol) and incubated in a boiling water bath for 5 min. These samples were loaded onto a discontinuous SDS-PAGE gel (6% acrylamide stacking gel, 12% acrylamide resolving gel), and electrophoresis was conducted for 30 min at 80 V followed by 1 h at 120 V.

Proteins were then transferred onto a nitrocellulose membrane at 270 mA for 1 h. The membrane was incubated in 10 mL blocking solution (3% w/v bovine serum albumin) overnight.
at 4 °C. The blocked membrane was then probed with horseradish peroxidase (HRP)-conjugated mouse monoclonal anti-6x His-tag antibody (Thermo Fisher, 1:35 000 dilution) to detect ALT proteins, and HRP-conjugated mouse monoclonal anti-GAPDH antibody (Invitrogen, 1:35 000 dilution) as a loading control. Antibody binding was detected with Supersignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher). Protein band signal intensity was measured using VisionCapt 15 software (Vilber-Lourmat).

2.3.7 – *Homology modelling and structure-based phylogenetic analysis of ALT proteins*

The hot dog-fold domain structures of selected ALTs were modelled using SWISS-MODEL Workspace v8.05, with the crystal structure of the single hot dog-fold thioesterase YbgC from *E. coli* (RCSB PDB Accession 1S5Ua) as the template (Biasini *et al.* 2014). Homology models were then aligned using the STAMP (STructural Alignment of Multiple Proteins) algorithm, and a structure-based phylogenetic tree was constructed in Visual Molecular Dynamics (VMD) v1.9.3 software, using the root-mean-square distances of aligned structures (Russell & Barton, 1992; Humphrey *et al*., 1996; Eargle *et al*., 2005).

2.4 – **RESULTS**

2.4.1 - *Identification and selection of ALT enzymes in diverse plant taxa*

ALT enzymes were identified in plant protein databases using the hot dog-fold domain sequences of *Arabidopsis thaliana* ALT1-4 as basic local alignment search queries. A total of 102 distinct ALT proteins in 71 species were found after removal of partial sequences and duplicate sequences in polyploid species, and these full-length predicted sequences were used to construct a maximum likelihood phylogenetic tree (Fig. 2.1). Analysis of this tree shows that ALTs are grouped mainly by occurrence in plant species or families (i.e. Brassicaceae, Poaceae,
etc.). This contrasts with the FatA- and FatB-type plant acyl-ACP thioesterases, which have a different phylogenetic origin than ALT enzymes, and are mainly grouped in a phylogenetic tree by their function (Pulsifer et al., 2014; Caswell et al., 2022). Numerous plant species, such as *Populus trichocarpa*, *Vitis vinifera* (grapevine), *Glycine max* (soybean) and *Brassica rapa*, possess only one ALT-encoding gene in their genome (Fig. 2.1). Others have multiple, paralogous ALT-encoding genes usually numbering from 2 to 4; *Sorghum bicolor* and *Setaria italica* (millet) of the Poaceae family possess 5 and 6 distinct *ALT* genes, respectively (Fig. 2.1).
Figure 2.1. Phylogenetic tree of 102 ALTs from 71 plant and green microalga species. ALT sequences were identified using basic local alignment searches of protein and translated RNA databases with *Arabidopsis* ALT1-4 protein sequences as queries. Thioesterases that were functionally characterized in this study are indicated with an asterisk.

Fifteen ALT enzymes from monocot, eudicot, lycophyte, green microalgae, and gymnosperm species were functionally characterized in this study through heterologous expression in *E. coli* bacteria. These were selected based on their phylogenetic relationships to widely survey the potential catalytic diversity of the ALT enzyme family both between and within species and taxa. Emphasis was placed on commercially grown and model plant species for which transcriptome profiling data is available. Like *Arabidopsis*, the monocot crop *Zea mays* (maize) has four ALT homologs in its genome. These four enzymes, which we named *ZmALT1*-4, were characterized to determine whether they exhibit a high level of catalytic diversity similar to *AtALT1*-4. Two ALT-encoding genes from the model plant *Brachypodium distachyon* (*BdALT1*-2), another member of the Poaceae, were also selected. Phylogenetic analysis of ALTs from the Poaceae shows that they can be divided into two distinct subclades, and *BdALT1*-2 are representative enzymes from each subclade, respectively (Fig. 2.1). ALT homologues from *Oryza sativa* subsp. *indica* (*OsALT1*), and the commercially grown eudicot species *Vitis vinifera* (*VvALT1*), *Glycine max* (*GmALT1*), and *Cannabis sativa* (*CsALT1*) were also selected for functional characterization. Finally, ALT homologues from the lycophyte *Selaginella moellendorffii*, the green microalga *Chlamydomonas reinhardtii*, and the only
extant member of the ancient gymnosperm family Ginkgoaceae, *Ginkgo biloba*, were chosen as representative ALTs from early-diverging plant lineages; these were named *SmALT1*, *CrALT1*, and *GbALT1* respectively (Fig. 2.1).

The substrate specificities of the selected ALT enzymes were characterized using the endogenous acyl-ACP pool of K27(DE3) *E. coli* as substrates. The K27(DE3) strain harbours a loss-of-function mutation in the *fadD* acyl-CoA synthetase gene that prevents degradation of fatty acids by β-oxidation, resulting in accumulation of fatty acids in the cell and their subsequent export into the culture medium (Overath *et al.*, 1969, Lü *et al.*, 2009). Lipids secreted into the culture medium were identified and quantified via GC-MS and GC-FID, respectively (Figs 2.2, 2.3, Appendix 4). *E. coli* harbouring an empty pET-28a vector was used as a negative control. Bacteria expressing the previously characterized thioesterases *AtALT1*, *AtALT3*, and *AtALT4* served as positive controls, as the products of these enzymes in *E. coli* encompass the entire range of medium-chain saturated and monounsaturated fatty acids (FAs) and β-keto fatty acids (β-keto FAs) that can be generated from the endogenous acyl-ACP substrate pool (Pulsifer *et al.*, 2014). As β-keto fatty acids are too unstable to be detected by GC-MS or GC-FID, these were decarboxylated to methylketones (MKs) by treating media samples with heat and acid prior to analysis (Fig. 2.2) (Yu *et al.*, 2010). Conversion of β-keto fatty acids to methylketones shortens the chain by one carbon.

2.4.2 – *Medium-chain 3-hydroxy fatty acids are observed as novel products of ALT enzymes in E. coli*

The GC-FID traces of lipid extracts from culture media of K27(DE3) *E. coli* strains expressing several ALTs, including *BdALT1*, *CsALT1*, *MtALT2*, *OsALT1*, *VvALT1*, and
*ZmA*LT2, displayed a large peak that was not present in lipid extracts from *E. coli* strains expressing *At*ALT1, *At*ALT2, or *At*ALT4, and only a minor product of *At*ALT3 expression (Fig. 2.2). GC-MS analysis initially showed this peak to consist of co-eluting compounds, the identities of which could not initially be determined (Fig. 2.3). Free fatty acids in culture media from *Vv*ALT1- and *Mt*ALT2-expressing bacteria were converted to fatty acid methyl esters, as conversion of fatty acids to their corresponding methyl esters creates a shift in column retention time. Sample esterification caused the peak in question to split into two, as shown in Fig. 2.3. The larger of the two peaks could still not be conclusively identified by its mass spectrum, but the smaller of the two peaks was identified as the methyl ester of a 3-hydroxy fatty acid by the characteristic base peak of m/z = 103 (Fig. 2.3). When NIST 17 mass spectral library was queried with this mass spectrum, 3-hydroxyhexadecanoic acid was returned as the closest match (Fig. 2.3, Appendix 4). However, due to loss of the molecular ion and some higher m/z fragments, the carbon chain-length of 3-hydroxy fatty acid methyl esters can be difficult to determine accurately by electron ionization (EI) GC-MS.
Fig. 2.2. Example GC-FID traces of lipids secreted into the media of K27(DE3) E. coli expressing the MtALT2 enzyme, as compared to E. coli expressing Arabidopsis AtALT3 or harbouring an empty pET-28a vector. \( \beta \)-keto fatty acids were decarboxylated to methylketones with heat and acid treatment prior to extraction of lipids and GC analysis. Unidentified peaks are indicated by a question mark. Compounds corresponding to peak numbers are reported in a table (FA = fatty acid, MK = methylketone, 3-OH FA = 3-hydroxy fatty acid, IST = internal standard).

<table>
<thead>
<tr>
<th>Peak</th>
<th>Compound Type</th>
<th>Compound Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7:0 MK</td>
<td>2-heptanone</td>
</tr>
<tr>
<td>2</td>
<td>6:0 FA</td>
<td>hexanoic acid</td>
</tr>
<tr>
<td>3</td>
<td>9:0 MK</td>
<td>2-nonanone</td>
</tr>
<tr>
<td>4</td>
<td>8:0 FA</td>
<td>octanoic acid</td>
</tr>
<tr>
<td>5</td>
<td>11:1 MK</td>
<td>(Z)-4-undecen-2-one</td>
</tr>
<tr>
<td>6</td>
<td>11:0 MK</td>
<td>2-undecanone</td>
</tr>
<tr>
<td>7</td>
<td>Indole</td>
<td>1H-indole</td>
</tr>
<tr>
<td>8</td>
<td>10:0 FA</td>
<td>decanoic acid</td>
</tr>
<tr>
<td>9</td>
<td>13:1 MK</td>
<td>(Z)-4-tridecen-2-one</td>
</tr>
<tr>
<td>10</td>
<td>13:0 MK</td>
<td>2-tridecanone</td>
</tr>
<tr>
<td>11</td>
<td>12:1 FA</td>
<td>(Z)-5 dodecenoic acid</td>
</tr>
<tr>
<td>12</td>
<td>12:0 FA</td>
<td>dodecanoic acid</td>
</tr>
<tr>
<td>13</td>
<td>15:1 MK</td>
<td>(Z)-8-pentadecen-2-one</td>
</tr>
<tr>
<td>14</td>
<td>15:0 MK</td>
<td>2-pentadecanoic acid</td>
</tr>
<tr>
<td>15</td>
<td>3-OH 12:0 FA</td>
<td>3-hydroxydodecanoic acid</td>
</tr>
<tr>
<td>16</td>
<td>14:1 FA</td>
<td>(Z)-7 tetradecanoic acid</td>
</tr>
<tr>
<td>17</td>
<td>14:0 FA</td>
<td>tetradecanoic acid</td>
</tr>
<tr>
<td>18</td>
<td>14:1 3-OH FA and 14:0 3-OH FA (co-eluting)</td>
<td>(Z)-3-hydroxy-7-tetradecenoic acid and 3-hydroxytetradecanoic acid (co-eluting)</td>
</tr>
<tr>
<td>19</td>
<td>16:1 FA</td>
<td>hexadecanoic acid</td>
</tr>
<tr>
<td>20</td>
<td>16:0 FA</td>
<td>(Z)-9 hexadecenoic acid</td>
</tr>
<tr>
<td>21</td>
<td>18:1 FA</td>
<td>octadecenoic acid</td>
</tr>
<tr>
<td>22</td>
<td>C24 IST</td>
<td>tetracosane</td>
</tr>
</tbody>
</table>
To aid in chain-length identification, trimethylsilyl (TMS) ethers were prepared from lipid extracts from \(Vv\)ALT1- and \(Mt\)ALT2-expressing \(E. coli\), since the addition of TMS groups to alcohols and acids produces larger molecular weight fragments. Both co-eluting compounds comprising the unknown peak (peak 18 in Fig. 2.2), (Z)-3-hydroxy-7-tetradecenoic acid (14:1 3-OH FA), and 3-hydroxytetradecanoic acid (14:0 3-OH FA), could be identified via the mass spectra of their TMS derivatives (Fig. 2.3, Appendix 4). (Z)-3-hydroxy-7-tetradecenoic acid, which is not present in the NIST 17 library, was identified from the molecular ion present in the mass spectrum of its 2-TMS derivative (m/z = 386) and fragments characteristic of the TMS derivatives of both a 3-hydroxy fatty acid (m/z = 73, m/z = 147, m/z = 233, M+ – 31) and a monounsaturated fatty acid (M+ – 169, M+ – 90). Another 3-hydroxy fatty acid, 3-hydroxydodecanoic acid, was also identified in transmethylated and silylated lipid extracts from bacterial cultures expressing \(Bd\)ALT1, \(Cs\)ALT1, \(Mt\)ALT2, and \(Vv\)ALT1, and, interestingly, \(At\)ALT1 (Fig. 2.3, Appendix 4).

Since monounsaturated 14:1 acyl chains are significantly more abundant in \(E. coli\) acyl-ACP pools than saturated 14:0 chains, a FID response correction factor for 7(Z)-3-hydroxy-7-tetradecenoic acid (14:1 3-OH FA) was applied when calculating the area of peaks consisting of co-eluting 14:1 and 14:0 3-OH FAs (numbered peak 18 in Fig. 2.2) in sample GC-FID traces. The molecular weight of (Z)-3-hydroxy-7-tetradecenoic acid was used with these corrected peak area values to determine the relative quantity (in nmol / unit OD) of 14-carbon 3-hydroxy fatty acids in spent media samples from \(E. coli\) expressing ALT enzymes.
Figure 2.3. Identification of 3-hydroxy fatty acids in lipid extracts from the culture media of K27(DE3)E. coli expressing ALT enzymes. When analyzed by GC-FID, lipid extracts from E. coli expressing several ALTs produced a peak that could not be conclusively identified by GC-MS or
comparison to retention standards. This peak is labelled with an arrow in the GC-FID chromatogram of fatty acids extracted from the culture media of VvALT1-expressing E. coli. Conversion of these fatty acids to methyl esters (VvALT1 FAMES) caused the unknown peak to split into two (peaks labelled A and B), indicating that it consisted of co-eluting compounds. GC-MS analysis demonstrated that Peak B corresponded to a 3-hydroxy fatty acid due to a characteristic m/z = 103 base peak, while the identity of Peak A still could not be determined. Silylation of lipid extracts from VvALT1-expressing E. coli allowed for both compounds to be conclusively identified by GC-MS. Peak A was identified as 3-hydroxytetradecanoic acid (3-OH 14:1 FA) based on the molecular ion present in the mass spectrum of its 2-TMS derivative (m/z = 386) and fragments characteristic of the TMS derivatives of both a 3-hydroxy fatty acid and a monounsaturated fatty acid. Peak B was confirmed as 3-hydroxytetradecanoic acid (3-OH 14:1 FA) by comparison to standard spectra from the NIST 17 Mass Spectral Library.

2.4.3 –ALT enzymes can be grouped according to their catalytic preferences in E. coli

The substrate preferences of ALTs investigated in this study varied in terms of chain length, oxidation state, and saturation level, both as compared to one another and to previously characterized ALTs (Yu et al., 2010; Pulsifer et al., 2014). All ALTs could be categorized into groups based on their catalytic preferences, with each group consisting of enzymes from diverse plant species and taxa. Group 1 contained enzymes that primarily generated 14-carbon β-ketofatty acids (which are converted to 13-carbon methylketones during the sample treatment process for detection by GC) and 14-carbon 3-hydroxy fatty acids, with a wide range of secondary products. Group 2 was made up of ALT enzymes that showed preference for 6-10 carbon fatty acyl- and β-keto fatty acyl-ACPs. Group 3 consisted of ALTs that mainly generated 12- and 14- carbon fatty acids, and Group 4 comprised enzymes that produced low, but statistically significant amounts of 16-carbon fatty acids as their primary products. Group 1 consisted of BdALT1/2, CsALT1, GmALT1, MtALT2, OsALT1, SmALT1, VvALT1, ZmALT2, and AtALT3. Group 2 consisted of three ALTs, MtALT1, ZmALT3, and AtALT4, Group 3 included only ZmALT1 and AtALT1, and Group 4 was comprised of CrALT1, GbALT1, and ZmALT4. Despite generating the same primary products, enzymes within each
functional group displayed significant variation in terms of the quantities of lipids and types of minor products generated when expressed in *E. coli*, as described in detail below.

The Group 1 ALTs were largest in number, and exhibited the broadest substrate specificity. They had preference for 14-carbon β-ketoacyl- and 3-hydroxyacyl-ACPs, but also generated a wide variety of secondary products ranging from 8-16 carbons in chain length (Figs. 2.4, 2.5). Of the Group 1 ALTs, *Bd*ALT2, *Cs*ALT1, *Os*ALT1, *Sm*ALT1, and *Vv*ALT1 favoured 14:1 β-keto acyl-ACP, while *Mt*ALT2 and *Bd*ALT1 had stronger affinity for 14-carbon 3-hydroxy fatty acyl-ACPs (Fig. 2.4). *Zm*ALT2 produced 14-carbon β-keto and 3-hydroxy FAs in similar amounts (3.06 ± 0.33 nmol / unit OD and 3.35 ± 0.17 nmol / unit OD respectively). *Bd*ALT1 and *Vv*ALT1 exhibited the greatest productivity in *E. coli* (Figs 2.4, 2.5). Cultures expressing *Bd*ALT1 generated, on average, approximately 8.56 ± 0.57 nmol/unit OD of C14 3-hydroxy FAs, 6.50 ± 0.54 nmol/unit OD 14:1 β-keto FA (converted to 13:1 MK), and 2.15 ± 0.27 nmol/unit OD 14:1 FA. Meanwhile, bacteria expressing *Vv*ALT1 generated 8.22 ± 0.48 nmol / unit OD C14 3-hydroxy FAs, 9.33 ± 0.61 nmol/unit OD 14:1 β-keto FA, and 6.08 ± 0.63 nmol / unit OD 14:1 FA (Figs. 2.4, 2.5).

ALT enzymes in this group also exhibited significant variation in their secondary and minor products when expressed in *E. coli* (Figs. 2.4, 2.5). Both *Bd*ALT1 and *Vv*ALT1 generated 12:1 FA, 14:0 β-keto FA, 3-hydroxy 12:0 FA, and 16:1 FA (Figs. 2.4, 2.5, Appendix 4). Although the activity profile of *Bd*ALT2 was similar to *Bd*ALT1, cells expressing *Bd*ALT2 accumulated proportionally larger quantities of fully reduced fatty acids. *Bd*ALT2-expressing bacteria also generated 8:0 FA and 8:0 β-keto FA as secondary products, while producing negligible quantities of C14 3-hydroxy FAs (Figs. 2.4, 2.5). Interestingly, *Mt*ALT2 produced 14:1 β-keto FA and 14:1 FA in nearly identical quantities (1.61 ± 0.23 nmol / unit OD and 1.42
± 0.27 nmol/unit OD, respectively) as secondary products, and similar to *BdALT2*, also generated some 8:0 FA / \(\beta\)-keto FA (Figs. 2.2, 2.4, 2.5). Minor products common to *CsALT1*, *OsALT1*, *SmALT1*, *ZmALT2*, and *AtALT3* included 8:0, 12:1, and 16:1 \(\beta\)-keto FAs (Fig. 3B). *AtALT3* differed from the other Group 1 enzymes in that *AtALT3*-expressing *E. coli* did not produce 14:1 FA in significant quantities; bacteria expressing this enzyme instead generated 1.71 ± 0.46 nmol/unit OD 12:0 \(\beta\)-keto FA (converted to 11:0 MK) as a secondary product.

*GmALT1* from soybean (*Glycine max*) initially showed no apparent thioesterase activity when expressed in K27(DE3) *E. coli*. The inactivity of our *GmALT1* expression construct was later traced to the use of the coding sequence from a splice variant that is apparently inactive, but nonetheless highly expressed in plant tissues according to transcriptomic data (Severin *et al.*, 2010). Tran *et al.* (2019) identified and characterized an active *GmALT1* isoform through heterologous expression in *E. coli* bacteria. The most abundant product of this active *GmALT1* isoform was 14:1 \(\beta\)-keto fatty acid (converted to 13:1 MK during sample treatment), therefore, *GmALT1* can be considered a Group 1 ALT enzyme (Tran *et al.*, 2019).

The major products of the Group 2 thioesterases (*MtALT1*, *ZmALT3*, and *AtALT4*) ranged from 6-10 carbons in chain length. *E. coli* expressing *AtALT4* generated 6:0 FA (1.36 ± 0.14 nmol/unit OD) and 8:0 FA (2.42 ± 0.15 nmol/unit OD), as previously reported by Pulsifer *et al.* (2014) (Figs. 2.4, 2.5, Appendix 4). Expression of *ZmALT3* resulted in the production of 2.43 ± 0.31 ng/unit OD of 8:0 FA, while expression of *MtALT1* generated more than three times this amount, at 8.48 ± 0.34 nmol/unit OD (Figs. 2.4, 2.5). Secondary products of *MtALT1* expression were 8:0, 10:0, and 14:1 \(\beta\)-keto FAs, as well as 6:0 FA (Figs. 2.4, 2.5). *MtALT1* was the only member of the Group 2 enzymes to generate a product over 10 carbons in chain length. *ZmALT3* was one of the most catalytically specific ALT enzymes characterized in this
experiment, only generating 8:0 fatty acyl-ACP in statistically significant quantities. Nearly all other ALTs studied produced significant amounts of more than one fully and/or partially reduced fatty acid species in *E. coli*.

*AtALT1* and *ZmALT1*, which made up Group 3, displayed strong activity toward 12:0 and 14:1 fatty acyl-ACP substrates (Fig. 2.4). These enzymes also generated 12:0 β-keto FA as a secondary product, as well as a broad range of minor fatty and β-keto fatty acid products (Figs. 2.4, 2.5). *ZmALT1*-expressing *E. coli* produced much larger quantities of lipids than *AtALT1*-expressing bacteria; expression of *AtALT1* generated 4.22 ± 0.69 nmol/unit OD 12:0 FA and 2.53 ± 0.56 nmol/unit OD 14:1 FA, while *ZmALT1*-expressing bacteria produced 17.15 ± 1.27 nmol/unit OD 12:0 FA and 21.8 ± 1.23 nmol/unit OD 14:1 FA. *AtALT1*- and *ZmALT1*-expressing cultures also greatly resembled one another in terms of their minor product profiles, generating these compounds in similar proportions relative to total lipid production, as shown in Figs. 2.4 and 2.5. However, some differences were still observed. *AtALT1* produced significant quantities of 8:0 FA and 12:0 3-OH FA in *E. coli*, while *ZmALT1* generated significant amounts of C14 3-OH FAs and 16:1 β-keto FA in bacteria (Figs. 2.4-2.5).

The Group 4 ALTs, which include *ZmALT4*, *CrALT1*, and *GbALT1*, displayed much lower thioesterase activity in *E. coli* than other ALT enzymes characterized in this study. *E. coli* strains expressing *CrALT1* and *ZmALT4* generated small (0.26-0.32 nmol/OD) but statistically significant quantities of 16:1 FA (Figs. 2.4, 2.5). *CrALT1* was the only ALT to produce 16:0 FA in significant amounts as compared to *E. coli* containing an empty vector. No FAs or β-keto FAs of other chain lengths were observed as products of *ZmALT4* and *CrALT1* (Figs. 2.4, 2.5). In contrast, *GbALT1* expression generated nearly three times as much 16:1 FA (0.77 ± 0.13 nmol/OD), and this enzyme acted on a wider range of acyl-ACP substrates, with
secondary products including 8:0 and 14:1 FAs, 14:0 and 14:1 β-keto FAs, and 16:0 and 16:1 β-keto FAs (Figs. 2.4, 2.5).
Fig. 2.4. Quantification of fully reduced and 3-hydroxy fatty acids secreted into the media of K27(DE3) E. coli cultures carrying an empty pET-28a vector, or expressing ALT enzymes. Free fatty acids were quantified by comparison of peak areas to a tetracosane (C24 alkane) internal standard using GC-FID.
Statistically significant differences in the amount of fatty acids in culture media of bacteria expressing acyl-lipid thioesterases as compared to bacteria carrying an empty pET-28a vector were determined by a two-tailed Student's t-test \( (n = 3, \ p < 0.001) \). Statistically significant quantities are marked with an asterisk (*). ALT enzymes are grouped based on their substrate chain-length and oxidation state preferences when expressed in \textit{E. coli}, as follows: Group 1 – preference for C14 β-ketoacyl- and / or 3-hydroxyacyl-ACPs, Group 2 – preference for C6-10 fatty acyl- and / or β-ketoacyl-ACPs, Group 3 – preference for C12-C14 fatty acyl-ACPs, Group 4 – preference for C16 fatty acyl-ACPs.
Fig. 2.5. Quantification of methylketones derived from β-keto fatty acids secreted into the media of K27(DE3) *E. coli* cultures carrying an empty pET-28a vector, or expressing ALT enzymes. β-keto fatty acids produced by ALTs were decarboxylated to stable methylketones via treatment of culture media with heat and acid prior to extraction of lipids and analysis by gas chromatography. Methylketones were quantified by comparison of peak areas to a tetracosane (C24 alkane) internal standard using GC-FID.
Statistically significant differences in the amount of methylketones culture media of bacteria expressing acyl-lipid thioesterases as compared to bacteria carrying an empty pET-28a vector were determined by a two-tailed Student's t-test \((n = 3, p < 0.001)\). Statistically significant quantities are marked with an asterisk (*). ALT enzymes are grouped based on their substrate chain-length and oxidation state preferences when expressed in \textit{E. coli}, as follows: Group 1 – preference for C14 β-ketoacyl- and / or 3-hydroxyacyl-ACPs, Group 2 – preference for C6-10 fatty acyl- and / or β-ketoacyl-ACPs, Group 3 – preference for C12-C14 fatty acyl-ACPs, Group 4 – preference for C16 fatty acyl-ACPs.

2.4.4 – ALT proteins are expressed at different levels in \textit{E. coli}

All ALT proteins were successfully expressed in K27(DE3) \textit{E. coli}, as shown by immunoblotting analysis of cell lysates from ALT-expressing bacterial strains (Fig. 2.6). Following lysis of bacterial cells by sonication, the crude lysates were centrifuged to separate the soluble and insoluble fractions. Immunoblotting analysis was performed with each fraction to determine whether the recombinant ALT proteins were aggregating into insoluble inclusion bodies. Target ALT protein bands were only observed within the soluble fraction of lysate samples, indicating that after the incubation period for induced cultures, the majority of the recombinant ALT enzyme likely still existed in a bioactive form rather than aggregating in inclusion bodies in K27(DE3) cells (data not shown). Quantification of ALT protein band intensity with respect to that of a GAPDH loading control revealed that the expression levels of ALTs in \textit{E. coli} varied significantly (Fig. 2.6, Table 2.1). While certain ALTs that generated large quantities of lipid in \textit{E. coli}, such as ZmALT1 and VvALT1, had low expression levels as compared to less active ALTs like GbALT1 and CrALT1, no overall correlation was ultimately found between the expression levels of ALT proteins and the total amount of medium-chain lipids (in ng/unit OD) produced by bacterial cultures expressing them (Table 2.1). This suggests that variations in the amounts of FAs and β-keto FAs produced by ALT-expressing \textit{E. coli} cultures result mainly from differences in the specific activities of ALT enzymes towards their
endogenous acyl-ACP substrates, rather than differences in protein expression levels. However, the relative expression levels of ALTs from monocots (ZmALT1-4, BdALT1-2, OsALT1) was generally much lower as compared to eudicot ALTs that produced similar quantities of lipids, indicating that sequence features specific to ALT genes from monocot plants may inhibit protein expression and/or accumulation in E. coli (Table 1).

**Figure 2.6.** Immunoblots of clarified cell lysates of (K27)DE3 E. coli expressing ALT enzymes. Twenty μg of total protein was loaded into each lane. GAPDH was used as a loading control. The membrane was probed with HRP-conjugated anti-GAPDH and anti-His tag antibodies. Images were captured using a Vilber Lourmat Chemi-Smart 3000 imaging station. GAPDH bands were visualized at 2 s exposure, while ALT protein bands were visualized at 15 s exposure. Molecular weights were estimated based on a Precision Plus Dual Colour Protein Ladder (Bio-Rad)
Table 2.1. Relative expression levels of ALT enzymes with respect to total lipid production in (K27)DE3 E. coli. Relative expression levels were determined via Western blotting by comparing ALT protein band intensity to that of a GAPDH loading control. Total lipid production was measured by GC-FID analysis of spent culture media (n = 3).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Relative ALT band intensity</th>
<th>Total lipid production in E. coli (μg / unit OD)</th>
<th>( \text{total lipid production} \ \text{relative protein expression} ) (μg / OD / relative ALT band intensity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AtALT1 )</td>
<td>2.57</td>
<td>3.29 ± 0.23</td>
<td>1.28 ± 0.09</td>
</tr>
<tr>
<td>( AtALT3 )</td>
<td>0.27</td>
<td>1.86 ± 0.18</td>
<td>6.89 ± 0.67</td>
</tr>
<tr>
<td>( AtALT4 )</td>
<td>1.17</td>
<td>1.01 ± 0.05</td>
<td>0.86 ± 0.05</td>
</tr>
<tr>
<td>( MtALT1 )</td>
<td>2.31</td>
<td>2.64 ± 0.14</td>
<td>1.15 ± 0.06</td>
</tr>
<tr>
<td>( MtALT2 )</td>
<td>2.73</td>
<td>2.10 ± 0.14</td>
<td>0.77 ± 0.05</td>
</tr>
<tr>
<td>( VvALT1 )</td>
<td>0.30</td>
<td>7.30 ± 0.03</td>
<td>24.34 ± 1.07</td>
</tr>
<tr>
<td>( CsALT1 )</td>
<td>2.62</td>
<td>1.86 ± 0.05</td>
<td>0.71 ± 0.15</td>
</tr>
<tr>
<td>( GmALT1 )</td>
<td>3.84</td>
<td>0.28 ± 0.08</td>
<td>0.73 ± 0.02</td>
</tr>
<tr>
<td>( CrALT1 )</td>
<td>2.74</td>
<td>0.32 ± 0.05</td>
<td>0.12 ± 0.02</td>
</tr>
<tr>
<td>( SmALT1 )</td>
<td>2.26</td>
<td>1.40 ± 0.06</td>
<td>0.62 ± 28.47</td>
</tr>
<tr>
<td>( GbALT1 )</td>
<td>0.31</td>
<td>0.78 ± 0.05</td>
<td>0.25 ± 0.16</td>
</tr>
<tr>
<td>( BdALT1 )</td>
<td>0.09</td>
<td>3.95 ± 0.16</td>
<td>43.91 ± 1.75</td>
</tr>
<tr>
<td>( BdALT2 )</td>
<td>0.17</td>
<td>2.61 ± 0.12</td>
<td>15.36 ± 0.71</td>
</tr>
<tr>
<td>( OsALT1 )</td>
<td>0.28</td>
<td>1.53 ± 0.12</td>
<td>5.46 ± 0.44</td>
</tr>
<tr>
<td>( ZmALT1 )</td>
<td>0.06</td>
<td>11.96 ± 0.40</td>
<td>199.40 ± 6.73</td>
</tr>
<tr>
<td>( ZmALT2 )</td>
<td>0.52</td>
<td>1.85 ± 0.11</td>
<td>3.55 ± 0.21</td>
</tr>
<tr>
<td>( ZmALT3 )</td>
<td>0.94</td>
<td>0.87 ± 0.07</td>
<td>0.92 ± 0.07</td>
</tr>
<tr>
<td>( ZmALT4 )</td>
<td>0.65</td>
<td>0.24 ± 0.03</td>
<td>0.37 ± 0.04</td>
</tr>
</tbody>
</table>

2.4.5 – ALT enzymes with exclusive preference for 6-10 carbon substrates appear to differ significantly from other ALTs in their tertiary structure

Three-dimensional homology models of the hot dog-fold domains of the ALT enzymes characterized in this study were constructed using the single hot dog-fold thioesterase YbgC from E. coli as a template (PDB ID: 1S5Ua), and a structural alignment of these models was
used to generate a phylogenetic tree. The topology of this structure-based tree was reflective of the catalytic groupings described in Section 2.4.2, and contrasts significantly from that of a sequence-based phylogenetic tree, in which ALTs are grouped by their occurrence in plant species and families (Figs. 2.1, 2.4, 2.5, 2.7). Most notably, the Group 2 ALTs (MtALT1, AtALT2, ZmALT3) are in a separate clade from the other ALTs characterized in this study. Structural alignment data shows that the tertiary structure of Group 2 ALTs diverges from the alignment consensus structure overall, particularly in regions spanning helices α1-α2 (amino acid residues 44-53 of the thioesterase domain), strands β4 – β5 of the β-sheet (aa78-98), and β8 – α3 at the C-terminus (aa122 onward) (Fig. 2.7). The Group 4 ALTs ZmALT4 and CrALT1, which exclusively produce 16-carbon fatty acids, occupy adjacent taxa, as do the Group 1 ALTs SmALT1, AtALT3, VvALT1, MtALT2, and ZmALT2. Both ZmALT4 and CrALT4 possess amino acid insertions in the turn region spanning β6 – β7 (aa94-98).

ALTs are ‘difficult’ proteins to construct homology models for, as no experimentally validated protein structures with over 30% sequence identity to ALTs are available as modelling templates (Forrest et al., 2006; Dolan et al., 2012). Therefore, it must be considered that some features of the structure-based phylogenetic tree may have resulted from variations in model quality among the modelled proteins. Improvements to ALT model quality using technologies that became available following publication of this study, and information learned from these improved models, are described in Chapter 4 of this thesis.
Figure 2.7. **Left:** Unrooted structure-based phylogenetic tree of ALT proteins. Homology models of ALTs were generated by SWISS-MODEL v.8.05, using the *E. coli* single hot dog-fold YbgC thioesterase as a template (PDB Accession 1S5Ua). Structures were aligned with STAMP (Russell & Barton, 1992), and a phylogenetic tree was generated using root-mean-square distances between aligned structures. Labels are coloured according to ALT substrate specificity in *E. coli:* Group 1 ALTs (blue) primarily generate 14-carbon β-keto/3-hydroxy fatty acids, Group 2 ALTs (orange) produce 6-10 carbon fatty/β-keto fatty acids, Group 3 ALTs (yellow) mainly generate 12-14 carbon fatty acids, and Group 4 ALTs (green) produce 16-carbon fatty acids. *At =* Arabidopsis thaliana, *Bd =* Brachypodium distachyon, *Cr =* Chlamydomonas reinhardtii, *Cs =* Cannabis sativa, *Gb =* Ginkgo biloba, *Gm =* Glycine max, *Mt =* Medicago truncatula, *Os =* Oryza sativa subsp. indica, *Sm =* Selaginella moellendorfii, *Vv =* Vitis vinifera, *Zm =* Zea mays. **Right:** Homology models of the hot dog-fold domain of the Group 2 ALT *ZmALT3,* and the Group 3 ALT *ZmALT1.* Regions of significant structural dissimilarity between Group 2 ALTs and ALTs from other catalytic groupings were identified at α-helices α1-α2 (orange), β-strands β5 – β6 (blue) and β8-α3 at the C-terminal end of the hot dog-fold thioesterase domain (pink).
2.5 – DISCUSSION

2.5.1 - ALTs have highly diverse substrate specificities

Screening the substrate specificities of 15 ALT enzymes from across the plant kingdom revealed that these enzymes have highly diverse substrate preferences. The selected thioesterases could be divided into four groups based on their primary products. Group 1 ALTs primarily generated 14:1 \( \beta \)-keto or 3-hydroxy FAs with a wide range of minor products, Group 2 ALTs showed preference for 6-10 carbon fatty and \( \beta \)-ketofatty acid chains, Group 3 ALTs predominantly produced 12- and 14-carbon FAs, and Group 4 ALTs mainly produced 16-carbon FAs. As predicted, previously characterized Arabidopsis ALTs and ShMKS2 do not represent the full range of catalytic diversity that exists among ALT enzymes. AtALT1-4 and ShMKS2 can be classified in Groups 1-3 based on their primary products, but ALTs with preference for 16-carbon FAs (CrALT1, GbALT1, and ZmALT4) are identified in this study.

All catalytic groups consisted of members from diverse plant taxa, and if more than one ALT from a plant species was profiled, these were nearly always classified into different catalytic groups. Notably, each of the four catalytic groups included one of the four Z. mays ALTs, and B. distachyon was the only plant species to possess more than one ALT enzyme belonging to the same group (Figs. 2.4, 2.5). Based on the enzymes profiled here, the catalytic diversity of ALTs within species is greater than the diversity observed between species and taxa. ALTs from distantly related plant species that share similar substrate specificity likely have similar biological roles, while the varying substrate specificities of ALTs from the same species is an indication that these enzymes carry out distinct biological functions.

No two ALTs shared the exact same activity profile in E. coli. Despite having the same
primary products, all ALTs within each catalytic group differed in the secondary products they generated or the relative quantities of lipids they produced. Given such broad catalytic diversity within the cross-section of ALT enzymes characterized here, continuing to screen the substrate specificities of ALTs from other plant species will likely reveal an even greater extent of functional diversity within this enzyme family. However, since a heterologous in vivo system was used to screen the substrate preferences of ALTs, the possibility must be considered that certain features of their activity profiles may be artifacts resulting from the abundance of specific substrates in the endogenous E. coli acyl-ACP pool, or the effects of heterologous ACP structure on thioesterase substrate selectivity (Cronan & Thomas, 2008; Yang et al., 2023). The structural conformation of acyl carrier protein has been shown to exert influence on acyl-ACP thioesterase substrate selectivity, independently of thioesterase structure (Yang et al., 2023). The medium-chain FATB-type thioesterase ChFATB2 from Cuphea hookeriana, for example, displays preference for C8:0 acyl chains when interacting with acyl-ACPs of plant origin, but its chain-length selectivity shifts towards C10:0 acyl chains when encountering with E. coli acyl-ACPs (Yang et al., 2023). This has been attributed to structural differences in non-conservative, flexible loop regions of bacterial and plant ACPs that affect the binding affinity of acyl-ACPs bearing various chain lengths to the thioesterase (Yang et al., 2023). Substrate abundance or the effects of heterologous ACP structure on ALT function may also account for the apparent lack of correlation between the expression levels of ALT proteins and the total quantities of lipids they produce when expressed in E. coli (Fig. 2.6, Table 1). While an in vitro enzyme activity assay would eliminate the possibility of these artifacts, synthesizing the full range of acyl-ACP substrates required for such an assay would be prohibitively costly.

The structure-based alignment and phylogenetic tree created from three-dimensional
models of the ALT enzymes characterized in this study revealed some information regarding possible relationships between the catalytic domain sequence and substrate specificity of ALTs. The Group 2 ALTs ($At$ALT4, $Mt$ALT1, $Zm$ALT3) diverged the most from the consensus structure, occupying a separate clade from other ALTs. They especially differed from the consensus structure in the second and third alpha-helices of the hot dog-fold thioesterase domain, as well as at amino acid residues 78-98 of the thioesterase domain, spanning strands $\beta 4$ – $\beta 5$ of the $\beta$-sheet (Fig. 2.7). The latter region was later found to exert significant effects on ALT chain-length selectivity, and is at least partially responsible for the Group 2 ALTs’ preference for 6-10 carbon acyl-ACP substrates (see Chapter 4) (Kalinger et al., 2023).

The Group 4 ALTs $Cr$ALT1 and $Zm$ALT4, which produce 16-carbon FAs only, possess an amino acid insertion in the turn region between their sixth and seventh beta-strands (Appendix 1). $Gb$ALT1, the third Group 4 ALT, does not possess an insertion in this region, however, $Gb$ALT1 also generates numerous secondary products in addition to 16-carbon FAs (Figs. 2.4, 2.5, Appendix 1). It is therefore possible that the presence of an amino acid insertion in this region precludes ALTs from acting on substrates other than 16:0 and 16:1 acyl-ACPs. Tertiary structure-based phylogenetic analysis yielded no concrete information regarding the substrate preferences of the Group 1 and Group 3 ALTs. The identification of amino acid motifs that influence an ALT’s preference for fully reduced or partially reduced acyl-ACP substrates, and dictate specificity for 12-14 carbon chain-lengths, is described in Chapter 4 of this thesis.

2.5.2 – ALT enzymes can act on 3-hydroxyacyl-ACP substrates

Heterologous expression of Group 1 and Group 3 ALTs resulted in the production of novel compounds, including 3-hydroxydodecanoic acid (3-OH 12:0 FA), (Z)-3-hydroxy-7-
tetradecenoic acid (3-OH 14:1 FA), and 3-hydroxytetradecanoic acid (3-OH 14:0 FA) (Figs. 2.2, 2.3, 2.4, Appendix 4). Certain ALTs, such as MtALT2 and BdALT1, generate these novel 3-hydroxy fatty acids as their primary products. Surprisingly, the data presented here showed that 12- and 14-carbon 3-OH FAs were also present in culture media from *E. coli* expressing the previously characterized *Arabidopsis* ALTs AtALT1 and AtALT3, respectively. As 3-OH FAs had not been previously reported as a product of *Arabidopsis* ALTs (Pulsifer *et al.*, 2014), 3-OH FAs detected in extracts from ALT-expressing bacteria were initially thought to arise from unwanted reduction of β-keto FAs in culture media during the sample treatment process. However, if this were true, a direct relationship between the amounts of β-keto FAs and the corresponding 3-OH FAs in culture media would be expected, and no such relationship was observed. For example, *AtALT3* and *GbALT1* both produce significant quantities of 14:0 β-keto FA (converted to 13:0 MK), but negligible 3-OH 14:0 FA, and 12-carbon 3-OH FAs were not detected as products of *CsALT1*, *SmALT1*, *BdALT2*, nor *ZmALT1*, which all generated significant quantities of 12-carbon β-keto fatty acids (converted to 11-carbon MKs) in *E. coli* (Figs. 2.3, 2.4, 2.5, Appendix 4). Therefore, the 3-OH FAs observed in the product profiles of certain ALTs are likely direct enzymatic products. This indicates that, at least when expressed in *E. coli*, some ALTs are capable of cleaving the thioester bond of the 3-hydroxy fatty acyl-ACP intermediate in the fatty acid elongation cycle, thereby releasing a free 3-hydroxy fatty acid (Janßen & Steinbüchel, 2014).

### 2.5.3 – Evolutionary origins of ALTs

In a phylogenetic tree of *Arabidopsis* ALTs, *AtALT3* is the first to diverge from a common ancestor. Given the widespread expression and broad substrate specificity of *AtALT3*, Pulsifer *et al.* (2014) hypothesized that ALT enzymes descended from an ancient multifunctional
enzyme, and that \( \text{AtALT1/2/4} \), which have more specific catalytic preferences, evolved through sub-functionalization. However, the results of the catalytic activity screen reported here suggest otherwise. Of the three ALTs from early-diverging plants that were investigated in this study, both \( \text{CrALT1} \) and \( \text{GbALT1} \) predominantly generated 16-carbon FAs. This suggests that the ALT enzyme family originated from an ancestral enzyme with preference for 16-carbon substrates. Preference for 16-carbon acyl-ACPs is associated with thioesterases involved in primary lipid metabolism, as palmitic acid is the most common constituent of plant lipids (Jones et al., 1995). ALT enzymes potentially descended from an ancestral, single hot dog-fold palmitoyl-ACP thioesterase, which later evolved specificity for medium-chain acyl-ACPs, and for \( \beta \)-keto and 3-hydroxy fatty acyl-ACPs.

2.5.4 – Potential biological roles of ALT enzymes

The medium-chain fatty and \( \beta \)-keto fatty acid metabolites produced by ALT enzymes have insecticidal and antimicrobial properties, and likely function in plant defence (Ben-Israel et al., 2009; Fridman et al., 2005; Pulsifer et al., 2014). In Arabidopsis, \( \text{AtALT3} \) is expressed in all plant tissues, where the wide range of \( \beta \)-ketofatty acids it produces are possibly involved in plant-wide defence (Pulsifer et al., 2014). The Group 1 ALTs identified in this study may also function in plant-wide insect defense given their broad substrate specificities. It is worth noting that the single ALT genes of \( \text{V. vinifera} \) and \( \text{G. max} \) also encode for Group 1 ALTs with broad, plant-wide expression patterns according to publicly available transcriptomic data (Fasoli et al., 2012; Severin et al., 2010; Libault et al., 2010; Tran et al., 2019). This suggests that in plants that possess only one ALT gene, the encoded enzyme plays a broad role in plant defence by producing a wide range of volatile compounds, while plants with multiple ALTs have evolved paralogs to fulfill specific roles.
The Group 2 ALTs (\textit{AtALT4}, \textit{MtALT1}, \textit{ZmALT3}) act on a much narrower range of substrates, pointing towards more specialized biological roles. \textit{ZmALT3} was one of the most catalytically specific ALTs, producing only 8:0 FA, indicating that it likely has a highly specialized function. \textit{AtALT4} is exclusively expressed in anthers, where the volatile 6-8 carbon FAs it generates may serve to attract insect pollinators (Pulsifer \textit{et al.}, 2014). Transcriptomic data also localizes \textit{ZmALT3} to floral anthers and silks, and \textit{MtALT1} to flowers (Sekhon \textit{et al.}, 2011; He \textit{et al.}, 2009). While \textit{A. thaliana}, \textit{M. truncatula}, and \textit{Z. mays} can easily reproduce through wind- and self-pollination and would not necessarily need to produce chemicals that signal pollinator species, bee populations living near agricultural fields will frequently forage maize pollen (Danner \textit{et al.}, 2014; Wheelock & O’Neal, 2016; Severson & Parry, 1981), and there are also recorded observations of various pollinating insects visiting the flowers of wild \textit{A. thaliana} and \textit{M. truncatula} populations (Chen \textit{et al.}, 2003; Snape & Lawrence, 1971; Jullien \textit{et al.}, 2021). Since close taxonomic relatives of \textit{A. thaliana} and \textit{M. truncatula}, such as \textit{Arabidopsis lyrata} and \textit{Medicago sativa} (alfalfa), must rely on outcrossing through insect pollination to set seed, the ability to produce pollinator-attracting volatiles is likely evolutionarily conserved within these genera (Nasrallah, 2000; Chen & Zuo, 2018; Chen \textit{et al.}, 2003). Profiling of volatile emissions from \textit{Arabidopsis thaliana} flowers has demonstrated that \textit{A. thaliana} produces a blend of volatile floral terpenes that are potential signalling cues for pollinating insects, which supports this statement (Chen \textit{et al.}, 2003).

It is also possible that the medium-chain fatty and β-keto fatty acids generated by the Group 2 ALTs might instead serve to attract beneficial insect bodyguards that prey on pest species, or guard floral parts against infection by pathogenic fungi or bacteria (Chen \textit{et al.}, 2003). \textit{AtALT2}, which was not profiled in this study as its products overlap with \textit{AtALT3}, is a
Group 2 ALT as it produces 8-10 β-keto FAs. AtALT2 is expressed in suberized tissues, specifically root endodermis and periderm (Pulsifer et al., 2014). Its expression additionally co-regulates with several genes belonging to the Pathogenesis-Related 10 (PR-10) family, which are often induced in response to biotic and abiotic stress conditions (reviewed in Fernandes et al., 2013). The products of AtALT2 could provide further protection against insect herbivory and pathogen attack in suberin-fortified plant tissues.

*Arabidopsis* ALT1 has been speculated to produce FAs that accumulate in cuticle, as it is expressed in stem and leaf epidermal cells and co-regulates with genes involved in cuticular wax biosynthesis (Pulsifer et al., 2014). However, the other Group 3 ALT, ZmALT1, has a much broader range of lipid products including β-keto and 3-hydroxy fatty acids, and is localized mainly to floral anthers and silks, similarly to ZmALT3. As Group 4 ALTs may be representative of the ancestral state of ALTs, they may not occupy highly specialized biological roles despite their narrow substrate specificity.

2.5.5 – *E. coli* is a useful platform for ALT protein expression

Existing data at the time this study was conducted suggested that the lipid products generated from expression of ALT enzymes in *E. coli* would likely be reflective of their products *in planta* in terms of oxidation state and chain length. For example, ShMKS2, an ALT from wild tomato, is known to generate 14:0 and 12:0 β-keto FAs in both *E. coli* and in leaf trichomes where it is naturally expressed. Expression in *E. coli* was also used to accurately predict the product range of the medium-chain-specific ChFATB2 from *Cuphea hookeriana* (Dehesh et al., 1996). However, the products of acyl-ACP thioesterase expression in *E. coli* would necessarily differ from those *in planta* in terms of saturation level. These differences
stem from the availability of unsaturated acyl-ACPs in *E. coli* as compared to plants. For instance, the primary product of the FATB thioesterase from *A. thaliana* is 16:0 FA *in planta*, and 16:1 FA in *E. coli* (Mayer & Shanklin, 2007). Desaturases in *E. coli* can act on acyl-ACPs of 10 carbons or more, while plant desaturases can only act on acyl-ACP chains comprising 16 or more carbons (Feng & Cronan, 2009; Thelen & Ohlrogge, 2002). Therefore, fatty and β-keto fatty acids shorter than 16 carbons are not found *in planta*. ALTs that produce 12- and 14-carbon monounsaturated fatty/β-keto fatty acids in *E. coli*, notably the Group 1 and 3 thioesterases, would likely generate the corresponding saturated fatty/β-keto fatty acids *in planta*. Chapter 3 of this thesis provides more information on how the substrate specificities of ALTs *in planta* correlate with their activity in *E. coli*.

Although the product profiles of ALTs in *E. coli* might not perfectly match those *in planta* due to the differing activities of available desaturases, and the influence of heterologous ACP structure on thioesterase activity as described above in Section 2.5.1, *E. coli* is still a useful platform for ALT expression. Monounsaturated 11-15 carbon MKs and the derivatives of unsaturated 12-16 carbon FAs are ideal blending agents for diesel biofuels (Goh *et al.*, 2012; Lu *et al.*, 2008). *E. coli* strains have been engineered for the purpose of overproducing medium-chain FAs and β-keto FAs, such as those described by Liu *et al.* (2016) and Goh *et al.* (2012, 2014). While most FA-overproducing *E. coli* strains reported in the literature are engineered to favour fatty acyl-CoA substrates over fatty acyl-ACPs, metabolic flux could conceivably be adjusted to accomplish the reverse (Steen *et al.*, 2010). Expression of a highly active ALT, such as *ZmALT1*, in a fatty/β-ketofatty acyl-ACP overproducing *E. coli* strain could result in medium-chain FA or MK production in excess of currently achievable titers. In a 2012 study by Park *et al.*, expression of the ALT enzyme *ShMKS2* and its companion decarboxylase
ShMKS1 increased MK production levels in a β-keto FA overproducing *E. coli* strain well beyond the previously reported maximum at the time. It would be worthwhile to search the genomes of plant species investigated in this study for ALT-associated decarboxylases, express them alongside the ALTs profiled here, and assess MK levels in culture media without heat and acid treatment.

2.5.6 – *ALTs generate products of biotechnological significance*

In addition to biofuels, medium-chain FAs, 3-hydroxy FAs, and MKs are industrially valuable as components of insecticides, flavourings, fragrances, surfactants, and pharmaceutical precursors (Ohlrogge, 1994; Antonious *et al.*, 2003; Guo *et al.*, 2008). However, these compounds, particularly MKs and 3-hydroxy FAs, are often chemically synthesized via oxidation of hydrocarbon feedstocks from non-renewable fossil carbon sources (Park *et al.*, 2012, Yan *et al.*, 2020). The development of bio-renewable sources of medium-chain FAs and MKs is essential to sustainably manufacture them for industrial applications. Oilseed crops are widely used as renewable sources of certain medium-chain FAs; for example, 12:0 FA is abundant in coconut oil, while *Cuphea hookeriana* has high levels of 8:0 FA and 10:0 FA in its seeds (Dehesh *et al.*, 1996). However, no similarly suitable natural source of MKs has been discovered. Also, monounsaturated medium-chain FAs, 3-hydroxy FAs, or MKs cannot be easily produced in plant systems, necessitating the engineering of microbes.

Given their widespread substrate specificities across the entire range of saturated and monounsaturated medium-chain fatty and β-ketofatty acyl-ACPs, ALT enzymes are ideal candidates to be engineered as bio-renewable sources of medium-chain FAs and MKs. Of particular interest are the Group 1 ALTs, which act on a wide range of acyl-ACP substrates.
Multiple enzyme variants could be derived from Group 1 ALTs that are specific to one or a few of the fatty acid, β-keto fatty acid, or 3-hydroxy fatty acid species generated by the wild-type enzyme. Expression of plant thioesterases has already been employed to great success to overproduce free fatty acids in *E. coli* or alter its lipid profile (Lu *et al.*, 2008; Steen *et al.*, 2010; Voelker & Davies, 1994). A bacterial or yeast strain expressing an ALT that has been engineered to produce significant quantities of a particular medium-chain fatty or β-ketofatty acid would require fewer resources than cultivation of a plant for the same purpose.

An aim of this study was to identify connections between the hot dog-fold domain sequence and substrate specificity of ALT enzymes, however, only a few conclusions could be drawn from the data presented here. More insight into relationships between ALT protein sequence and substrate specificity was gained through targeted mutagenesis experiments, where protein sequence motifs were exchanged between ALT enzymes with opposite chain-length and oxidation-state preferences. This work is described in Chapter 4.

2.6 – CONCLUSIONS

This study demonstrates that the substrate specificities of ALT enzymes from wide-ranging plant taxa vary considerably in terms of chain length, oxidation state, and saturation level. Based on their primary fatty acid, β-keto fatty acid, and 3-hydroxy fatty acid products, ALTs can be classified into four groups that consist of members from diverse plant species and taxa. Each of the 15 thioesterases profiled in this study differed widely in the quantities of lipids and types of minor products they produced when expressed *E. coli*. The broad catalytic diversity of ALT enzymes indicates that they occupy specialized and diverse biological roles, however, analysis of their *in planta* activities and gene expression patterns is necessary to conclusively determine
their biological functions. The products of ALT enzymes are industrially valuable, as they are used in the manufacture of biofuels, insecticides, and pharmaceuticals. Given their functional diversity, ALTs have strong potential for metabolic engineering as biorenewable sources of valuable medium-chain fatty acids.
Chapter 3: Production of C6-C14 medium-chain fatty acids in seeds and leaves via overexpression of ALT-type acyl-ACP thioesterases

3.1 – ABSTRACT

ACYL-LIPID THIOESTERASE (ALT) enzymes can generate a wide range of medium-chain fatty acids, methylketone (MK) precursors, and 3-hydroxy fatty acids when expressed heterologously in *Escherichia coli* bacteria. While this makes ALT-type thioesterases attractive as metabolic engineering targets to increase production of high-value medium-chain fatty acids and MKs in plant systems, the behaviour of ALT enzymes *in planta* is not well understood. To profile the substrate specificities of ALT-type thioesterases in different plant tissue types, *AtALT1-4* from *Arabidopsis thaliana*, which have widely varied chain length and oxidation state preferences in *E. coli*, were overexpressed in *Arabidopsis* seeds, *Camelina sativa* seeds, and *Nicotiana benthamiana* leaves. Seed-specific overexpression of ALT enzymes led to medium-chain fatty acid accumulation in *Arabidopsis* and *Camelina* seed triacylglycerols, and transient overexpression in *N. benthamiana* demonstrated that the substrate preferences of ALT-type thioesterases *in planta* generally agree with those previously determined in *E. coli*. *AtALT1* and *AtALT4* overexpression in leaves and seeds resulted in the accumulation of 12–14 carbon-length fatty acids and 6–8 carbon-length fatty acids, respectively. While it was difficult to completely profile the products of ALT-type thioesterases that generate MK precursors (i.e. β-keto fatty acids), these results nonetheless show that ALT enzymes are catalytically diverse *in planta*. The knowledge gained from this study is an advancement towards being able to use ALT-type thioesterases as metabolic engineering tools to modify the fatty acid profiles of oilseed crops, other plants, and microorganisms.
3.2 – INTRODUCTION

Medium-chain fatty acids (MCFAs) and their chemical derivatives have commercial applications in the manufacture of biofuels, plasticizers, pharmaceuticals, detergents, and food additives, among other products (Dyer et al., 2008). However, the majority of medium-chain fatty acids used in industry are generated from fossil hydrocarbon feedstocks, or from crops such as coconut and oil palm, which can only be grown in tropical regions and are mainly cultivated through agricultural practices that harm local biodiversity (Phippen et al., 2006). Development of other, sustainable sources of MCFAs and their derivatives is therefore of importance, and much research has focused on engineering transgenic oilseed crops for this purpose (reviewed in Baud, 2018).

In plants, de novo fatty acid synthesis mostly occurs in plastids. Fatty acyl chain synthesis is terminated by thioesterase enzymes, which cleave the thioester bond linking the growing acyl chain to the phosphopantetheinate cofactor carried by acyl carrier protein (ACP) (reviewed in Ohlrogge & Jaworski, 1997). The plant acyl-ACP thioesterases thus far characterized belong to the hot dog-fold protein superfamily. Plants possess two major classes of acyl-ACP thioesterase: the Fat-type thioesterases, which have a double hot dog-fold thioesterase domain, and ACYL-LIPID THIOESTERASES, or ALT-type thioesterases, which consist of a single hot dog-fold domain (Ben-Israel et al., 2009; Mayer & Shanklin, 2007; Pulsifer et al., 2014; Yu et al., 2010). Fat-type thioesterases are ubiquitous as they function in primary lipid metabolism, and are divided into FatA and FatB subtypes, which predominantly release 18:1 and 16:0 fatty acids, respectively (Jones et al., 1995). In most oilseed crops, such as Brassica napus (e.g. canola), sunflower, and soybean, seed triacylglycerols consist almost entirely of fatty acyl chain-lengths greater than 16 carbons (Eccleston & Ohlrogge, 1998; Reske et al., 1997).
However, species that accumulate large quantities of MCFAs in their seed oil, such as *Cuphea spp.* and coconut, possess specialized FatB-type thioesterases that have specificity for medium-chain (8–14 carbon) fatty acyl-ACP substrates (Gunstone & Pollard, 2001; Graham & Kleiman, 1992; Dehesh *et al.*, 1996). Transgenic oilseed crops can be engineered to accumulate high levels of MCFAs via seed-specific overexpression of medium-chain specific FatB thioesterases, although MCFA content is lower in these genetically modified plants than in the species that naturally produce them (Dehesh *et al.*, 1996; Kim *et al.*, 2015; Tjellström *et al.*, 2013; Voelker *et al.*, 1996). Overexpression of specialized FatB enzymes has also been used to drive MCFA production in non-seed organs. In *Nicotiana benthamiana* plants engineered to store 15% of their fatty acid content in leaves as triacylglycerols (TAG), overexpression of medium-chain FatB-type thioesterases results in MCFAs accounting for 20–40 mol% of leaf TAG fatty acids (Reynolds *et al.*, 2015).

While few plant species possess specialized FatB-type thioesterases with activity towards medium-chain acyl-ACP substrates, ALT-type thioesterases have been identified in all major plant taxa (Kalinger *et al.*, 2018; see Chapter 2). ALT enzymes are medium-chain acyl-ACP thioesterases capable of interrupting fatty acid biosynthesis to generate fully reduced MCFAs, β-keto MCFAs, and 3-hydroxy MCFAs, and unlike FAT-type thioesterases, they are likely involved in secondary rather than primary lipid metabolism (Fridman *et al.*, 2005; Pulsifer *et al.*, 2014; Yu *et al.*, 2010; Zheng *et al.*, 2013). As described in the previous chapter, a survey of the substrate specificities of ALT-type thioesterases from a wide variety of plant and algal species in *Escherichia coli* indicated that ALT-type thioesterases are catalytically diverse. A single plant species will often possess two to four paralogous ALTs, which typically have
widely varied substrate chain length and oxidation state preferences when heterologously expressed in *E. coli* (Pulsifer *et al.*, 2014; Kalinger *et al.*, 2018; see Chapter 2).

This study explores whether ALT-type thioesterases exhibit the same degree of catalytic diversity *in planta* as they do in an *E. coli* system. Before this study, the substrate specificities of ALT-type thioesterases *in planta* had not been examined, with the exception of METHYLKETONE SYNTHASE 2 (*ShMKS2*) from wild tomato (*Solanum habrochaites* subsp. *glabratum*) (Yu *et al.*, 2010). *ShMKS2* mostly generates 12:0 and 14:0 β-keto fatty acids, which are converted to 11:0 and 13:0 methylketones (MKs) by METHYLKETONE SYNTHASE 1 (*ShMKS1*), a plastid-localized decarboxylase (Ben-Israel *et al.*, 2009; Yu *et al.*, 2010). These MKs are stored in leaf trichomes, where they protect the plant against herbivorous insects (Fridman *et al.*, 2005; Yu *et al.*, 2010).

*ShMKS1* homologues are unique to the Solanaceae family, giving rise to the hypothesis that the conversion of β-keto fatty acid products of ALT-like thioesterases to MKs is a mechanism that evolved specifically in solanaceous plants, and that ALT-type thioesterases from other plant families may serve functions other than MK biosynthesis (Khuet *et al.*, 2019; Yu *et al.*, 2010). However, as medium-chain MKs have been identified in a variety of non-solanaceous plants, including American cranberry, raspberry, and members of the Rutaceae (rue) family, it is possible that ALT enzymes in other plant families are in fact involved in MK biosynthesis alongside decarboxylases that do not resemble *ShMKS1* (Forney & Markovetz, 1971; Burdock, 2010; Ogunbinu *et al.*, 2007). The presence of paralogous ALT-type thioesterases within species, as well as their highly varied activity profiles when heterologously expressed in *E. coli*, suggests that ALT enzymes occupy specialized and diverse biological roles.
(Kalinger et al., 2018; see Chapter 2). If this catalytic diversity is preserved in planta, it would lend support to this hypothesis.

Significant effort has been directed towards the production of 8:0 and 10:0 fatty acids in transgenic oilseeds, as these chain lengths, which are components of plasticizers and biofuels, are not naturally produced in commercially grown oilseed crops (Dehesh et al., 1996; Kim et al., 2015). Certain ALT-type thioesterases produce 8:0 and 10:0 fatty acids in E. coli, and assuming that their substrate preferences in planta are the same, they could potentially be expressed in transgenic oilseeds to boost the production of these fatty acids. The ability of ALT-type thioesterases to generate the β-keto fatty acid precursors of MKs is also attractive from a biotechnological perspective. Medium-chain MKs are used in the manufacture of fragrances, flavorings, and diesel biofuels, and have insecticidal and antifungal properties (Goh et al., 2012; McDowell et al., 1988; Yuan et al., 2012; Zheng et al., 2013; reviewed in Pohl et al., 2011). ALT-type thioesterases are potential renewable sources of these compounds, which are currently almost exclusively generated for industrial use from petroleum feedstocks (Park et al., 2012). Some ALT enzymes also produce 3-hydroxy fatty acids in E. coli, which have applications in the manufacture of lubricants and surfactants (Kalinger et al., 2018; Khuat et al., 2019; see Chapter 2). The biotechnological potential of ALT-type thioesterases lies in their wide array of products compared to other acyl-ACP thioesterases, and an understanding of the catalytic activity of ALT-type thioesterases in planta is important for the use of these enzymes for metabolic engineering purposes.

The model plant Arabidopsis thaliana possesses four paralogous ALT enzymes, named AtALT1-4, each with unique substrate specificity in E. coli (Pulsifer et al., 2014). In this work, we profiled the direct enzymatic products of AtALT1-4 in planta via overexpression in
Arabidopsis and Camelina sativa seeds, and in N. benthamiana leaves. Seed-specific ALT overexpression led to MCFA accumulation in seed triacylglycerols. While only fully reduced fatty acids would be expected to be incorporated into seed TAG, we developed an experimental system where both fully reduced and β-keto fatty acids produced by ALT-type thioesterases could be detected in N. benthamiana leaves. Transient expression in N. benthamiana leaves revealed that the substrate chain length and oxidation state preferences of AtALT1-4 *in planta* generally agree with their substrate preferences in *E. coli*, providing evidence of the catalytic diversity of ALT-type thioesterases *in planta*.

### 3.3 – MATERIALS AND METHODS

#### 3.3.1 – Plant Growth Conditions

*Arabidopsis thaliana* (Col-0), *Camelina sativa* (L.) Crantz cultivar “Suneson”, and *Nicotiana benthamiana* plants were grown in soil (Pro-Mix BX General Purpose Growing Medium, Premier Tech, Rivière-du-Loup, QC, Canada) at 21°C and at 180–200 μmol/m² light intensity under long-day conditions (16 h/8 h photoperiod) in an Enconair Bigfoot Series growth chamber (Enconair, Winnipeg, MB, Canada). Seeds of all plant species were disinfected before sowing via incubation in ethanol/water (7:3, by vol) for 1 min, followed by incubation in household bleach/water (1:1, by vol) for 5 min. Seeds were then rinsed five times by gentle inversion in sterile distilled water. *Arabidopsis* seeds were cold-stratified at 4°C in the dark for 5 days before planting.

*Arabidopsis* plants used for *Agrobacterium*-mediated transformation (as described below) were grown in 4-inch diameter pots, with four plants of the same genotype per pot. Individual *N. benthamiana* and *C. sativa* plants were grown in 8-inch diameter pots. For *Arabidopsis* and
**C. sativa**, the pots were irrigated with 20–20-20 nitrogen-phosphorus-potassium fertilizer supplemented with Plant-Prod chelated micronutrient mix (Plant Products, Laval, QC, Canada) once every 4 weeks, while for *N. benthamiana*, the pots were irrigated with this fertilizer mixture once every 2 weeks.

### 3.3.2 - Seed-specific expression of AtALT1-4 in Arabidopsis and Camelina

The complete coding sequences of *AtALT1* (*At1g35290*), *AtALT2* (*At1g35250*), *AtALT3* (*At1g68260*), and *AtALT4* (*At1g68280*) were each amplified by PCR and inserted into the pBinGlyRed plasmid (Nguyen et al., 2015) between the EcoRI and NheI restriction sites, downstream of the seed-specific glycinin promoter from soybean (Appendices 2-3). These plasmids were transformed into *Agrobacterium tumefaciens* GV3101::pMP90. The floral dip method (Clough & Bent, 1998; Liu et al., 2009) was used to transform Arabidopsis Col-0 wild-type and the Col-0 *aae15 aae16* double mutant (Tjellström et al., 2013), as well as *C. sativa* cultivar “Suneson.” The *aae15 aae16* mutant harbours T-DNA insertions in *At4g14070* (*AAE15*) and *At3g23790* (*AAE16*), which encode acyl-ACP synthetase enzymes responsible for responsible for re-activating MCFA s that have been released prematurely by acyl-ACP thioesterases, allowing them to re-enter the fatty acid synthesis cycle and undergo further elongation (Koo et al., 2005). MCFA accumulation is thus expected to be greater when a medium-chain acyl-ACP thioesterase is expressed in seeds of plants with an *aae15/16* mutant background as compared to a wild-type background.

The T-DNA region of the pBinGlyRed plasmid contains a gene encoding the DsRED fluorescent protein, expressed under control of the cassava vein mosaic virus (CsVMV) promoter, as a selectable marker (Nguyen et al., 2015; Verdaguer et al., 1998). Positive
transformants were identified by examining seeds under green light for DsRED fluorescence. Seed fluorescence was used as the positive selection criteria for T1, T2, and T3 seeds. Homozygous T3 seeds were used for seed TAG composition analysis. For all genotypes (*Arabidopsis* or *C. sativa*), seed TAG composition was analyzed using three independent transgenic lines (with n = 3 plants per line).

3.3.3 – *Fatty acid composition analysis of Arabidopsis and Camelina seed oils*

Seeds (15-30mg) were weighed, then dried in a dessicator for 2 days, after which sample dry weights were recorded. Seeds were crushed with a glass rod, and 1 mL of 2.5% v/v sulfuric acid in methanol was added to each sample, along with 10 μL of 50 μg/μL glyceryl triheptadecanoate (tri17:0 TAG) in toluene as an internal standard, and 0.2% w/v butylated hydroxytoluene in methanol to prevent oxidation of unsaturated fatty acids during transmethylation. Samples were then heated at 85 °C for 90 min in sealed tubes to convert fatty acyl chains in seed TAG to fatty acid methyl esters (FAMEs). Samples were allowed to cool to room temperature, after which 1.5 mL of 0.9% w/v NaCl solution and 150 μL of *n*-hexane were added to each sample. Samples were vortexed vigorously, and then centrifuged for 2 min at 4500 x g to separate phases. The organic phase was transferred directly to a GC vial containing a glass insert. One μL of this organic phase was used for analysis by gas chromatography with flame ionization detection (GC-FID).

Fatty acid methyl esters were separated on an Agilent GC-FATWAX capillary column (30 m length, 0.25 mm ID, 0.25 μm film thickness) using helium as the carrier gas, with a constant flow of 1.5 mL/min. Column oven temperature was initially held at 50°C for 3 min, increased to 230°C at a rate of 5°C/min, then held at 230°C for 7 min, for a total run time per sample of 46 min. Injector and detector temperatures were maintained at 250°C. Peaks were
integrated using Agilent Technologies OpenLab CDS ChemStation software and normalized with respect to 17:0 internal standard peak area, and the mol % of each fatty acid component of seed TAG was calculated. For each genotype tested, seed TAG composition was analyzed from three individually transformed transgenic lines. For each transgenic line, seed TAG analysis was performed on samples from three individual plants, and the calculated mol % values were averaged over these three samples to represent the line (i.e. n = 3 replicates per transgenic line).

For gas chromatography-mass spectrometry (GC–MS) identification of seed TAG components, fatty acid methyl esters were separated on an HP-5 capillary column (30 m length, 0.25 mm ID, 0.25 μm film thickness) using helium as the carrier gas, with a constant flow of 1.5 mL/min. The same column oven temperature program as for GC-FID analysis was used. The injector temperature was maintained at 250°C, while the detector temperature was maintained at 325°C. EI mass spectra were obtained using an Agilent 5977E GC/MSD operating at an ionization energy of 70 eV.

3.3.4 – Genetic construct assembly for transient ALT gene expression in N. benthamiana leaves

The complete coding sequences of Arabidopsis ALT1, ALT2, ALT3, and ALT4, Solanum habrochaites subsp. glabratum MKS1 (GenBank Accession: AY701574.1), and MKS2 (GenBank Accession: GU987106.1) were each amplified using primers listed in Appendix 2, Table S2.2, to allow for insertion into the pENTR D-TOPO directional entry vector (Invitrogen®). ShMKS1 and ShMKS2 CDS sequences were amplified from cDNA synthesized from total RNA isolated from 26-day-old S. habrochaites leaves. The resulting pENTR D-TOPO plasmids were transformed into DH5α E. coli. Plasmid DNA from transformants was then extracted via alkaline lysis. The insert sequence for each of the seven constructs was
verified via Sanger sequencing using the universal M13F (−20) and M13R primers (Appendix 2). Inserts were then transferred from pENTR D-TOPO to the pK7WG2D Gateway® destination vector (Karimi et al., 2002) using a Gateway LR Clonase II recombination reaction, performed according to manufacturer's instructions (Invitrogen®). The resulting pK7WG2D constructs containing the coding regions under the 35S promoter were transformed into DH5α E. coli, and plasmid DNA from transformants was extracted via alkaline lysis. Each of the seven constructs was then transformed into the GV3101::pMP90 strain of Agrobacterium tumefaciens by electroporation.

3.3.5 – N. benthamiana Transient Assay

Transient expression of ALT enzymes in N. benthamiana leaves was performed according to Sparkes et al. (2006), with a few modifications. Agrobacterium tumefaciens strains harboring binary vectors (pK7WG2D-35S::AtALT1/2/3/4, pK7WG2D-35S::ShMKS1, pK7WG2D-35S::ShMKS2, empty pK7WG2D, pXZP393-35S::CpFatB2, or pBin19-35S::p19 viral suppressor of silencing) were grown overnight at 30°C and 200 rpm in LB broth supplemented with appropriate antibiotics. Cells were pelleted and washed twice in two volumes of infiltration buffer (5 mM MES pH 5.6, 2 mM Na3PO4•12 H2O, 100 μM acetosyringone), then resuspended in one volume of infiltration buffer, and incubated at 30°C for 2 h with gentle shaking. The OD_{600} of cultures were measured and adjusted to 0.1 with infiltration buffer before infiltration. To prevent co-suppression, Agrobacterium cultures were mixed with an Agrobacterium strain expressing the p19 viral suppressor of gene silencing before infiltration (Doan et al., 2012). Agrobacterium cultures carrying 35S::ShMKS2, 35S::CpFatB2, and each of the four Arabidopsis ALTs were infiltrated by syringe either in a 1:1 ratio with 35S::p19, or in a 1:1:1 ratio with 35S::p19 and 35S::ShMKS1 into the underside of leaves from 5-week-old N.
*benthamiana* plants. Leaves infiltrated with the empty pK7WG2D vector served as a negative control. The second and third leaves from the apical meristem were selected for injection, and injections were repeated on three separate plants for each strain or a combination thereof. Plants were left to grow under the conditions described above (Section 3.3.1), and leaf tissue was harvested at 6 days post-infiltration.

### 3.3.6 – Total lipid extractions from *N. benthamiana* leaves and GC analysis of fatty acids

Leaf tissue samples (approximately 300–600 mg) were excised from *Agrobacterium*-infiltrated *N. benthamiana*, weighed, flash frozen in liquid nitrogen, ground with a mortar and pestle, and transferred to glass tubes. Then, 3.75 mL of chloroform/methanol (2:1 by vol) and 1 mL of 1 mM EDTA in 0.15 M acetic acid was added to each sample. Samples were vortexed vigorously, after which 1.25 mL of chloroform and 1.25 mL of 0.88% w/v KCl solution was added. The samples were vortexed again, and centrifuged at 3000×g for 5 min. The lower chloroform phase was transferred to clean glass tubes. Then, 3 mL of 1 M HCl in methanol was added to each sample, with 25 μg of heptadecanoic acid (17:0 FA) and 10 μg of 2-heptadecanone (17:0 MK) as internal standards. Samples were incubated at 80°C for 3 h to transmethylate fatty acyl chains. After cooling to room temperature, 200 μL of hexane was added to each sample, and samples were vortexed vigorously. The upper phase was then transferred to clean GC vials containing glass inserts and 0.8 μL of this extract was injected onto the column for subsequent GC-FID and GC–MS analysis.

For both GC-FID and GC–MS analysis, fatty acid methyl esters were separated on an HP-5 capillary column (30 m length, 0.25 mm ID, 0.25 μm film thickness) using helium as the carrier gas, with a constant flow of 1.5 mL/min. Column oven temperature was initially held at
50°C for 3 min, increased to 320°C at a rate of 7°C/min, then held at 320°C for 5 min, for a total run time per sample of about 46 min. The injector temperature was maintained at 250°C, while the detector temperature was maintained at 325°C. EI mass spectra were obtained using an Agilent 5977E GC/MSD operating at ionization energy of 70 eV.

3.3.7 – Statistical Analysis of Data

For *Arabidopsis* and *C. sativa* transgenic plants expressing thioesterases under the seed-specific soybean glycinin promoter, seeds of three individual plants from three independently transformed transgenic lines were used for TAG analysis, with a total of nine samples per genotype. The statistical significance of differences in seed TAG composition between the three transgenic lines, and between transgenic and control plants, was determined using one-way ANOVA, followed by Tukey's HSD test (95% family-wise confidence level, p < 0.01). Levene's test (p < 0.05) was used to confirm equality of variances between Col-0 wild-type and *aae15 aae16* groups. A two-sample Student's t-test (p < 0.05) was used to compare total MCFA accumulation (in mol %) between *Arabidopsis* plants expressing each thioesterase in a Col-0 wild-type versus an *aae15 aae16* double mutant background.

The statistical significance of differences in MCFA or MK content (in nmol/g fresh leaf weight) between *N. benthamiana* leaves expressing each ALT-type thioesterase either alone or in combination with the *ShMKS1* decarboxylase, between leaves expressing ALT-type thioesterases and those expressing *ShMKS1* alone, and between leaves expressing ALT-type thioesterases and those infiltrated with empty pK7WG2D, was determined using one-way ANOVA, followed by Tukey's HSD test (95% family-wise confidence level, p < 0.05). Levene's test (p < 0.05) was used to confirm equality of variances between groups. R statistical
analysis software (version 4.0.3) was used to perform all statistical analyses.

3.4 – RESULTS

3.4.1 – Seed-specific overexpression of ALT-type thioesterases increases medium-chain fatty acid content in seed triacylglycerol

The major aim of this study was to profile the in planta substrate specificities of ALT-type thioesterases via overexpression of AtALT1-4 in two tissue types (seed and leaf tissues). We hypothesized that when ALT enzymes are overexpressed in seeds, the MCFAs released from acyl-ACP by ALT enzymes would be incorporated into seed triacylglycerols, as is the case when medium-chain FatB-type thioesterases are expressed in seeds (Tjellström et al., 2013). Since the fatty acid synthesis cycle is interrupted by the activity of thioesterases, a concomitant decrease in long-chain fatty acid content in seed TAG was also expected (Tjellström et al., 2013). We also predicted that the substrate chain length and oxidation state preferences of ALT enzymes expressed in seeds would largely match that determined through their heterologous expression in the K27 strain of E. coli bacteria, given the functional and evolutionary relationship between plant and bacterial fatty acid synthases (Kalinger et al., 2018; Pulsifer et al., 2014; Voelker et al., 1996). When expressed in K27 E. coli, AtALT1 mainly generates 12:0 and 14:0 fatty acids, AtALT2 primarily produces 8:0 and 10:0 β-keto fatty acids with trace amounts of 10–14 carbon fatty acids, AtALT3 generates a wide range of 8–16 carbon β-keto fatty acids and trace amounts of fully reduced and 3-hydroxy fatty acids, and AtALT4 produces 6:0 and 8:0 fatty acids (Pulsifer et al., 2014; Kalinger et al., 2018; Table 1). Since β-keto fatty acids and their derivatives are not expected to be incorporated into seed triacylglycerols, we hypothesized that seed-specific expression of AtALT1 and AtALT4 would significantly alter
seed TAG composition, while expression of \textit{AtALT2} and \textit{AtALT3} would have less pronounced effects.

The coding sequences of each of the four \textit{ALT} genes from \textit{Arabidopsis} were inserted into the pBinGlyRed plasmid, placing them under control of the seed-specific glycinin promoter from soybean. Each of these four constructs was then transformed into Col-0 wild-type (WT) and an \textit{aae15 aae16 (aae15/16)} double mutant background. The \textit{aae15/16} mutant, described by Tjellström \textit{et al.} (2013), is deficient in the plastidial AAE15 and AAE16 acyl-ACP synthetase enzymes. AAE15 and AAE16 are responsible for re-activating MCFAs that have been released prematurely by acyl-ACP thioesterases, allowing them to re-enter the fatty acid synthesis cycle and undergo further elongation (Koo \textit{et al.}, 2005). MCFA accumulation is thus expected to be greater when a medium-chain acyl-ACP thioesterase is expressed in seeds of plants with an \textit{aae15/16} mutant background as compared to a wild-type background.

Expression of \textit{AtALT1}-4 in \textit{Arabidopsis} seeds resulted in increased MCFA content in TAG, with greater overall MCFA accumulation in \textit{aae15/16} plants than Col-0 WT plants (Fig. 3.1). Mean total MCFA accumulation was significantly higher in \textit{aae15/16} plants expressing \textit{AtALT2}, \textit{AtALT3}, and \textit{AtALT4} as compared to Col-0 WT plants expressing these thioesterases (Fig. 3.1). For Col-0 WT plants expressing \textit{AtALT1}, \textit{AtALT2}, \textit{AtALT3}, and \textit{AtALT4}, respectively, median MCFA accumulation in seeds totaled 8.05, 0.82, 0.15, and 4.75 mol\% of seed TAG, as compared to 8.81, 1.03, 0.63, and 7.65 mol\% for \textit{aae15/16} plants (Fig. 3.1). These results mirror those described by Tjellström \textit{et al.} (2013) using FatB-type acyl-ACP thioesterases. The identities of the MCFAs produced as a consequence of ALT expression in seeds did not vary between Col-0 WT and the \textit{aae15/16} backgrounds, therefore, only the results of ALT expression in the \textit{aae15/16} mutant background are reported on henceforward.
**Table 3.1.** Summary of the products of AtALT1-4 when heterologously expressed in *E. coli*, as compared to their *in planta* products when overexpressed in *Arabidopsis* seeds and *Nicotiana benthamiana* leaves

<table>
<thead>
<tr>
<th>ALT enzyme</th>
<th>Products observed in <em>Escherichia coli</em></th>
<th>Products observed <em>in planta</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AtALT1</em></td>
<td>12:0 FA, 12:1 FA, 14:0 FA, 14:1 FA</td>
<td>12:0 FA, 14:0 FA</td>
</tr>
<tr>
<td><em>AtALT2</em></td>
<td>10:0 FA (trace), 12:0 FA (trace), 14:1 FA (trace)</td>
<td>10:0 FA, 12:0 FA, 14:0 FA</td>
</tr>
<tr>
<td></td>
<td>7:0 MK, 9:0 MK</td>
<td>13:0 MK (with ShMKS1)</td>
</tr>
<tr>
<td><em>AtALT3</em></td>
<td>8:0 FA, 12:0 FA, 14:1 FA, 14:0 FA</td>
<td>14:0 FA</td>
</tr>
<tr>
<td></td>
<td>7:0 MK, 9:0 MK, 11:1 MK, 13:1 MK, 13:0 MK, 15:1 MK</td>
<td>13:0 MK (with ShMKS1)</td>
</tr>
<tr>
<td><em>AtALT4</em></td>
<td>6:0 FA, 8:0 FA</td>
<td>6:0 FA, 8:0 FA</td>
</tr>
</tbody>
</table>
Figure 3.1. Box-and-whisker plots depicting medium-chain fatty acid content in seed TAG of Arabidopsis expressing AtALT1 (a), AtALT2 (b), AtALT3 (c), or AtALT4 (d) in seeds of a Col-0 wild-type background versus an aae15 aae16 (aae15/16) double mutant background. ALT coding sequences were under control of the seed-specific glycinin promoter from soybean. MCFA content is expressed as average mol % of medium-chain (C6–14) fatty acids in seed TAG. Seed samples from three individual plants from each of three individually transformed transgenic lines were used for GC-FID analysis, for a total of nine samples per genotype. Whisker elements represent the maximum and minimum values of each sample set. Student's t-test results were: (a) \( p = 0.16 \), (b) \( p = 0.008 \), (c) \( p = 0.01 \), (d) \( p = 0.009 \), where \( p \leq 0.05 \) was considered significant.

In aae15/16 plants with no seed-localized ALT expression, trace amounts of 14:0 fatty acid were detected in seed oil (0.22 ± 0.01 mol% of seed TAG), but no 6-12 carbon MCFA s were observed (Figs. 3.2, 3.3, Table 3.1). Seed-specific expression of AtALT1 increased the amount of 14:0 in seed TAG to 1.55 ± 0.35 mol% on average (Figs. 3.2a, 3.3). In addition, 12:0 fatty acid was a major component of TAG in AtALT1-expressing seeds, comprising approximately
9.46 ± 2.12 mol% of TAG (Table 3.1, Figs. 3.2a, 3.3). Notably, one of the three AtALT1-expressing transgenic lines, aae15/16 glycinin::ALT1-line 6, accumulated significantly more 12:0 and 14:0 in seed oil than the other two lines tested (Fig. 3.2a). Differences between transgenic lines are expected and generally result from positioning of the T-DNA within the genome and the number of T-DNA copies inserted (Wilson et al., 1990). A decrease in 16:0 content, as well as in 18:2, the most abundant acyl chain component of Arabidopsis seed TAG, was observed in the AtALT1-expressing plants (Figs. 3.2a, 3.3). Previously, Tjellström et al. (2013) demonstrated that the expression of 12:0- and 14:0-specific acyl-ACP thioesterases in Arabidopsis seeds also leads to a decrease in the proportion of 18:1 and 18:3 fatty acids in TAG; however, this trend was not observed in plants expressing AtALT1 (Fig. 3.2a). It is possible that the activity of AtALT1 is too low to direct resources away from the production of 18:1 and 18:3. Expression of the 12:0-specific thioesterase UcFatB2 from Umbellifera californica and the 14:0-specific thioesterase CpuFatB2 from Cuphea pulcherrima increases MCFA content in Arabidopsis seed TAG to 43.0 and 41.9 mol%, respectively, while MCFA content in AtALT1-expressing seeds was 11.02 ± 1.53 mol% on average (Fig. 3.2a) (Tjellström et al., 2013).
Figure 3.2. Seed TAG composition from aae15 aae16 Arabidopsis overexpressing AtALT1 (a), AtALT2 (b), AtALT3 (c), or AtALT4 (d) under control of the seed-specific soybean glycinin promoter. Gray bars represent aae15/16 plants with no seed-specific ALT overexpression. The amount of each fatty acid species is expressed as mol % of total seed TAG content. Individual seed TAG components were quantified by comparison of GC-FID peak areas to an internal tri17:0-TAG standard. Three independent homozygous lines were tested for each genotype. Values represent the average ± SD for samples from three individual plants per line. Asterisks (*) indicate values that are significantly different from control, and letters are used to indicate where transgenic lines differ significantly from one another. Statistical significance was determined using one-way ANOVA, followed by Tukey's HSD test (95% family-wise confidence level, $p < 0.01$).
Figure 3.3. Representative GC-FID chromatograms of transmethylated seed TAG from *Arabidopsis* plants overexpressing *AtALT1–4* in seeds of the *aae15 aae16* (*aae15/16*) mutant background under control of the seed-specific soybean glycinin promoter. FAME identities were determined by comparison to authentic standards and by GC–MS. Butylated hydroxytoluene (peak is labeled) was added to TAG samples to prevent oxidation of unsaturated fatty acids during transmethylation.
When expressed in *E. coli*, *AtALT2* mostly produces 8:0 and 10:0 β-keto fatty acids, with trace amounts of 10:0–14:0 fatty acids (Pulsifer *et al*., 2014; Table 3.1). ALT enzymes with demonstrated preference for β-keto fatty acids were not expected to have significant effects on seed oil composition. β-keto fatty acids are unstable and their derivatives are not present in *Arabidopsis* seed oils, and seeds likely lack the enzymatic machinery to convert them to a more stable form, such as MKs. The medium-chain β-keto fatty acids released by seed-expressed ALT enzymes are likely re-activated and returned to the fatty acid synthesis cycle, or turned over via β-oxidation. Therefore, only the fully reduced fatty acid products of seed-expressed ALTs would be observed, as once released from the plastid, these can be activated to acyl-CoA thioesters and incorporated into seed TAG. *AtALT2*-expressing seeds produced small amounts of 10:0 (0.82 ± 0.08 mol%), 12:0 (0.42 ± 0.02 mol%), and 14:0 (0.67 ± 0.01 mol%) fatty acids (Figs. 3.2b, 3.3, Table 3.1). While some *AtALT2*-expressing lines showed slight increases or decreases in the amounts of certain long-chain fatty acid species, this was not consistent between lines (Fig. 3.2b). As such, it can be concluded that aside from trace MCFA accumulation, *AtALT2* expression did not otherwise alter seed oil composition.

*AtALT3* expression did not have any pronounced effects on seed oil composition. *AtALT3* has very broad substrate specificity in *E. coli*, producing a wide range of β-keto fatty acids from 8–16 carbons in chain length, trace amounts of some free fatty acids, particularly 8:0 and 14:0, and small quantities of 14:1 3-hydroxy fatty acid (Kalinger *et al*., 2018). Interestingly, *AtALT3*-expressing seeds did not produce 8:0 FA and did not accumulate more 14:0 FA in TAG than control seeds (Table 3.1, Figs. 3.2c, 3.3). Two of the three *AtALT3*-expressing seed lines (Lines 2 and 3) accumulated trace amounts of 6:0 fatty acid (<0.1 mol%) with *AtALT3*-Line 4, accumulating significantly more 6:0 (1.28 ± 0.37 mol%) (Fig. 3.2c). Aside from a slight
decrease in 18:3 content, any other effects on seed oil composition were not consistently observed between transgenic lines (Fig. 3.2c).

Expression of *AtALT4* resulted in the incorporation of 6:0 (2.66 ± 0.27 mol%) and 8:0 (4.22 ± 0.30 mol%) fatty acids in TAG (Table 3.1, Figs. 3.2d, 3.3). More variability in MCFA accumulation was observed between transgenic lines expressing *AtALT4* than among those expressing the other ALT enzymes. A concomitant increase in 18:1 fatty acid in TAG was also observed, with *AtALT4*-expressing seeds containing nearly double the amount of 18:1 in TAG than control seeds (Fig. 3.2d). 18:2 and 18:3 content in TAG of *AtALT4*-expressing seeds was lower than in controls. This phenomenon was previously observed in seeds expressing the 8:0-specific FatB thioesterases *ChFatB2* and *CpuFatB3* from *Cuphea spp.* by (Tjellström et al., 2013). Transgenic lines expressing *AtALT4* in seeds also accumulated less total seed oil as compared to control plants and transgenic lines expressing the other ALT enzymes. Examination of *AtALT4*-expressing seeds under a stereomicroscope revealed that a large portion of seeds were visibly deformed (Fig. 3.4). Germination analyses indicated that a greater than normal proportion of *AtALT4*-expressing *Arabidopsis* seeds were non-viable, with only 71% of seeds germinating as compared to 97% for *aae15/16* control seeds.
Figure 3.4. AtALT4-expressing *Arabidopsis* and *Camelina sativa* seeds exhibit morphological deformities (indicated with arrows) as compared to untransformed seeds. A. Seeds of *aae15 aae16* untransformed *Arabidopsis* control, and seeds overexpressing AtALT4 in an *aae15 aae16* background. B. Seeds of untransformed *Camelina sativa* var. “Suneson” control, and seeds overexpressing AtALT4 in a “Suneson” WT background. Images were captured with a Zeiss SteREO Discovery v20 stereomicroscope and AxioVision Rel. 4.8. software under transmitted light, bright field.

MCFA accumulation in TAG also occurred due to AtALT1 and AtALT4 overexpression in *Camelina sativa* seeds, albeit to a much lesser extent than in *Arabidopsis* (Fig. 3.5, Appendix 4). AtALT1 expression in *Camelina seeds* resulted in the accumulation of small amounts of 12:0 (1.88 ± 0.08 mol%) and 14:0 (0.48 ± 0.01 mol%) fatty acids (Fig. 3.5a), while overexpression of AtALT4 led to accumulation of small quantities of 6:0 (0.06 ± 0.03 × 10−1 mol%), 8:0 (0.39 ± 0.02 mol%), and 10:0 (0.19 ± 0.01 mol%) fatty acids (Fig. 3.5b). AtALT1-expressing *Camelina* seeds displayed an approximate 20% decrease in 18:2 content in seed
TAG, similarly to what was observed in AtALT1-overexpressing Arabidopsis (Figs. 3.2a, 3.5a). However, unlike in Arabidopsis, overexpression of AtALT1 in Camelina seeds resulted in a 15% decrease in 18:1 content, along with a 14% increase in 18:3 content in seed TAG (Figs. 3.2a, 3.5a). AtALT4 expression had similar effects on the overall TAG composition and morphology of both Camelina and Arabidopsis seeds, with AtALT4-expressing Camelina seeds displaying notable deformities, and accumulating nearly twice as much 18:1 Δ9 fatty acid in TAG as untransformed seeds alongside an approximately 34% decrease in the major TAG component (18:3 fatty acid in Camelina) (Figs. 3.2b, 3.4, 3.5b). However, unlike in Arabidopsis, where AtALT4-expressing seeds produced nearly double the amount of both forms of 18:1 fatty acid (Δ9 and Δ11) in seed oil, AtALT4-expressing Camelina seeds showed a 37% decrease in 18:1 Δ11 content as compared to wild-type (Fig. 3.5b).
Figure 3.5. Fatty acid composition of seed TAG from *Camelina sativa* (L.) *Crantz* cultivar ‘Suneson’ overexpressing *AtALT1* (a) or *AtALT4* (b). Gray bars represent seeds from untransformed plants (control). The amount of each fatty acid species is expressed as mol % of total seed TAG content. Individual seed TAG components were quantified by comparison of GC-FID peak areas to an internal tri17:0-TAG standard. Values represent the average ± SD for seed samples from three individual homozygous plants per transgenic line. Statistical significance was determined via one-way ANOVA followed by Tukey’s HSD test (95% family-wise confidence level, *p* < 0.01). Asterisks (*) indicate values that are significantly different from control, and letters indicate where transgenic lines differ significantly from one another.
3.4.2 – MCFA and MK production in *N. benthamiana* leaves as a result of transient overexpression of ALT-type thioesterases

Seed-specific overexpression of *Arabidopsis* ALT enzymes demonstrated that these enzymes have diverse chain length preferences *in planta*. AtALT1-expressing seeds accumulated 12:0 and 14:0 FA in seed TAG, while AtALT4-expressing seeds accumulated 6:0 and 8:0 FA (Table 3.1, Figs. 3.2, 3.3., 3.5). However, AtALT2 and AtALT3, which prefer β-keto fatty acyl-ACP substrates over fully reduced substrates in *E. coli*, had minimal effects on seed oil composition (Figs. 3.2, 3.3) It was highly likely that these ALT enzymes were still active in seeds, but any β-keto fatty acid products they would generate would not be incorporated into seed TAG. In an effort to observe any products of AtALT1-4 that could not be observed when these thioesterases were expressed in seeds, we transiently expressed AtALT1-4 in the leaves of *N. benthamiana*. This was also performed to determine whether the activity profiles of ALT enzymes vary due to the differences in acyl-ACP pools in different plant tissues (Post-Beittenmiller *et al.*, 1991).

For transient overexpression in *N. benthamiana*, *Agrobacterium* strains carrying AtALT1-4 were infiltrated either alone or in combination with a strain carrying 35S::*ShMKS1* from *Solanum habrochaites* subsp. *glabratum*, in an attempt to visualize β-keto fatty acids produced by ALT-type thioesterases. The plastid-localized METHYLKETONE SYNTHASE 1 (*ShMKS1*) decarboxylase from *S. habrochaites* converts the 12:0 and 14:0 β-keto fatty acids generated by the ALT-like thioesterase METHYLKETONE SYNTHASE 2 (*ShMKS2*) to 11- and 13-carbon methylketones *in planta*, which are stored in glandular trichomes for defence against insect herbivores (Yu *et al.*, 2010). It was therefore hypothesized that overexpression of the plastid-localized *ShMKS1* decarboxylase should result in any β-keto fatty acids generated
by *Arabidopsis* ALT enzymes being converted to stable MKs, at least for β-keto fatty acids with 12- and 14-carbon chains (Yu *et al.*, 2010; Yu & Pichersky, 2014). The stable methylketone derivatives of β-keto fatty acids can be resolved using our GC-FID and GC-MS analysis methods. As positive controls for this experimental system, *N. benthamiana* leaves were infiltrated with *ShMKS2* both alone and in combination with *ShMKS1*. In addition, *ShMKS1* should not interfere with the production of MCFA by ALT-type thioesterases or their accumulation in leaf tissues. As a control for successful transient gene overexpression, *N. benthamiana* leaves were infiltrated with a strain carrying 35S::*CpFatB2* from *Cuphea palustris*. Overexpression of *CpFatB2* in *N. benthamiana* leaves results in the accumulation of large amounts of 14:0 fatty acid (Ding *et al.*, 2014).

At 6 days post-infiltration, extensive necrosis was observed in leaf tissue infiltrated with 35S::*CpFatB2* (Fig. 3.6). Free MCFA are toxic to plant cells, and as large amounts of MCFA accumulate in leaf tissue, a portion of them are incorporated into phosphatidylcholine, leading to increased membrane fluidity and consequently membrane leakage (Reynolds *et al.*, 2015). Overexpression of *CpFatB2* in *N. benthamiana* leaves was therefore expected to lead to cell death as a consequence of high MCFA accumulation (Ding *et al.*, 2014). At 6 days post-infiltration, mild chlorosis was observed where leaves were infiltrated with 35S::*ShMKS2*, 35S::*AtALT1* or 35S::*AtALT4*, either alone or in combination with 35S::*ShMKS1*, and 35S::*AtALT2* or 35S::*AtALT3* in combination with 35S::*ShMKS1* (Fig. 3.6). Leaf areas infiltrated with empty pK7WG2D, 35S::*ShMKS1* alone, 35S::*AtALT2* alone or 35S::*AtALT3* alone maintained a healthy appearance (Fig. 3.6).
Figure 3.6. Phenotypes of *N. benthamiana* leaves transiently expressing the FAT-type thioesterase *CpFatB2* or the ALT-like thioesterases *AtALT1-4* and *ShMKS2*, either alone or in combination with the *ShMKS1* decarboxylase. Photographs were captured at 6 days post-infiltration with *Agrobacterium tumefaciens*.

Small amounts of 8:0 (9.68 ± 1.69 nmol/g fresh leaf weight (FLW)), 12:0 (2.86 ± 1.04 nmol/g FLW), and 14:0 (10.01 ± 2.11 nmol/g FLW) fatty acids accumulated in leaves infiltrated with the empty pK7WG2D vector, and no MKs were detected (Fig. 3.7, Appendices 4-5). Overexpression of the *ShMKS1* decarboxylase alone did not significantly alter MCFA or MK content in leaves as compared to empty vector controls (Fig. 3.7). Small amounts of 2-tridecanone (13:0 MK, 10.80 ± 6.42 nmol/g FLW) were detected in *ShMKS2*-expressing leaves, demonstrating that some of the 14:0 β-keto fatty acids generated by *ShMKS2* were successfully converted to MKs by an endogenous *N. benthamiana* decarboxylase (possibly an *ShMKS1* homolog) and stored in leaf tissue (Fig. 3.7e, Appendices 4-5). The primary product of *ShMKS2* overexpression was 14:0 FA (190.09 ± 59.87 nmol/g FLW) (Fig.
Myristic acid accumulation as a consequence of \( ShMKS2 \) expression was expected, as \( ShMKS2 \) is known to hydrolyze myristoyl-ACP (Yu & Pichersky, 2014). However, \( ShMKS2 \) shows much stronger preference for 14:0 \( \beta \)-keto acyl-ACPs, and 13:0 MK is the primary product generated when \( ShMKS2 \) is overexpressed constitutively in \( N. tabacum \) seedlings, even in the absence of \( ShMKS1 \) (Yu & Pichersky, 2014). This suggests that decarboxylase activity in \( N. benthamiana \) either occurs at a much slower rate than in \( N. tabacum \), or that the enzymatic machinery capable of MK production is expressed at much lower levels in the leaves of 5-week-old plants as compared to seedlings. When \( ShMKS2 \) was co-expressed with \( ShMKS1 \) in \( N. benthamiana \) leaves, as shown in Fig. 3.7e, 13:0 MK production in leaves increased more than 60-fold. In addition, 2-undecanone (11:0 MK, 18.72 ± 2.76 nmol/g FLW) and 2-pentadecanone (15:0 MK, 40.22 ± 15.9 nmol/g FLW) were also detected in these leaves (Fig. 3.7e, Appendices 4-5). These results indicate that overexpression of an \( ShMKS1 \)-like decarboxylase is required to efficiently process the \( \beta \)-keto FAs produced as a consequence of \( ShMKS2 \) overexpression in \( N. benthamiana \) leaves (Fig. 3.7e). Interestingly, 14:0 and 12:0 FA accumulation also increased overall in leaves expressing both \( ShMKS1 \) and \( ShMKS2 \), although the amount of 14:0 FA varied greatly between individual leaf samples (Fig. 3.7e). This contrasts previous findings where 14:0 FA accumulation was shown to decrease in \( N. tabacum \) seedlings overexpressing \( ShMKS1 \) and \( ShMKS2 \) together in comparison to plants expressing \( ShMKS2 \) alone (Yu and Pichersky, 2014).
Figure 3.7. Quantification of medium-chain fatty acids and methylketones (MK) detected in *N. benthamiana* leaves transiently expressing *AtALT1* (a), *AtALT2* (b), *AtALT3* (c), *AtALT4* (d), *ShMKS2* (e), or *CpFatB2* (f), either alone or in combination with the *ShMKS1* decarboxylase. Gray bars represent leaves infiltrated with *Agrobacterium* harboring an empty pK7WG2D vector, and black bars represent leaves infiltrated with *Agrobacterium* expressing the *ShMKS1* decarboxylase alone. All genes were expressed under control of the 35S promoter. The amounts of each fatty acid and MK species are expressed as nmol/g fresh leaf weight (FLW). Values represent the average ± SD for tissue samples from three independent leaf infiltrations of each strain or strain combination, each performed on a different plant. Asterisks (*) indicate values that are significantly different from the empty vector control, circles (●) indicate values that are significantly different from leaves expressing *ShMKS1* alone, and diamonds (♦) indicate values that are significantly different from both the empty vector control and leaves expressing *ShMKS1* alone. Letters indicate further statistically significant differences between samples. Statistical significance was determined using one-way ANOVA, followed by Tukey's HSD test (95% family-wise confidence level, *p* < 0.05).
As expected, *CpFatB2*-overexpressing leaves accumulated large levels of 14:0 FA, generating $2367.30 \pm 327.68 \text{ nmol/g FLW}$ on average (Fig. 3.7f, Appendices 4-5). Interestingly, small but statistically significant amounts of 13:0 MK ($10.03 \pm 2.97 \text{ nmol/g FLW}$) were also detected as a consequence of *CpFatB2* overexpression (Fig. 3.7f, Appendices 4-5). *AtALT1* overexpression in leaves resulted in the accumulation of more than 3 times the amount of 12:0 and 14:0 FAs than the empty vector control (Fig. 3.7a, Appendices 4-5). 13:0 MK was also occasionally detected in leaves expressing both *AtALT1* and *ShMKS1*, however, this was not consistently observed across samples (Fig. 3.7a). Overexpression of *AtALT4* led to 6:0 FA accumulation in leaves, which was not detected in the empty vector control nor in leaves expressing *ShMKS1* alone (Fig. 3.7d, Appendices 4-5). *AtALT4*-expressing leaves generated $53.00 \pm 15.92 \text{ nmol 6:0/g FLW}$ on average, and leaves expressing *AtALT4* and *ShMKS1* together accumulated $43.13 \pm 16.98 \text{ nmol 6:0/g FLW}$ (Fig. 3.7d). However, unlike what was observed in *Arabidopsis* seeds, 8:0 FA content in *N. benthamiana* leaves remained unchanged as a consequence of *AtALT4* expression (Fig. 3.7d). Trace amounts of 13:0 MK were consistently detected when *ShMKS1* was co-expressed with *AtALT4*, but this was not significant when compared to leaves expressing *ShMKS1* alone (Fig. 3.7d). Taken together with the results of *ALT* expression in *Arabidopsis* and Camelina seeds, these results show that the *in planta* chain length and oxidation state preferences of *AtALT1* and *AtALT4* remain fairly constant in different tissue types and species, and are in good agreement with their substrate specificities in *E. coli* (Table 1).

*N. benthamiana* leaves overexpressing *AtALT2* and *AtALT3* without *ShMKS1* did not show any changes in lipid composition as compared to leaves infiltrated only with *ShMKS1*-expressing bacteria, or bacteria harboring the empty pK7WG2D vector (Fig. 3.7b,c.). Co-
expression of either of these ALT enzymes with ShMKS1 led to the accumulation of small amounts of 13:0 MK (3.70 ± 2.07 nmol/g FLW for AtALT2 and 4.99 ± 1.61 nmol/g FLW for AtALT3) (Table 1, Fig. 3.7b,c, Appendix 4). While N. benthamiana leaves infiltrated with 35S::AtALT3 did not accumulate more 14:0 FA than control leaves, interestingly, co-expression of AtALT3 with ShMKS1 caused 14:0 FA accumulation to nearly double as compared to the empty vector (Fig. 3.7c, Appendix 4). When overexpressed in E. coli, AtALT2 and AtALT3 both generate C8-10 β-keto fatty acids, which are precursors of 7:0 and 9:0 MK (Table 1). Comparing the chromatograms of control N. benthamiana lipid extracts to authentic 7:0 and 9:0 MK retention standards indicates that these compounds should be readily distinguishable in the GC-FID traces if present (Appendix 4). In E. coli, AtALT2 produces trace amounts of 12:0 and 14:0 FA, and while this was also observed in Arabidopsis seeds, this was not seen in N. benthamiana leaves (Table 1). Also, in addition to 14:0 FA, AtALT3 produces 8:0 and 12:0 FA in E. coli, neither of which were observed as products of AtALT3 expression in Arabidopsis seeds or N. benthamiana leaves (Table 1).

3.5 – DISCUSSION

3.5.1 – ALT enzymes show diverse chain-length and oxidation state preferences in planta

Previous experiments involving the expression of ALT-type thioesterases from various plant species in E. coli showed that these enzymes are catalytically diverse in terms of chain-length and oxidation-state substrate specificities (Kalinger et al., 2018; see Chapter 2). The activity profiles of FatB-type thioesterases in E. coli are generally reflective of their in planta substrate preferences, however, it was not known before this study whether ALT-type thioesterases would be as catalytically diverse in planta as they are in E. coli. We demonstrate
here that *AtALT1*-4 are indeed also catalytically diverse *in planta*, and that the *in planta* substrate specificities of ALT-type thioesterases with preference for fully reduced acyl-ACP chains generally match their substrate specificities in *E. coli*. *AtALT1*, which generates 12- and 14- carbon fatty acids in *E. coli*, produces 12:0 and 14:0 fatty acids when overexpressed in both *Arabidopsis* and *C. sativa* seeds, and in *N. benthamiana* leaves. The only differences between *AtALT1* substrate specificity *in planta* and in *E. coli* are in saturation level preference; the primary products of *AtALT1* in *E. coli* are 12:1 and 14:1 monounsaturated fatty acids (Pulsifer *et al.*, 2014). This was expected because *E. coli* naturally synthesizes 10–14 carbon monounsaturated fatty acids, while desaturases in plants generally act on chain lengths of 16 carbons or more. Overexpression of *AtALT4* led to accumulation of 6:0 and 8:0 fatty acids in *Arabidopsis* and *C. sativa* seed TAG and *N. benthamiana* leaves, which are also its products in *E. coli*. Trace amounts of 10:0, 12:0, and 14:0 fatty acids are observed as a consequence of *AtALT2* overexpression in both *E. coli* and *Arabidopsis* seeds, however, this was not seen in *N. benthamiana* leaves, likely due to low activity or low expression of *Arabidopsis* ALT enzymes in *N. benthamiana* leaves.

For ALT-type thioesterases with a known preference for β-keto fatty acyl-ACP substrates in *E. coli* (see Chapter 2), it was much more difficult to compare their *in planta* substrate specificities with their substrate preferences in bacteria. However, the results of ALT overexpression in *N. benthamiana* still highlight that no two ALT enzymes studied to date are exactly alike. Co-expression of the ALT-like thioesterase *ShMKS2* with its companion decarboxylase *ShMKS1* in *N. benthamiana* leaves resulted in the production of 13:0 MK (a derivative of 14:0 β-keto fatty acid) and 14:0 fatty acid, as well as small amounts of 11:0 and 15:0 MKs. This matches what is observed when *ShMKS2* and *ShMKS1* are co-expressed in *E.
coli (Yu et al., 2010; Yu & Pichersky, 2014). When AtALT3 was co-expressed with the ShMKs1 decarboxylase in N. benthamiana leaves, only small amounts of 13:0 MK and 14:0 fatty acid were detected. 14-carbon-length β-keto fatty acid is the primary product of AtALT3 in E. coli, which suggests that the chain length and oxidation preferences of AtALT3 in planta and in bacteria are similar (Pulsifer et al., 2014). However, other free fatty acid products of AtALT3 in E. coli, including 8:0 and 12:0 fatty acids, were neither observed in AtALT3-overexpressing Arabidopsis seeds nor in N. benthamiana leaves, indicating that the chain length preferences of AtALT3 in planta and in E. coli do not completely agree. Of the four Arabidopsis ALT-type thioesterases, AtALT3 is the most similar to ShMKs2, sharing 62.81% amino acid sequence identity. Since the activity profile of ShMKs2 in N. benthamiana is very similar to its activity in E. coli, both when it is expressed alone or in combination with ShMKs1, similar results were expected for AtALT3. However, it is possible that family- and species-specific sequence features of Arabidopsis ALTs render them less active when expressed in Solanaceous species, or that Arabidopsis ALT genes are not expressed as strongly in N. benthamiana as ShMKs2. These issues could potentially be overcome through codon optimization of Arabidopsis ALT genes for N. benthamiana.

The 8:0 and 10:0 β-keto FAs produced when AtALT2 and AtALT3 are expressed in E. coli were not detected when these ALT enzymes were co-expressed with ShMKs1 in N. benthamiana, which suggests that ShMKs1 nor endogenous decarboxylases present in N. benthamiana leaves have the ability to act on 8–10 carbon β-keto fatty acid substrates. We also hypothesized that the 8:0 and 10:0 β-keto fatty acid precursors of 7- and 9-carbon MKs, respectively, may be produced in very small quantities in N. benthamiana leaves by AtALT2 and AtALT3, such that more time post-Agrobacterium infiltration may be needed for detectable
levels of these compounds to accumulate. However, we tested this hypothesis and observed no significant differences in the lipid profile of leaf tissue harvested at 6, 9, 12, and 14 days post-infiltration with Agrobacterium harbouring either 35S::AtALT2 or 35S::AtALT3 in combination with 35S::ShMKS1 (data not shown).

Understanding the catalytic behavior of ALT-type thioesterases in plant tissues is an important step in using these enzymes as part of a metabolic engineering toolkit for sustainable production of medium-chain fatty acids and MKs. However, this study also uncovered some roadblocks to using ALT-type thioesterases for biotechnological purposes, including low specific activity compared to FatB-type thioesterases, lack of enzymatic machinery to efficiently convert the β-keto fatty acid products of ALT-type thioesterases to MKs, and the deleterious effects of 6:0 FA production in seeds. These issues are addressed in the following sections.

3.5.2 – Recombinant ALT-like thioesterases could be co-expressed with medium-chain specific acyltransferases to increase MCFA accumulation in seed TAG

When AtALT1 and AtALT4 were overexpressed in aae15 aae16 Arabidopsis and in Camelina sativa, the average combined 12:0 + 14:0 fatty acid content in seed TAG never exceeded 12 mol% of seed TAG in Arabidopsis and 2.5 mol% of seed TAG in Camelina. Also, the average combined 6:0 + 8:0 fatty acid content in seed TAG never exceeded 7 mol% in Arabidopsis and only 0.45 mol% in Camelina (Figs. 3.2, 3.5). Since overexpression of FatB-type thioesterases in seeds results in much greater accumulation of MCFA in seed TAG, this suggests that the activity and/or the expression levels of ALT-type thioesterases is significantly lower than that of medium chain-specific FatB-type thioesterases (Kim et al., 2015; Tjellström
et al., 2013).

Low MCFA content in ALT-overexpressing Arabidopsis and Camelina seeds, even in the absence of acyl synthetase activity in aae15 aae16 Arabidopsis, and the negative effects of AtALT4 overexpression on Arabidopsis seed viability indicates that the endogenous acyltransferases in Arabidopsis and Camelina do not efficiently incorporate MCFA into seed TAG. For medium-chain fatty acids to be incorporated into seed triacylglycerols, following their release from the plastid, they must first be activated to acyl-CoA esters, transported to the endoplasmic reticulum, and then they must be esterified onto a glycerol backbone. Glycerol-3-phosphate acyltransferases, lysophosphatidic acid acyltransferase, and diacylglycerol acyltransferase sequentially catalyze the addition of acyl-CoA esters onto the glycerol backbone (reviewed in Yuan & Li, 2020). However, in most plants, including Arabidopsis and Camelina, lysophosphatidic acid acyltransferases (LPAAT) show stringent specificity for unsaturated fatty acyl-CoA substrates (such as 18:1-, 18:2-, and 18:3-CoA), and no or very low selectivity for medium-chain fatty acyl-CoA molecules (Kim et al., 2005; Nlandu Mputu et al., 2009). Diacylglycerol acyltransferase (DGAT) catalyzes the rate-limiting step of TAG biosynthesis, and in most plants, DGAT also does not show selectivity for medium-chain fatty acyl-CoA substrates nor diacylglycerol-containing medium-chain fatty acyl groups (reviewed in Yuan & Li, 2020). Combined with the apparent low activity of natural ALT-type thioesterases as compared to FatB-type thioesterases, this creates significant roadblocks to using ALT-type thioesterases for developing oilseed crops with high MCFA content in TAG.

Several steps could be taken to increase MCFA accumulation in seeds expressing ALT-like thioesterases, the first being to engineer the ALT enzymes themselves for greater specific activity. As ALT enzymes with preference for fully reduced fatty acyl-ACP substrates have
similar activity profiles in planta as that in E. coli, as shown in this study, it is highly likely that any increase in enzyme activity found in E. coli would also be observed in planta. Domain-swapping approaches, iterative computational redesign of protein structure, and randomized mutagenesis and screening have all been employed to increase the activity of other medium-chain acyl-ACP thioesterases towards specific substrates (Jing et al., 2018; Ziesack et al., 2018; Grisewood et al., 2017; Hernández Lozada et al., 2018). Such approaches can applied to single hot-dog fold, ALT-type thioesterases, with domain-swapping and randomized mutagenesis likely being the most appropriate (see Chapter 4). While high-quality hypothetical models of ALT thioesterase domain structure can be generated computationally, the crystal structure of an ALT-type thioesterase has yet to be solved.

Co-expression of ALT enzymes with acyltransferases that have high specificity for MCFA could also be another means of boosting MCFA accumulation. Lysophosphatidic acid acyltransferases with preference for all medium-chain lengths ranging from 8:0–14:0, and diacylglycerol acyltransferases with specificity for certain medium-chain lengths have been characterized in numerous species that naturally incorporate MCFA into their seed oil, such as Cuphea spp., palm (Elaeis guineensis), and coconut (Cocos nucifera). Coconut LPAAT2 and palm kernel DGAT2, which specifically incorporate 12:0 fatty acid at the sn-2 and sn-3 positions of TAG, respectively, could be overexpressed alongside AtALT1 to boost 12:0 accumulation in seed TAG, while class B lysophosphatidic acid acyltransferase from Cuphea pulcherrima, which is highly specific for 14:0 chains, could be expressed alongside 14:0-specific ALT enzymes such as AtALT1 and ShMKS2 (Aymé et al., 2014; Kim et al., 2015; Knutzon et al., 1999). Cuphea viscosissima LPAAT2 specifically incorporates 8:0 and 10:0 at the sn-2 position of TAG, and it could be used to increase 8:0 accumulation in oilseeds
expressing ALT-like thioesterases with similar substrate specificity as \textit{AtALT4}. Combined expression of \textit{C. viscosissima} LPAAT2 with a recently characterized 10:0-specific DGAT from \textit{C. pulcherrima} and a recombinant ALT-like thioesterase could also be a means of engineering oilseed plants for high 10:0 content (Iskandarov \textit{et al.}, 2017).

Although ALT-like thioesterases do not seem to naturally show preference for 10:0 acyl-ACPs over other chain lengths, an engineered ALT with high specificity for 10:0 fatty acyl-ACP could likely be generated using techniques that have previously been applied to medium-chain FatB-type thioesterases. While FatB-type thioesterases possess a double hot-dog fold thioesterase domain, the acyl-ACP chain length specificity of FatB enzymes is entirely dictated by its N-terminal hotdog fold domain. The substrate-binding pocket residues of the medium-chain FatB \textit{UcFatB2} from \textit{Umbellularia californica} have been elucidated, and other regions of the N-terminal thioesterase domain have been shown to directly influence the acyl-ACP chain length preference of \textit{CpFatB2} (Feng \textit{et al.}, 2017; Jing \textit{et al.}, 2018; Ziesack \textit{et al.}, 2018). Domain-swapping approaches guided by computational modeling of the docking interactions between medium-chain FatB-type thioesterases with their acyl-ACP substrates has been successfully used to generate chimeric FatB-type thioesterases with altered chain-length specificities compared to the wild-type enzymes (Feng \textit{et al.}, 2017; Jing \textit{et al.}, 2018; Ziesack \textit{et al.}, 2018). ALT-type thioesterases are good candidates for engineering using similar rational design methods, as their substrate preferences may be dictated by structurally analogous amino acid residues. Chapter 4 of this thesis describes the use of such an approach to identify amino acid residues that influence ALT substrate specificity.

3.5.3 – \textit{AtALT4} overexpression leads to significant restructuring of seed TAG composition and morphological defects in \textit{Arabidopsis} and \textit{Camelina} seeds.
The effects of AtALT4 overexpression in Arabidopsis and Camelina seeds are of particular interest, as the ability to generate 6:0 FA appears to be unique to ALT-like thioesterases. Before this study, no data existed on the effects of 6:0 FA accumulation in Arabidopsis seeds. Also, no medium-chain FatB-type thioesterases with high specificity for 6:0-ACP have been reported. Arabidopsis seeds overexpressing AtALT4 accumulate less total seed oil than control seeds or seeds expressing any of the other ALT-like thioesterases. Both Arabidopsis and Camelina plants expressing AtALT4 in seeds have significantly altered seed TAG composition, with nearly double the amount of 18:1 content in seed TAG while 18:2 and 18:3 content are reduced (Figs. 3.2d, 3.3, 3.5b). Elevated 18:1 fatty acid content in AtALT4-expressing seeds may be a counterbalance to increased membrane and TAG fluidity due to 6:0 and 8:0 fatty acid accumulation (Garces et al., 1994; Tjellström et al., 2013). When oilseed plants, such as sunflower, are grown at low temperatures, 18:1 content in TAG decreases and 18:2 and 18:3 content increases, indicating that 18:1 from TAG is desaturated to compensate for decreased membrane fluidity in the cold (Garces et al., 1994).

A significant proportion of AtALT4-expressing Arabidopsis and Camelina seeds were non-viable and were noticeably deformed when examined under a stereomicroscope (Fig. 3.4). The amount of 8:0 FA produced by AtALT4 should not have such a drastic effect on seed viability, as Arabidopsis seeds expressing 8:0-specific FatB-type thioesterases have been shown to accumulate nearly twice as much 8:0 in TAG without any reported effects on seed morphology (Tjellström et al., 2013). It is possible that the production of 6:0 FA by AtALT4 is responsible for the observed effects. C6 chains may not be incorporated efficiently into TAG by Arabidopsis acyltransferases, leading to a significant proportion of the 6:0 FA generated by AtALT4 to be redirected to β-oxidation, or incorporated into non-storage lipids, such as phospholipids.
Overproduction of fatty acids that cannot be efficiently incorporated into seed TAG results in decreased total seed oil content, as it creates a futile cycle where the fatty acid synthesis cycle is interrupted only for the resulting product to be immediately directed to \(\beta\)-oxidation once released from the plastid (Poirier et al., 1999). This may be why \(At\)ALT4-expressing seeds accumulate less total seed oil than seeds expressing any of the other ALT enzymes. However, the reduced total oil content and morphological defects seen in these seeds could also be the result of disruptions to carbon flux at other points in the TAG biosynthetic pathway, which are in some way caused by \(At\)ALT4 overexpression.

A way to investigate whether the accumulation of 6:0 by \(At\)ALT4 is in fact the cause of the observed negative effects on seed viability would be to co-express plant acyltransferases with specificity for 6:0 acyl chains, so that this chain length could be efficiently incorporated into seed TAG. However, very few plant species incorporate this chain length into their seed oils. A genetically altered strain of \textit{Cuphea viscosissima}, VS-320, accumulates some 6:0 acyl chains in seed TAG, with 6:0 fatty acid accounting for 4.2 mol% of seed oil composition (Geller et al., 1999). This suggests that there is either a FatB-type or ALT-type thioesterase with 6:0 specificity in this strain. Much of the 6:0 content in this strain of \textit{C. viscosissima} is in the form of tricaproin (tri-6:0 TAG), which has only ever been detected in trace amounts in two other tropical plant species, indicating that \textit{C. viscosissima} VS-320 acyltransferases have some activity for this chain length (Geller et al., 1999). The transcriptome of developing seeds from wild-type \textit{C. viscosissima} is available, and numerous genes involved in seed oil biosynthesis, including lysophosphatidic acid acyltransferase 2 (LPAAT2), have been identified and characterized (Kim et al., 2015). Obtaining the LPAAT2 sequence from the VS-320 \textit{C. viscosissima} strain and characterizing this acyltransferase may therefore be of interest.
3.5.4 – Production of large amounts of MKs in N. benthamiana by ShMKS1 / ShMKS2 co-expression, but not ShMKS1 / AtALT co-expression, suggests a physical interaction between ShMKS1 and ShMKS2

Our results suggest that ShMKS2 interacts physically with its companion decarboxylase ShMKS1 in a way that Arabidopsis ALT proteins are not capable of. That is, the small amounts of MKs produced by AtALT2 and AtALT3 when co-expressed with ShMKS1 could be evidence of weak interactions with ShMKS1, while ShMKS1 could form a metabolon with ShMKS2 to efficiently convert the β-keto fatty acids it generates to MKs. While ALT-type thioesterases from many non-solanaceous plants, such as AtALT3, share high sequence identity with ShMKS2, ShMKS1-like decarboxylases have not yet been found outside the Solanaceae family. However, since medium-chain MKs are components of leaf extracts, resin, and/or fruit and flower aromas from several non-Solanaceous plants, including Eclipta indica L., Pilocarpus jaborandi, citrus, coffee, and Commiphora rostrata (Burdock, 2010; Ogunbinu et al., 2007; Sanchez-Recillas et al., 2017; Prokopy et al., 1998; Cruz-López et al., 2016; Evans & Becerra, 2006), alternate enzymatic machinery to synthesize these compounds must exist in other plant taxa. Methylketones produced by plants outside of the Solanaceae may still be derived from the products of ALT-type thioesterases, through the activity of enzymes that are not close ShMKS1 orthologues. Identification and characterization of candidate ketoacyl-ACP decarboxylase genes with similar localization and gene expression patterns as ALT genes in non-Solanaceous species, such as Arabidopsis, might lead to interesting findings.

Also, like other hot-dog fold thioesterases, ALT enzymes probably cannot function as monomers (Pidugu et al., 2009). It is not known whether any endogenous ALT-like/MKS2-like thioesterase in N. benthamiana could be interacting with transiently expressed Arabidopsis
ALT enzymes, and what effect this might have on AtALT activity. Further study of the physical interactions of ALT-like/MKS2-like thioesterases with one another and with other proteins is warranted, as it would likely reveal valuable information regarding the biological functions of these enzymes.

3.6 – CONCLUSIONS

The biotechnological potential of ALT-type thioesterases as sustainable sources of industrially valuable medium-chain fatty acids and β-keto fatty acid derivatives, such as MKs, largely lies in their catalytic diversity. Unlike the FAT-type thioesterases, which typically act on only one or rarely two acyl-ACP substrates even when heterologously expressed in E. coli bacteria, many ALT-type thioesterases produce a variety of major and minor products. A single ALT enzyme could therefore be used to produce a range of desired products in planta. Alternatively, it could be engineered for exclusive specificity towards any one of the acyl-ACP or β-keto-acyl-ACP substrates it normally acts on. By profiling the in planta activities of select ALT enzymes expressed in different tissue types (seeds and leaves), the findings of this study demonstrate that the catalytic diversity of ALT enzymes observed in E. coli is also displayed in planta. The behaviour of ALT-type thioesterases in different biological systems is important knowledge to use ALT enzymes as metabolic engineering tools to tailor the lipid profiles of crops or heterologous hosts (e.g. microbes), and our findings will also help guide future efforts to determine the precise biological roles of ALT-type thioesterases from diverse plant species.
Chapter 4: Structural determinants of substrate specificity in ALT-type acyl-ACP thioesterases

4.1 – ABSTRACT

The diverse substrate chain-length and oxidation state preferences of ALT-type thioesterases set them apart from other plant acyl-ACP thioesterases. While ALTs show promise as metabolic engineering tools to produce high-value medium-chain fatty acids and methylketones in bacterial or plant systems, the structural basis of ALT chain-length and oxidation state selectivity needs to be explored for these applications. Here, we describe the results of a targeted motif-swapping approach to uncover connections between ALT protein sequence and substrate specificity. Guided by comparative motif searches and computational modelling, we exchanged regions of amino acid sequence between ALT-type thioesterases from *Arabidopsis thaliana*, *Medicago truncatula*, and *Zea mays* to create chimeric ALT proteins. Comparing the activity profiles of chimeric ALTs in *E. coli* to their wild-type counterparts led to the identification of interacting regions within the ALT hot dog-fold thioesterase domain that shape substrate specificity and enzyme activity. Notably, the presence of a 31-CQH[G/C]RH-36 motif on the central α-helix was shown to shift chain-length specificity towards 12–14 carbon chains, and to be a core determinant of substrate specificity in ALTs with preference for 12–14 carbon 3-hydroxyacyl- and β-ketoacyl-ACP substrates. For an ALT containing this motif to be functional, an additional 108-KXXA-111 motif and a compatible sequence spanning aa77–93 of the surrounding β-sheet must also be present, demonstrating that interactions between residues in these regions of the catalytic domain are critical to thioesterase activity. The behaviour of the chimeric ALTs also indicated that aa77–
93 play a significant role in dictating chain-length specificity, while aa91–96 influence oxidation state selectivity. Additionally, aa64–67 on the hot dog-fold β-sheet were important in enabling an ALT to act on 3-hydroxyacyl-ACP substrates. By revealing connections between thioesterase sequence and substrate specificity, this study is an advancement towards engineering recombinant ALTs with product profiles suited for specific applications.

4.2 – INTRODUCTION

Besides cleaving fully reduced, 6-14 carbon acyl chains to release MCFAs, ALTs can act on 3-hydroxyacyl-ACP and β-ketoacyl-ACP intermediates of fatty acid biosynthesis to produce medium-chain 3-hydroxy and β-keto fatty acids (see Chapter 2). While medium-chain FatBs have only been found so far in species with high medium-chain triglyceride content in their seed oil, most plant species, including green microalgae, possess ALT-type thioesterases that are not exclusively localized to seeds (Pulsifer et al., 2014; Kalinger et al., 2018). Heterologous expression of ALT-type thioesterases from 11 diverse plant species in E. coli, as described in Chapter 2, demonstrated that while certain ALTs share general chain-length and oxidation state preferences (i.e. preference for 6–8 carbon or 12–14 carbon chains, preference for β-keto fatty acyl-ACPs or fully reduced acyl-ACPs), no two ALTs share the exact same product profiles in a bacterial system (Pulsifer et al., 2014; Kalinger et al., 2018). Typically, an ALT enzyme will act on substrates of varied oxidation states and multiple chain lengths. For instance, an ALT with preference for C14-β-ketoacyl-ACPs may also display substantial activity towards fully reduced C8–16 acyl-ACPs (Kalinger et al., 2018). ALTs are therefore potential sustainable sources of many valuable MCFAs, 3-hydroxy fatty acids, and of methylketones, as the β-keto fatty acids made by ALTs can be easily decarboxylated either chemically or enzymatically (Yu et al., 2010).
The ability of ALTs to generate medium-chain 3-hydroxy fatty acids and methylketone precursors is especially intriguing, as engineering a sustainable means of producing these compounds remains a challenge to date. Methylketones are widely used in the flavouring and fragrance industries and as blending agents for biofuels, while 3-hydroxy fatty acids have applications as components of lubricants and surfactants, and currently both are largely sourced from fossil hydrocarbons (Yan et al., 2020; Youssef et al., 2005; Solís-Solís et al., 2007; Guo et al., 2008; Goh et al., 2012; Park et al., 2012; Hanko et al., 2018). Since the products of ALTs lend themselves to numerous industrial uses, engineering recombinant ALTs with targeted activity profiles would be a worthwhile endeavour. However, a detailed understanding of what dictates their chain length and oxidation state preferences is required for this. The naturally broad substrate specificities of ALTs currently limit their use as metabolic engineering tools, since an ALT will generate some less desirable products alongside compounds of interest for specific applications.

Protein mutagenesis through targeted “domain-swapping,” where a sequence fragment belonging to a protein of interest is replaced with one from a structurally similar, yet functionally different protein, has been used to shed light on the basis of chain-length specificity in many acyl-ACP and acyl-CoA thioesterase enzymes. By exchanging combinations of similarly-sized sequence fragments between two medium-chain FatB enzymes with differing chain length preferences from the oilseed plant Cuphea viscossima, Jing et al. (2018) determined that the N-terminal hot dog-fold domain of medium-chain specific FatB-type thioesterases is responsible for dictating chain length preference, while the C-terminal domain cleaves the thioester bond connecting the fatty acyl chain to ACP. Ziesack et al. (2018) created a chimeric thioesterase with greatly increased affinity for 8:0 acyl-ACP, and another that
exhibited both medium- and long-chain thioesterase activities, by sequentially exchanging sequence fragments between 8:0-specific and 14:0-specific FatB thioesterases from another Cuphea species.

According to the ThYme (thioester-active enzyme) database, ALT enzymes belong to a phylogenetic clade that includes bacterial single hot-dog fold thioesterases such as the YbgC and YbaW medium-chain acyl-CoA thioesterases, and 4-hydroxybenzoyl-CoA thioesterases (Caswell et al., 2022). Little research has been done on how protein sequence relates to substrate specificity in this clade, and an iterative domain-swapping approach was an attractive starting point for investigating the determinants of substrate specificity in ALTs. However, since ALTs are relatively small proteins with a thioesterase domain of ~150 residues, the above-described methods for FatB-type thioesterases are not as suitable for ALTs. Smaller sequence fragments could be exchanged between paralogous ALTs to yield informative results, but this would be inefficient given their high degree of sequence identity (usually ~60–90%), and the large number of recombinant proteins that would need to be analyzed. Additionally, while separate hot-dog fold domains are responsible for chain-length specificity and catalytic action in FatB thioesterases, the sequence of a single hot-dog fold domain necessarily dictates chain-length specificity, oxidation state specificity, and enzyme activity in ALT-type thioesterases, with any mutations having the potential to influence all three (Jing et al., 2018). The promiscuity of ALT enzymes presents another complication. FatB-type thioesterases usually generate only one or two fatty acid chain-lengths as major products, and high-throughput screening methods such as colourimetric assays or overexpression in microbial fatty acid auxotrophs can be used to preliminarily identify mutants with altered substrate specificity or activity (Hernández-Lozada et al., 2018). The ability of individual ALTs to accommodate many substrates of various chain-
lengths, and especially various oxidation states, means a single screening technique will not capture all mutants with interesting product profiles.

These challenges demanded a more finely targeted approach to ALT mutagenesis. Through comparative motif searches, we identified short motifs shared among ALTs with similar chain-length and oxidation state specificities in *E. coli* – specifically, a six-amino-acid (aa) motif on the central alpha-helix of the hot dog-fold domain and a four-amino-acid motif nearby on the underlying beta-sheet. Computational models of the predicted three-dimensional structure of ALTs suggested that residues in these motifs contribute to substrate binding cavity formation. To investigate their potential role in influencing substrate specificity, the identified regions of sequence were exchanged between suitable pairs of ALT enzymes from the model plant *Arabidopsis thaliana*, the eudicot *Medicago truncatula*, and the monocot *Zea mays* to create chimeric ALTs. We report on how these motifs and other regions of the hot dog-fold domain influence ALT chain-length and oxidation state specificities, as determined by comparing the product profiles of chimeric ALT enzymes in *E. coli* to those of ALTs found in nature.

### 4.3 – MATERIALS AND METHODS

#### 4.3.1 - Sequence data collection

The nucleotide coding sequences and amino acid sequences of ALT-type thioesterases were obtained from the NCBI GenBank and the Cannabis Genome Browser Gateway ([http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway](http://genome.ccbr.utoronto.ca/cgi-bin/hgGateway)). Online accession numbers of all sequences mentioned in this chapter are listed in Chapter 2, Section 2.3.1.
4.3.2 - Assembly of DNA constructs for the expression of chimeric ALT proteins in K27(DE3) E. coli

Double restriction digests of pET-28a plasmid DNA using restriction endonucleases BamHI and HindIII were prepared. Digested plasmid DNA was run on a 0.8% agarose gel, and the linearized plasmid was excised and purified using a NucleoSpin Gel and PCR Purification kit (Macherey-Nagel).

DNA sequences encoding the AtALT3-A, AtALT4-A, AtALT4-B, ZmALT1-B, ZmALT3-B, and MtALT1-B mutants were created by overlap-extension PCR using primers listed in Appendix 2, and previously made constructs described in Chapter 2, Section 3 as templates. These constructs contained the wild-type ALT protein coding sequence (either AtALT3, AtALT4, ZmALT1, ZmALT3, MtALT1, or MtALT2) in the pET-28a expression vector, truncated to exclude the coding region for the N-terminal region encoding the plastid targeting sequence. DNA sequences encoding the AtALT4-AB, AtALT4-AC, AtALT4-ABC, AtALT4-ABCD, AtALT4-ABCDE, MtALT1-AB, and MtALT1-AC mutants were assembled from DNA fragments amplified from previously made constructs encoding mutant ALT-type thioesterases, using PCR primers listed in Appendix 2. DNA sequences encoding all other mutant ALTs described in this chapter were chemically synthesized. The protein-coding sequences of all mutant ALTs described in this chapter are listed in Appendix 3. 16-bp extensions were added to the 5’ and 3’ ends of each ALT-encoding sequence, such that the DNA could be ligated into the above digested and purified pET28a plasmid DNA using the In-Fusion High-Speed Cloning Kit (Takara-Clontech). The 16-bp extensions were complementary to the pET-28a vector sequence at the site of linearization and introduced a BamHI restriction site immediately before the start codon of each gene and a HindIII restriction site after the stop codon. Expression in
the pET-28a vector also introduced a T7-epitope tag at the N-terminus of the ALT protein. The isolated plasmid DNA was then transformed into the K27(DE3) strain of *E. coli* previously described in Chapter 2 (Lü *et al.*, 2009).

4.3.3 - *Extraction of fatty acids produced by ALT-expressing E. coli*

Fully reduced, 3-hydroxy, and β-keto fatty acids produced by ALT-expressing *E. coli* were extracted and profiled as previously described in Chapter 2, Section 2.3.3, with minor modifications. K27(DE3) *E. coli* transformed with pET-28a (empty vector) or with pET-28a containing wild-type or chimeric ALT coding sequences, were grown in 50 mL of kanamycin-containing LB media at 37 °C to an OD600 of ~ 0.4. Protein expression was then induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Induced cells were then grown for 20 h at 18 °C with 200 rpm shaking.

Bacterial cultures were harvested in glass tubes via centrifugation at 4000 x g for 5 min. Two mL of culture supernatant was mixed with 2 mL of 20 mM H₂SO₄, and then 2 μL of a 20 μg/μL solution of heptadecanoic acid (17:0 FA) dissolved in toluene was added to each sample as an internal standard for a total of 40 μg 17:0 FA per sample. Samples were then incubated at 75 °C for 30 min to decarboxylate β-keto fatty acids in culture media to methylketones. After cooling the samples at room temperature, lipids were then extracted into 150 μL hexane by vortexing vigorously for 15 s. One μL of the upper hexane layer was used for gas chromatographic analysis.

4.3.4 - *Identification and quantification of fatty acids and methylketones produced by ALT-expressing E. coli via GC-FID and GC-MS*

Lipids were profiled using an Agilent 7820A gas chromatograph equipped with an HP-
5MS column (30 m length, 0.25 mm inner diameter, 0.25 µm film thickness) and a flame ionization detector according to Chapter 2, Section 2.3.4, with some modifications. Samples were injected with a 7:1 split ratio, with the injector temperature maintained at 250 °C and the detector temperature at 325 °C. The carrier gas was helium at a constant flow rate of 2 mL/min. The column oven temperature was initially held at 50 °C for 8 min, then increased at a rate of 15 °C per min up to 325 °C, which was held for 4 min. Peaks were integrated using Agilent Technologies OpenLab CDS ChemStation software and normalized with respect to the 17:0 FA internal standard peak area, and the amount of each fatty acid and methylketone species in culture media was calculated in units of nmol per unit OD and mol%. Fatty acid and methylketone species were identified by electron impact (EI) GC-MS (ionization energy: 70 eV) and by comparison to authentic retention standards. The temperature program, column, and carrier gas flow rate used for GC-MS analysis were the same as described above for GC-FID. Mass spectra of fully reduced fatty acids and saturated methylketones were compared to standard spectra in the National Institute of Standards and Technology’s NIST 17 Mass Spectral Library. Monounsaturated methylketones were identified by comparison to spectra previously reported by Goh et al. (2012). EI mass spectra used for compound identification are shown in Appendix 4. Free 3-hydroxy fatty acids were identified via GC-MS of silylated samples as described in Chapter 2, Section 2.3.5.

4.3.5 – Statistical analysis of GC-FID data

Lipid production by *E. coli* strains either expressing ALTs or harbouring an empty pET28-vector was quantified using spent media from three individual bacterial cultures. Quantities of fatty acids secreted into culture media, and of methylketones derived from secreted β-keto fatty acids, are expressed as the average across cultures (n=3). Data normality was confirmed by
Shapiro-Wilk’s test ($p < 0.05$). The statistical significance of differences in fatty acid and methylketone production in *E. coli* strains expressing ALT-like thioesterases as compared to a strain harbouring an empty pET-28a vector were determined via a right-tailed Student’s t-test ($p < 0.05$). If sample variances were determined to be equal by Levene’s test ($p \geq 0.05$), the significance of differences in fatty acid and methylketone accumulation among groups of strains expressing mutant and wild-type ALT enzymes were assessed by one-way ANOVA followed by Tukey’s Highly Significant Differences test ($n = 3$, 95% family-wise confidence level, $p < 0.01$). Otherwise, the statistical significance of differences among strains was determined by a Kruskal-Wallis test followed by Dunn’s test of significance ($n = 3$, 95% family-wide confidence level, $p < 0.01$).

4.3.6 - Confirmation of ALT-like protein expression via Tris-Tricine SDS-PAGE and immunoblotting

Cell lysates from induced K27(DE3) *E. coli* cultures either expressing ALT-type thioesterases or harbouring an empty pET-28a vector were prepared by the method described in Chapter 2, Section 2.2.6. Tris-Tricine SDS-PAGE of crude cell lysates was then performed according to Khuat *et al.* (2019), with 10 µg of total protein loaded per sample lane. Proteins were then transferred onto nitrocellulose membranes at 270 mA for 25 min. Membranes were incubated in Ponceau S staining solution (0.5% w/v Ponceau S, 5% v/v glacial acetic acid in distilled H2O) for 10 min, de-stained in distilled water for 10 min, and imaged under white light prior to blocking and immunodetection. Membranes were then incubated in bovine serum albumin (BSA) blocking solution (3% w/v BSA in Tris buffered saline + 0.01% Tween-20) for 1 h at room temperature with gentle shaking. The blocked membrane was probed with mouse monoclonal anti-T7 tag primary antibody (Millipore Sigma catalogue no. 69522, lot no.
3020009, diluted 1:30000) overnight at 4 °C. The membrane was then probed with goat anti-
mouse horseradish peroxidase-conjugated secondary antibody (Millipore Sigma catalogue no.
12–349, lot no. 3174787, diluted 1:250000) for 1 h at room temperature with gentle shaking.
Antibody binding was detected with SuperSignal West Pico Plus Chemiluminescent Substrate
(Thermo Fisher), and images were captured using Bio-Rad ImageLab 6.1 software.

4.3.7 - Structural modelling of ALT proteins

The modelled three-dimensional structures of *Arabidopsis thaliana* ALT3/4 (UNIPROT
IDs Q8W583, F4HX80), *Medicago truncatula* MtALT1/2 (UNIPROT IDs G7K1I0, G7K1I9)
and *Zea mays* ALT3 (UNIPROT ID A0A1D6HF47), were retrieved from the AlphaFold
Protein Structure Database (https://alphafold.ebi.ac.uk/) (Jumper *et al.*, 2021; Varadi *et al*.,
2022). The structure of *Zea mays* ALT1 was modelled using the AlphaFold CoLab web
notebook (Jumper *et al*., 2021; Varadi *et al*., 2022). Regions of each model with poor per-
residue confidence scores (pI/DDT < 70) or that presented with disordered secondary structure
were removed from the models using UCSF Chimera 1.15 software (Pettersen *et al*., 2021;
Yang *et al*., 2012). This excluded the predicted N-terminal plastid targeting peptide sequence
and C-terminal residues beyond position 136 of the consensus ALT thioesterase domain
sequence. As shown in a protein sequence alignment of ALT-like thioesterases characterized
in *E. coli*, homology among ALTs deteriorates beyond aa137 (Appendix 1) (Larkin *et al*.,
2007). This nonconservative region is variable in length, and even absent in certain ALTs with
demonstrated activity in *E. coli*, such as ZmALT3 (Appendix 1).

The global and local quality of the monomeric models output by AlphaFold following
omission of low-confidence regions was estimated using the Discrete Optimized Protein Energy
(DOPE) quality scoring metric in Chimera 1.15’s MODELLER 10.1 extension (Yang et al., 2012). All models had initial z-DOPE scores $< -1.0$, which is indicative of a reliable overall fold prediction. Ramachandran and rotamer outliers, steric clashes, and other geometric outliers were detected using MolProbity 4.2 (Chen et al., 2010). Regions with drops in local model quality and outlier residues were then selectively minimized in UCSF Chimera 1.15 (Yang et al., 2012). Any Ramachandran and rotamer outliers that persisted after selective minimization were corrected manually in Chimera, and loop regions with lower quality scores as compared to the model’s global average were refined via the MODELLER extension’s LoopModel protocol (Yang et al., 2012). Global z-DOPE scoring and local DOPE scoring were measured after each iteration of selective minimization and loop refinement. Selective refinement was discontinued when it did not lower the model’s z-DOPE score by $-0.02$ or more. Finally, global energy minimization of the model was performed to eliminate serious steric clashes. Following refinement and minimization, all models had z-DOPE scores $-1.28 \geq -1.55$. Refined monomeric ALT protein models, with residues coloured according to modelling confidence, are included in Appendix 6.

The HSYMDOCK server was then used to construct homo-oligomeric assemblies from the refined monomeric models (Yan et al., 2018). The predicted homo-oligomeric unit for all modelled ALTs was a homotetramer with dihedral symmetry, similarly to other phylogenetically related single hot dog-fold bacterial thioesterases with solved crystal structures, including the *E. coli* acyl-ACP thioesterase YbgC (accession no. 5T06 on the RCSB PDB) (Caswell et al., 2021; Watanabe et al., 2016). For each ALT, the structure with the lowest free energy score out of 50 homotetrameric structures generated by HSYMDOCK was used for further analysis. The top homotetrameric models for each ALT consistently had docking scores
< - 700, indicative of favourable interactions at binding interfaces (Yan et al., 2018). Highenergy residues in the ALT tetramer models were then corrected using the FoldX 5.0 suite’s RepairPDB command (Delgado Blanco et al., 2019). Repaired models were used for further inspection and visualization in UCSF ChimeraX 1.2.5, and for simulation of single- or multiple-residue mutations using the FoldX 5.0 suite’s BuildModel command (Delgado Blanco et al., 2019; Pettersen et al., 2021).

4.3.8 - *In silico* docking of *E. coli* acyl-ACPs with *AtALT4*

Interaction models of ALTs and *E. coli* C6-, C10-, and C14- acyl-ACPs (2FAC, 2FAE, and 3EJB on the RCSB PDB) were generated using the ClusPro docking server (Vajda et al., 2017; Kozakov et al., 2017; Roujeinikova, 2006). Distance restraints to guide docking were set based on prior knowledge of the interaction of acyl-ACPs with thioesterase proteins. The conserved serine residue at position 36 (S36) of ACP must be oriented towards a conserved catalytic aspartic acid residue on the thioesterase’s active loop, which is D14 of the truncated hot dog-fold domain sequence of ALT-like thioesterases (Pulsifer et al., 2014; Jing et al., 2018; Ziesack et al., 2018). To accommodate the phosphopantetheine arm that connects the acyl chain with S36 of ACP, S36 was specified to be within 13.5 Å of D14 on the ALT’s active site loop (Jing et al., 2018). ClusPro generated 10,000 models for each input pairing, which were clustered into the top 10 maximum likelihood conformations. The top 10 binding conformations for each ALT-acyl ACP input pair were then inspected in UCSF ChimeraX 1.2.5 software to ensure satisfaction of distance restraints. The top-scoring binding conformation that appeared among all input pairs was taken as the representative ALT-acyl ACP interaction model.
4.4 – RESULTS

4.4.1 - Two short amino acid motifs are consistently found in ALTs that have preference for β-ketoacyl-ACP and 3-hydroxyacyl-ACP substrates

The hot-dog fold domain sequences of 19 ALT proteins from diverse plant species were ordered according to their known chain-length and oxidation state preferences in E. coli, with the aim of identifying residues that were conserved among ALT-type thioesterases with similar product profiles (Pulsifer et al., 2014; Kalinger et al., 2018, see Chapter 2). Comparative motif searches conducted using the STREME program (minimum motif length = 3aa) did not identify any significant features when aligned ALTs were grouped by chain-length specificity (6–10 carbon versus 12–16 carbon chains preferred) (Bailey, 2021). However, when ALT sequences were grouped according to their oxidation state preferences, two motifs located at amino acid residues (aa) 31–36 and 108–111 of the catalytic domain consistently occurred in thioesterases with known preference for β-keto and 3-OH fatty acyl-ACPs over fully reduced fatty acyl-ACPs (Fig. 4.1A). The first of these motifs reads 31-CQH[G/C]RH-36 (Fig. 4.1A). In ALT-type thioesterases that primarily produce fully reduced fatty acids in E. coli, this motif is altered at one or more residues. These substitutions occur at variable positions, even among ALTs with similar chain-length preferences. For instance, while A. thaliana ALT1 and Z. mays ALT1 both favour C12-C14 chains, aa31–36 of AtALT1 read “CQH[Q]H”, differing only from the 31-CQH[G/C]RH-36 motif at position 35, while aa31–36 of ZmALT1 read “LHSGRD” (Fig. 4.1A). The second conserved motif identified reads 108-[K/R/Q][A/G][I/T][A/G]-111. Since the vast majority of ALTs that prefer β-keto and 3-OH acyl-ACP substrates possess Lys at position 108 and Ala at position 111, this motif will henceforward be referred to as 108-KXXA-111. Notably, in ALT proteins with preference for fully reduced acyl-ACPs, the Ala111 or
Gly111 residue is usually replaced with Val111 (Fig. 4.1A).

The structural positions of the identified motifs were investigated using computational models of the predicted ALT active unit (Figs. 4.1B, 4.2) (Jumper et al., 2021; Varadi et al., 2022; Yan et al., 2018; Pettersen et al., 2021; Yang et al., 2012; Chen et al., 2010). The predicted tertiary structure of the ALT thioesterase domain generated by AlphaFold 2.0 consists of a 20-residue central alpha-helix surrounded by a 5–6 strand antiparallel beta-sheet, similar to its closest homologue with a known crystal structure, the E. coli YbgC thioesterase (PDB ID 5T06) (Fig. 4.2A) (Jumper et al., 2021; Varadi et al., 2022; Watanabe et al., 2016). As single hot-dog fold thioesterases form active units of homo- or heterodimers, tetramers, or higher-order oligomers in vivo, the HSYMDOCK symmetric multimer docking program was used to predict the multimeric structure of ALT enzymes (Yan et al., 2018). The most likely homooligomeric assembly predicted by HSYMDOCK was a tetramer with dihedral symmetry, again resembling EcYbgC (Figs. 4.1B, 4.2A) (Yan et al., 2018; Watanabe et al., 2016). According to these models, aa31–36 are part of the thioesterase domain’s central α-helix, a core feature of hot dog-fold domain structure. Residues 108–111 are nearby on the surrounding β-sheet, with the side-chain of aa35 in close proximity (≤5 Å) to aa111 (Fig. 4.1B). Additionally, the side chains of residue 32 and Asn24, the latter of which is conserved in nearly all ALT enzymes, are within 5 Å of one another (Fig. 4.1B). When models of the ALT active unit are superposed with that of EcYbgC, Asn24 aligns structurally with a predicted catalytic His25 residue of YbgC (Caswell et al., 2021; Watanabe et al., 2016; Pidugu et al., 2009) (Fig. 4.2B). Given that the two motifs found to assort with oxidation state preference in ALTs occupy central positions within the thioesterase domain and are in proximity of one another and key active site residues, we hypothesized that they may be major determinants of substrate specificity in ALT-type
thioesterases.

4.4.2 - Disruption of the 31-CQH[G/C]RH-36 motif on the central alpha-helix of the ALT thioesterase domain significantly alters chain-length and oxidation state preference

A targeted motif-swapping approach was used to investigate whether the motifs identified above influence ALT substrate specificity. The 31-CQH[G/C]RH-36 motif was disrupted in an ALT that normally prefers β-keto and 3-hydroxy acyl-ACP substrates in *E. coli*, and introduced into an ALT that normally shows preference for fully reduced substrates. Pairs of mutant ALT proteins were synthesized where aa31–36 of the hot dog-fold domain were exchanged between one ALT with preference for 14:1 β-ketoacyl- and 3-hydroxyacyl-ACPs, and another ALT from the same species with preference for 8:0 acyl-ACP (Figs. 4.3, 4.4). The rest of the catalytic domain sequence was unaltered. Specifically, sequence was exchanged between MtALT2 and MtALT1 from *Medicago truncatula*, AtALT3 and AtALT4 from *Arabidopsis thaliana*, and ZmALT2 and ZmALT3 from *Zea mays* to create mutant ALTs. These mutants were annotated with the modifier -A, e.g. “MtALT2-A”. In MtALT1, aa31–36 read 31-CQHCGD-36, while in ZmALT3, aa31-36 read 31-IEIARQ-36. AtALT4 and AtALT3 only differ within aa31–36 at position 35, and so an AtALT3 R35M mutant (AtALT3-A) and an AtALT4 M35R mutant (AtALT4-A) were generated by overlap-extension PCR (Figs. 4.1A, 4.3, 4.4, Appendix 3).
Figure 4.1. Two amino acid motifs occur in ALTs that prefer β-keto and 3-hydroxy acyl-ACP substrates. A. Motifs that assort with oxidation state specificity (31-CQH[G/C]RH-36, 108-KXXA-111) are highlighted on a partial alignment of ALT sequences from diverse plant species, which were previously characterized in E. coli using endogenous acyl-ACP pools (see Chapter 2) (Pulsifer et al., 2014; Yu et al., 2010; Kalinger et al., 2018). Residues are numbered beginning from the N-terminus of the hot-dog fold domain; a predicted N-terminal plastid targeting sequence has been excluded. Protein sequences were aligned using ClustalW and sorted by oxidation state specificity in E. coli, and motifs of interest were identified using STREME (Larkin et al., 2007; Bailey, 2021). At = Arabidopsis thaliana, Bd = Brachypodium distachyon, Cr = Chlamydomonas reinhardtii, Cs = Cannabis sativa, Gm = Glycine max, Mt = Medicago truncatula, Os = Oryza sativa subsp. indica, Sh = Solanum hirsutum subsp. glabratum, Sm = Solanum melongena, Vv = Vitis vinifera, Zm = Zea mays. B. The aa31–36 and aa108–111 are indicated on a three-dimensional model of the predicted Arabidopsis thaliana ALT4 homotetramer, generated with AlphaFold 2.0 and HSYMDOCK (Jumper et al., 2021; Varadi et al., 2022; Yan et al., 2018). Examination of this model at the amino acid level shows that the side-chains of certain residues within these motifs are within potential interacting distance of one another (≤5 Å) and a predicted catalytic site residue, N24. Models were visualized in ChimeraX 1.2.5 software (Pettersen et al., 2021)
Figure 4.2. Predicted catalytic and acyl-binding cavity residues in modelled ALTs align with those of *E. coli* YbgC and *Umbellularia californica* FatB. A. Monomer models of the hot-dog fold domains of *Mt*ALT1 and the *E. coli* acyl-CoA thioesterase YbgC, with numbered α-helices and β-strands, and superimposition of ALT homotetramer models (grey) with the crystal structure of the *E. coli* YbgC homotetramer (yellow). ALT monomers were modelled by AlphaFold 2.0, and homotetrameric models were assembled with HYSDMDOCK (Jumper et al., 2021; Varadi et al., 2022; Yan et al., 2018). The crystal structure of *Ec*YgbC was retrieved from the RCSB PDB (PDB ID: 5T06) (Watanabe et al., 2016). B. Comparison of predicted catalytic residues and acyl-binding cavity-forming residues in modelled ALTs and the crystal structure of *Ec*YbgC. *Mt*ALT1 is used as a representative example. Catalytic triad residues belong to two neighbouring subunits. Molecular surfaces formed by acyl-binding cavity residues are coloured according to hydrophobicity (blue = hydrophilic, white = amphipathic, yellow = hydrophobic) C. Comparison of predicted acyl-binding cavity structure in modelled ALTs, and the crystal structure of the N-terminal hot-dog fold domain of *Uc*FatB. *Mt*ALT1 is used as a representative example. Ribbon structure of α1-α2 of the hot-dog fold domain is hidden from models to increase visibility of key residues. The crystal structure of *Uc*FatB was retrieved from the RCSB PDB (PDB ID: 5X04) (Xue & Feng, 2017) and the N-terminal hot dog-fold domain (residues 100-247) were isolated in ChimeraX 1.2.5 software (Pettersen et al., 2021).

Altering the 31-CQH[G/C]RH-36 motif in an ALT with preference for C14 3-OH and β-keto acyl-ACPs led to large-scale changes in oxidation state and chain-length specificity. The *Mt*ALT2-A and *At*ALT3-A mutants adopted similar substrate specificities to *Mt*ALT1 and *At*ALT4, respectively, now showing strong preference for fully reduced substrates and for 8:0 acyl-ACP chains (Figs. 4.3, 4.5). *Mt*ALT1 primarily generates 8:0 FA (71.5 ± 5.8 mol% total product), with 8:0 β-keto FA, the precursor of 7:0 MK, as a secondary product (10.3 ± 1.5 mol%), while the major products of *Mt*ALT2 in *E. coli* are 3-OH 14:1 FA (34.6 ± 2.7 mol%) and the 14:1 β-keto FA precursor of 13:1 MK (11.8 ± 1.1 mol%) (Figs. 4.3, 4.5). The *Mt*ALT2-A mutant lost its affinity for > 10 carbon acyl chains, with 8:0 FA and the precursor of 7:0 MK now comprising 75.4 ± 8.3 mol% and 12.8 ± 2.5 mol%, respectively, of its total product yield in *E. coli* (Figs. 4.3, 4.4, 4.5). The primary products of the *At*ALT3-A mutant were 8:0 FA (69.0 ± 1.8 mol% total product) and 6:0 FA (16.7 ± 0.5 mol% total product), similar to *At*ALT4 (Figs. 4.3, 4.5) (Pulsifer et al., 2014; Kalinger et al., 2018). Additionally,
cells expressing AtALT3-A could no longer generate any β-keto or 3-hydroxy fatty acids, aside from small amounts of 8:0 β-keto FA (1.4 ± 0.3 mol% total product), which is a minor product of AtALT4 (6.5 ± 1.3 mol% total product). Cultures expressing the MtALT2-A and AtALT3-A variants accumulated nearly 40 times as much 8:0 FA / mL OD$_{600}$ than those expressing wild-type MtALT2 and AtALT3 (Figs. 4.3, 4.5).

Conversely, and rather surprisingly, introducing the 31-CQH[G/C]RH-36 motif into AtALT4, MtALT1, or ZmALT3 did not give the resulting mutants the ability to generate 12–16 carbon β-keto or 3-OH fatty acids. *E. coli* strains expressing the AtALT4-A, MtALT1-A, and ZmALT3-A mutants did not produce significantly increased quantities of any FAs or MK precursors as compared to a strain harbouring an empty pET-28a vector, suggesting that these thioesterases were inactive (Fig. 4.3). Immunoblotting of induced cell lysates ruled out deficiencies in recombinant protein expression as the cause of these observations (Fig. 4.6).

Additional MtALT1 and AtALT4 variants were created to assess whether the second motif found to assort with ALT oxidation state specificity, 108-KXXA-111, was required to accommodate the 31-CQH[G/C]RH-36 motif and produce active mutant thioesterases. Residues 108–111 of MtALT1 and AtALT4 were exchanged both alone and in combination with aa31–36 from MtALT2 or AtALT3, respectively (Figs. 4.4, 4.5). These residues were shown to influence oxidation state and chain-length specificity by fine-tuning the thioesterase’s affinity for substrates that it is naturally capable of accommodating. When the 108-KXXA-111 motif was introduced alone into AtALT4, the resulting AtALT4-B mutant generally retained the chain-length preferences of the wild-type enzyme, producing 8:0 FA as its primary product and 6:0 FA as its secondary product (Fig. 4.5). While 8:0 FA made up a slightly greater proportion (53.8 ± 18.3 mol%) of this variant’s total products as compared to wild-type AtALT4, total 6:0
FA accumulation (24.2 ± 4.9 mol%) was slightly decreased (Fig. 4.5). The MtALT1-B variant displayed significantly increased production of 6:0 FA and 8:0 β-keto fatty acid, with a concomitant decrease in 8:0 FA production. (Fig. 4.5). The proportion of 6:0 FA generated by this mutant increased by 15.1 mol% compared to wild-type MtALT1, while that of the 8:0 β-keto FA precursor of 7:0 MK increased by 7.7 mol% (Fig. 4.5). This was balanced by a simultaneous 25.7 mol% decrease in 8:0 FA accumulation. Attempts to shift mutant ALT substrate specificity towards ≥C12 3-hydroxyacyl and β-ketoacyl-ACPs remained unsuccessful, as introduction of the 108-KXXA-111 motif alongside the 31-CQH[G/C]RH-36 motif in either the AtALT4 or MtALT1 backbone still resulted in inactivated mutant enzymes.
Figure 4.3. GC-FID chromatograms of secreted lipids from K27(DE3) E. coli strains expressing wild-type or chimeric ALTs. β-keto fatty acids produced by ALTs were decarboxylated to methylketones with heat and acid treatment prior to identification and quantification by GC-FID and GC-MS. Compounds corresponding to peaks are labelled above the chromatograms (FA = fully reduced fatty acids, MK = methylketone, 3-OH FA = 3-hydroxy fatty acid). In the MtALT2-A mutant, aa31–36 have been replaced with 31-CQHCGD-36 from MtALT1, while the rest of the MtALT2 protein backbone was left unaltered. The AtALT3-A mutant contains a single R35M substitution, such that aa31–36 reads 31-CQHGMR-36, like AtALT4. In the ZmALT1-A mutant, 31-LHSGRD-36 has been replaced with 31-IEIARQ-36 from ZmALT3, with the rest of the protein backbone remaining intact.
4.4.3 - Exchanging residues 31–36 and 108–111 of the thioesterase domain between ALT enzymes that do not possess the 31-CQH[G/C]RH-36 and 108-KXXA-111 motifs also results in changes to chain-length and oxidation state specificity

To further explore the effects of aa31–36 and aa108–111 on ALT substrate specificity, sequence encoding these residues was also exchanged between ZmALT1 and ZmALT3 from Zea mays, which possess neither the 31-CQH[G/C]RH-36 nor 108-KXXA-111 motifs (Figs. 4.1, 4.4, 4.5). ZmALT3 primarily generates 8:0 FA and acts only on fully reduced fatty acyl-ACPs in E. coli, while ZmALT1 produces the precursors of C11-C15 MKs (49.5 ± 4.3 nmol / mL OD) and trace amounts of C14 3-OH FAs (2.5 ± 0.4 nmol / mL OD) alongside its major products of 14:1 FA and 12:0 FA (Figs. 4.3, 4.5) (Kalinger et al., 2018). When aa31–36 were exchanged between ZmALT1/3, both chimeric mutants, ZmALT1-A and ZmALT3-A, were functional (Figs. 4.3, 4.5). ZmALT3-A took on an activity profile that closely resembled ZmALT1, with its chain-length preferences shifted toward C12–14 acyl chains. Its most abundant products in E. coli were 14:1 FA (43.2 ± 6.5 mol%) and 12:0 FA (21.3 ± 3.0 mol%) (Figs. 4.3, 4.5). Notably, it also gained the capability to act on β-keto acyl-ACPs like ZmALT1, as C13–15 MK precursors represented 15.9 ± 0.8 mol% of its product output (Fig. 4.5). On the other hand, replacing aa31–36 in the ZmALT1 backbone with the corresponding sequence from ZmALT3 caused its chain-length preferences to shift towards longer, rather than shorter chain-lengths, with 16:1 FA as its major product in E. coli (33.8 ± 0.7 mol%) (Fig. 4.5).

ZmALT1 and ZmALT3 only differ at position 108 within aa108–111 of the thioesterase domain. ZmALT1 reads 108-RATV-111, while ZmALT3 reads 108-TATV-111 (Figs. 4.1A, 4.4). A R108T ZmALT1 mutant (ZmALT1-B) and T108R ZmALT3 mutant (ZmALT3-B) were
created via overlap-extension PCR. The ZmALT1-B mutant retained its preference for fully reduced 12–14 carbon acyl-ACPs, but also displayed increased overall affinity for β-ketoacyl- and 3-hydroxyacyl-ACPs as a result of the R108T substitution (Figs. 4.4, 4.5). This was unexpected, as aa108–111 of ZmALT1-B were now identical to those of ZmALT3, which is not naturally capable of producing MK precursors or 3-OH FAs in E. coli. The β-keto fatty acid precursors of 11–13 carbon MKs now made up 17.8 ± 3.3 mol% of the ZmALT1-B mutant’s products in E. coli as compared to 11.8 ± 1.5 mol% for wild-type ZmALT1, and 14:1 3-OH fatty acid, which is only a trace product of ZmALT1, now comprised 9.5 ± 1.3 mol% of the ZmALT1-B mutant’s product output (Fig. 4.5). This was balanced by a large decrease in affinity for 16-carbon acyl-ACP substrates (Fig. 4.5). The ZmALT3-B mutant did not display thioesterase activity despite being expressed successfully in E. coli, indicating that ZmALT3 cannot accommodate a T108R substitution (Figs. 4.5, 4.6).

The behaviour of these ZmALT1 and ZmALT3 variants generally aligned with that of the previously described chimeras; mutating aa108–111 could modulate chain-length and oxidation state preferences by adjusting the variant’s affinity for each of its natural substrates, while altering aa31–36 could enable the mutant to act on certain acyl-ACPs or prevent it from binding substrates that the wild-type ALT normally accommodates.
Figure 4.4. Sequence fragments exchanged between ALT pairs to create chimeric thioesterases and their contributions to substrate specificity. Regions exchanged between ALT enzymes are annotated on pairwise alignments of the hot-dog fold domain sequences of AtALT3/4 from Arabidopsis thaliana, MtALT1/2 from Medicago truncatula, and ZmALT1/3 from Zea mays, and highlighted on a three-dimensional model of the predicted ALT homotetramer. These are annotated as follows: A = aa31–36, B = aa108–111, C = aa78–93, D = aa94–96, E = aa64–67. A summary of how each region influences ALT chain-length and oxidation specificity, as could be deduced from the behaviour of mutant ALT enzymes, is also provided. Residues predicted to be of particular importance in dictating ALT substrate specificity based on experimental results or computational models of protein structure are marked with asterisks. Sequence pairs were aligned using the Needleman-Wünsch algorithm, and models of ALT structure were generated with AlphaFold 2.0 and HSYMDOCK (Jumper et al., 2021; Varadi et al., 2022; Yan et al., 2018; Rose & Eisenmenger, 1991).
Figure 4.5. Structure, substrate specificity, and productivity of wild-type and mutant ALTs in *E. coli*. The structural composition of chimeric ALT proteins is represented by coloured blocks. Sequence fragments exchanged between ALT pairs to create chimeric enzymes are annotated as follows: A = aa31–36, B = aa108–111, C = aa78–93, D = aa94–96, E = aa64–67. Residue positions are numbered from the N-terminus of the hot-dog fold domain; a predicted N-terminal plastid targeting sequence has been excluded. Production of individual fatty acid and methylketone species (in mol% of total products), on average, by wild-type and chimeric ALTs in *E. coli* is summarized in three tables, with darker cell shading representing greater proportions. Lowercase letters approximate sample means. Mean values in each column sharing a common letter did not differ significantly according to the appropriate statistical test (Tukey’s HSD test or Dunn’s test at the α = 0.01 significance level). Cells without letters represent values that did not differ statistically from zero or from an *E. coli* strain harbouring an empty pET-28a vector according to a right-tailed Student’s t-test (p < 0.05, adjusted using the Holm-Šidák correction). β-keto fatty acids produced by ALTs were decarboxylated to methylketones prior to GC-MS and GC-FID analysis. Green bars represent the total productivity of ALTs in units of nmol / mL OD600. Quantities reported are the average of triplicate samples (n = 3), and error on these values represents ± SE. ALT variants that did not display thioesterase activity in K27(DE3) *E. coli* were omitted from this figure.
Figure 4.6. Recombinant ALT-like thioesterases are expressed successfully in K27(DE3) E. coli. 
Left: Immunoblots of crude cell lysates from K27(DE3) E. coli strains expressing wild-type or mutant ALT thioesterases show a prominent band at ~20 kDa, which is not present in samples from a strain harbouring an empty pET-28a vector. 10 μg of total protein was loaded into each lane. Expressed ALTs carried an N-terminal T7-tag, and ALT protein bands were detected by probing the membrane with an anti-T7 mouse monoclonal primary antibody and an anti-mouse horseradish peroxidase-conjugated secondary antibody. All membranes were imaged at 59 s exposure using a BioRad ChemiDoc XRS+ system with ImageLab v6.0.1 software. Images taken from different membranes are separated. Right: Membranes used for immunodetection of ALT proteins stained with Ponceau S following transfer from Tris-Tricine SDS-PAGE gels. All Ponceau-stained membranes were imaged under white light, with an exposure time of 1/30 s.
4.4.4 - Residues 78–93 of the ALT hot dog-fold domain influence substrate chain-length specificity

Aside from 31-CQH[G/C]RH-36 and 108-KXXA-111, no other motifs that consistently assorted with substrate preference in ALT enzymes could be identified through comparative motif searches. Therefore, regions of the thioesterase domain that were generally less well-conserved among paralogous ALTs were prioritized next as mutagenesis targets. Residues 78–93 of the ALT hot dog-fold domain, which span β3-β4, represent one such region, and residues 78–93 of MtALT1, ZmALT3, and ZmALT1 were exchanged with the corresponding sequence from MtALT2, ZmALT1, and ZmALT3, respectively (Figs. 4.4, 4.5, Appendix 3). In all three chimeras, named MtALT1-C, ZmALT1-C, and ZmALT3-C, dramatic changes in chain-length specificity were observed.

Replacing aa77–93 with sequence from ALTs that prefer 12–16 carbon substrates caused the chain-length preferences of the MtALT1-C and ZmALT3-C chimeras to shift away from 8:0 acyl chains (Figs. 4.4, 4.5, Appendix 4). C12–14 FAs and C16 FAs made up 51.7 ± 6.3 mol% and 36.8 ± 7.9 mol%, respectively, of the products of MtALT1-C in E. coli, and over half of the ZmALT3-C mutant’s output consisted of C12–16 FAs (58.9 ± 2.2 mol%) (Fig. 4.5). Bacteria expressing the MtALT1-C mutant also began to produce small but significant amounts (≤3.6 mol%) of the precursors of C11-C15 methylketones, which are natural products of MtALT2. The opposite effect on chain-length specificity was observed in the ZmALT1-C chimera, which adopted preference for 8:0 acyl chains (Fig. 4.5). 8:0 FA, which is only a trace product of ZmALT1 in E. coli (0.8 ± 0.1 mol% total products), was the most abundant compound produced by the ZmALT1-C mutant, representing 42.9 ± 5.2 mol% of its products (Fig. 4.5). While average total lipid accumulation (per mL OD) in E. coli cultures expressing
the ZmALT3-C chimera was approximately three times greater than in cultures expressing wild-type ZmALT3, bacteria expressing the ZmALT1-C chimera had a 20-times decrease in comparison to cultures expressing ZmALT1 (Fig. 4.5). Immunoblotting analysis revealed that ZmALT3-C and ZmALT1-C accumulate to similar levels as the wild-type enzymes in E. coli, indicating that these observations reflect changes in mutant thioesterase activity rather than fluctuations in recombinant protein expression (Fig. 4.6).

AtALT3 and AtALT4 only differ at positions 79, 80, and 83 within aa78–93, so these residues were exchanged with those from AtALT3 to create the AtALT4-C mutant (Figs. 4.4, 4.5, Appendix 3). Unlike MtALT1-C and ZmALT3-C, however, the chain-length preferences of this mutant remained unaltered. AtALT4-C continued to produce 8:0 and 6:0 FAs as its major products, in similar proportions to wild-type AtALT4 (Fig. 4.5).

4.4.5 - The 108-KXXA-111 motif and compatible sequence at aa78–93 of the ALT hot dog-fold domain are required to accommodate the 31-CQH[G/C]RH-36 motif

Given that aa78–93 of the ALT catalytic domain exerted influence on chain-length specificity in some cases, the effects of mutating these residues in conjunction with previously explored motifs were investigated. Replacing aa31–36 of the existing AtALT4-C and MtALT1-C chimeras, where aa78–93 had already been mutated, with 31-CQHGRH-36 led to a loss of thioesterase activity. Surprisingly, subsequent introduction of the 108-KXXA-111 motif in these variants, such that it and the 31-CQHGRH-36 motif were both intact, rescued thioesterase activity (Figs. 4.4, 4.5). The re-activated chimeras were named MtALT1-ABC and AtALT4-ABC.
Since AtALT3/4 share identical sequence beyond position 83 within aa78–93, but MtALT1/2 differ at several positions beyond aa83, additional MtALT1 chimeras were created to assess whether aa84–93 played any role in the restoration of thioesterase activity in the AtALT4-ABC and MtALT1-ABC variants. This was found to be the case, since an MtALT1 variant where only aa78–83 were exchanged with the corresponding sequence from MtALT2 alongside aa31–36 and aa108–111 did not show signs of activity. Thioesterase activity was restored in a chimera where aa78–90, aa31–36, and 108–111 were identical to MtALT2, however, the productivity of this chimera (MtALT1-ABC2) was severely compromised (Figs. 4.4, 4.5, Appendices 3-4).

The AtALT4-ABC chimera was the first AtALT4 variant to have a substantially different activity profile from wild-type AtALT4 (Fig. 4.5). While 8:0 FA remained the major product of AtALT4-ABC, this variant’s affinity for 12–14 carbon acyl chains increased significantly. It also no longer produced 6:0 FA, which accounts for 33.6 ± 9.3 mol% of the products of wild-type AtALT4 (Fig. 4.4). The proportion of 3-hydroxy fatty acid products generated by this variant (3.7 ± 1.9 mol%) was now statistically identical to wild-type AtALT3 (4.7 ± 1.7 mol%) (Fig. 4.5). Interestingly, total FA + MK accumulation in E. coli expressing the AtALT4-ABC variant was nearly ten times greater than in AtALT4-expressing bacteria, increasing to 244.0 ± 15.8 nmol / mL OD from 25.0 ± 2.7 nmol / mL OD, although immunoblotting analysis demonstrated that this was due to increased protein accumulation in E. coli (Figs. 4.5 and 4.6).

Comparing the activity profiles of the MtALT1-ABC and MtALT1-ABC2 chimeras to one another and to the MtALT1-C mutant also indicated that residues 91 and 93 of the ALT thioesterase domain influence selectivity towards 3-hydroxy acyl-ACP substrates (Figs. 4.4, 4.5). MtALT1-ABC and MtALT1-ABC2 only differ at these two positions, with the MtALT1-
ABC chimera possessing Phe91 and His93 from MtALT2 instead of Leu91 and Gln93 (Figs. 4.4, 4.5, Appendix 3). 3-OH 14:1 FA, which is the primary product of MtALT2, made up 16.1 ± 2.0 mol% of products in E. coli expressing MtALT1-ABC, while bacteria expressing MtALT1-ABC2 only produced it in trace amounts (1.9 ± 0.5 mol%) (Figs. 4.4, 4.5). 3-hydroxy fatty acids were not detected in samples from bacteria expressing the MtALT1-C variant, which contains aa78–93 identical to MtALT2, but not the 31-CQHGRH-36 motif.

4.4.6 - Residues 94–96 and 64–67 of the ALT catalytic domain also contribute to oxidation state specificity

Further investigations were conducted with the AtALT3/4 pair, as it was unclear why the AtALT4-ABC mutant’s substrate preferences still did not resemble those of AtALT3 as closely as MtALT1-ABC resembled MtALT2. AtALT3/4 differ at only 15 positions within their 140aa hot dog-fold catalytic domain, as shown in Fig. 4.3. Excluding previously explored regions, most differences in catalytic domain sequence between AtALT3/4 occur within aa64–67, aa94–96, and aa140–144 (Fig. 4.4). While MtALT1/2 share identical sequence at aa64–67 and aa94–96, AtALT4 contains the sequences 64-KFLS-67 and 94-SIL-96, and AtALT3 possesses 64-NFLA-67 and 94-FIF-96 in these regions. Residues 94–96 (D) in the existing AtALT4-ABC mutant backbone were replaced with those from AtALT3 to create the AtALT4-ABCD variant, and both aa64–67 (E) and aa94–96 (D) were replaced with the corresponding sequence from AtALT3 to create AtALT4-ABCDE (Figs. 4.4 and 4.5, Appendix 3). Residues aa140–144 were omitted from further analysis due to a breakdown in sequence conservation and modelling confidence at the ALT C-terminus. These residues are typically modelled with low confidence by AlphaFold, and sometimes, with disordered secondary
structure, resulting in their exclusion from models of the predicted ALT homotetramer (Figs. 4.1B, 4.2A, Appendix 6).

Although total lipid production by the *AtALT4*-ABCD and *AtALT4*-ABCDE variants in *E. coli* was compromised in comparison to wild-type *AtALT4*, their substrate specificities now approximated *AtALT3* more closely (Figs. 4.5 and 4.7). 14-carbon products were now the most abundant in samples from the bacterial strains expressing *AtALT4*-ABCD and *AtALT4*-ABCDE, comprising $38.8 \pm 5.2$ mol% and $49.6 \pm 14.9$ mol% of total products, respectively (Figs. 4.5 and 4.7, Appendix 4). Notably, these mutants displayed significantly increased affinity for ≥10C β-keto fatty acyl-ACPs, which are the preferred substrates of *AtALT3* in *E. coli*. Although C13 MK precursors did not represent as substantial a proportion of their product output as in *AtALT3*, these compounds, which are not normally produced by *AtALT4*, now comprised $12.8 \pm 1.8$ mol% and $11.9 \pm 2.7$ mol% of the products of *AtALT4*-ABCD and *AtALT4*-ABCDE, respectively (Figs. 4.5, 4.7). The introduction of aa64–67 (E) from *AtALT3* into the *AtALT4*-ABCDE mutant also caused its affinity for 3-OH fatty acyl-ACPs to increase significantly. 14-carbon 3-OH FAs accounted for $16.7 \pm 7.8$ mol% of this variant’s total products in *E. coli* (Figs. 4.5, 4.7, Appendix 4). Production of 6–8 carbon FAs by strains expressing these variants did not differ from the empty vector control strain.

Interestingly, the *AtALT4*-ABCD and *AtALT4*-ABCDE chimeras also displayed significant alterations in substrate saturation level preference. While most ALTs prefer monosaturated, C14:1 acyl chains over saturated C14 acyl chains in *E. coli*, these variants prefer the latter, as evidenced by GC-FID chromatograms of lipid extracts from bacterial cultures expressing these mutants (Fig. 4.7, Appendices 4-5).
Figure 4.7. GC-FID chromatograms of secreted lipids from K27(DE3) E. coli strains expressing AtALT4 variants or AtALT3. To create the AtALT4-ABCDE chimera, sequence encoding residues 31–36, 64–67, 77–93, and 108–111 of the AtALT4 thioesterase domain was exchanged with sequence encoding the corresponding residues from AtALT3. In addition to the previously listed mutations, residues 94–96 were also replaced with those from AtALT3 in the AtALT4-ABCDE chimera. β-keto fatty acids produced by ALTs were converted to methylketones with heat and acid treatment prior to identification and quantification by GC-MS and GC-FID. Chromatograms of fatty acid and methylketone retention standard mixtures, and of secreted lipids from K27(DE3) E. coli harbouring an empty vector, are shown for comparison. Compound identities of numbered peaks are listed below chromatograms (FA = fully reduced fatty acids, MK = methylketone, 3-OH FA = 3-hydroxy fatty acid).
4.4.7 - Additional AtALT3/4 chimeras reveal more information regarding the influence of identified regions of the ALT catalytic domain on substrate specificity

Three more AtALT4 variants were constructed to assess the combinatorial effects of aa108–111, aa78–96, and aa64–67 of the ALT thioesterase domain on substrate specificity in the absence of the 31-CQH[G/C]RH-36 motif. Wild-type Met35 was restored within the existing AtALT4-ABC, AtALT4-ABCD, and AtALT4-ABCDE mutant backbones, such that the 31-CQH[G/C]RH-36 motif was no longer present and other mutations remained intact. These chimeras were named AtALT4-BC, AtALT4-BCD, and AtALT4-BCDE (Fig. 4.5, Appendix 3). The chain-length preferences of AtALT4-BC, which contained 108-KGIA-111 and aa78–93 from AtALT3, largely resembled those of the wild-type AtALT4 enzyme and the previously described AtALT4-B and AtALT4-C mutants. However, bacteria expressing AtALT4-BC now produced 14:1 3-OH FA and the precursors of C9, C11, and C15 MKs in similar proportions to AtALT4-ABC (Fig. 4.5, Appendices 4-5).

The introduction of 94-FIF-96 in the AtALT4-BCD variant slightly increased its affinity for 12–14 carbon chains while simultaneously decreasing its affinity for 6:0 FA (Figs. 4.4, 4.5, Appendix 3). The proportion of 12–14 carbon products generated by AtALT4-BCD in E. coli increased by approximately 9.0 ± 3.1 mol% compared to the AtALT4-BC variant, and 6:0 FA no longer represented a statistically significant proportion of its product output (Fig. 4.5). The inclusion of 64-KFLA-67 from AtALT3 in the AtALT4-BCDE variant reversed the observed changes in chain-length selectivity, and further boosted 8:0 β-keto fatty acid production to 17.0 ± 3.6 mol% of total products, the greatest proportion for any AtALT4 variant (Figs. 4.4, 4.5, Appendix 3). Unlike E. coli expressing AtALT4-ABCDE, the strain expressing AtALT4-BCDE was incapable of producing 3-hydroxy fatty acids. All three of the
above-described chimeras still produced 8:0 FA as their primary product, similar to AtALT4, and none were capable of producing C14 β-keto FAs in significant amounts (Fig. 4.5).

4.4.8 - Computational modelling of ALT variants and in silico docking with E. coli ACPs illustrates how residues that influence substrate specificity contribute to acyl binding cavity structure and interactions with ACP partners

To make more informed predictions regarding which residues are the most important within the structural regions shown to influence ALT substrate specificity, models of select chimeric ALTs were constructed, and in silico docking simulations were performed between the ALT homotetramer and E. coli acyl carrier protein partners (Jumper et al., 2021; Varadi et al., 2022; Yang et al., 2012; Chen et al., 2010; Vajda et al., 2017; Kozakov et al., 2017; Roujenikova, 2006; Cryle & Schlichting, 2008; Delgado Blanco et al., 2019). According to computational models of the wild-type ALTs used to create the chimeric mutants described in this chapter, residues in areas that influence ALT substrate preferences, specifically residues 35 and 36 on the central α-helix, and residues 91 and 111 nearby on the β-sheet, contribute to a central hydrophobic pocket likely to be the thioesterase’s acyl binding cavity (Figs. 4.2B, 4.4, 4.8). Alignment of these models with the known crystal structure of the E. coli YbgC thioesterase supports this, with predicted cavity-forming residues in ALT models aligning structurally with cavity-forming residues in EcYbgC (Fig. 4.2B) (Watanabe et al., 2016). Several predicted cavity-forming residues in ALTs are also structurally analogous to known acyl-binding cavity residues in the N-terminal hot-dog fold domain of the Umbellularia californica FatB thioesterase (Fig. 4.2C) (Feng et al., 2017; Xue & Feng, 2017).
Figure 4.8. Predicted acyl binding cavity structures of modelled wild-type and mutant ALTs. Computational models of the predicted ALT active unit indicate that introduction of the 31-CQH[G/C]RH-36 motif leads to disruptions in the acyl-binding cavity structure of \textit{At}ALT4 and \textit{Mt}ALT1, while altering the 108-KXXA-111 motif compromises substrate-binding cavity depth in \textit{At}ALT3 and \textit{Mt}ALT2. Ribbon structure of α1-α2 of the hot-dog fold thioesterase domain (aa24–54) is hidden for increased visibility of individual residues. Predicted cavity-forming residues are shown as stick models, and molecular surfaces formed by these residues are coloured according to hydrophobicity (yellow = hydrophobic, white = amphipathic, blue = hydrophilic). Models of wild-type ALT monomers were created by AlphaFold 2.0, and tetramer assemblies were constructed with HSYMDOCK (Jumper \textit{et al.}, 2021; Varadi \textit{et al.}, 2022; Yan \textit{et al.}, 2018). The effects of amino acid mutations were simulated using the FoldX 5.0 suite (Delgado Blanco \textit{et al.}, 2019). Models were visualized in ChimeraX 1.2.5 software (Pettersen \textit{et al.}, 2021).

The FoldX 5.0 suite was used to simulate the effects of certain mutations on the modelled ALT catalytic domain structure (Delgado Blanco \textit{et al.}, 2019). Comparison of models of the inactive \textit{At}ALT4-A, \textit{Mt}ALT1-A, and \textit{Zm}ALT3-A mutants to their wild-type counterparts showed that major disruptions to the predicted acyl binding cavity structure occur as the result of an M35R substitution in \textit{At}ALT4, or the introduction of 34-GRH-36 into \textit{Mt}ALT1 (Fig. 4.8). Hypothetical \textit{At}ALT3 and \textit{Mt}ALT2 variants where 108-KGIA-111 of \textit{At}ALT3 were replaced with 108-KATV-111 from \textit{At}ALT4, and 108-KATA-111 of \textit{Mt}ALT2 were replaced with 108-KTTV-111 from \textit{Mt}ALT1, were also modelled (Fig. 4.8). In these variants, acyl binding cavity depth also appears severely compromised due to displacement of the R35 and H36 side-chains resulting from the introduction of Val111 (Fig. 4.8).

The maximum likelihood docking conformation obtained through docking simulations of the modelled \textit{At}ALT4 homotetramer with \textit{E. coli} holo-ACPs (PDB IDs 2FAC, 2FAE, and 3EJB) demonstrates that residues within regions that influence ALT substrate preferences, specifically aa83–88 and T110 from one monomer subunit, and aa64–67 from the neighbouring subunit, are part of a positively charged patch on the thioesterase surface that forms the predicted interaction interface with acyl carrier protein (Fig. 4.9) (Vajda \textit{et al.}, 2017; Kozakov
et al., 2017; Roujeinikova, 2006; Cryle & Schlichting, 2008). Residues R83, R88, and N64 are shown in this model to contribute to inter-chain hydrogen-bonding interactions with residues belonging to ACP (Fig. 4.9). Changes in chain-length and oxidation state specificities that are observed when these regions of the ALT catalytic domain are mutated may result, in part, from effects on how they interact with ACPs that deliver substrates.
Figure 4.9. Residues within regions that influence ALT substrate specificity may contribute to interactions with ACPs. A. Ribbon model of the maximum likelihood docking confirmation of *E. coli* acyl carrier protein (yellow, PDB ID 2FAC) with the *AtALT4* homotetramer (grey) (Roujeinikova, 2006). The predicted active site residue of *AtALT4*, Asp17, is located at 13.5 Å from of S36 on ACP, which is necessary to accommodate rotation of the tethered acyl chain carrier by ACP about the phosphopantethiene arm (Jing *et al.*, 2018; Ziesack *et al.*, 2018). Docking of *AtALT4* with *E. coli* ACPs was performed using the ClusPro server, and docking conformations were visualized in ChimeraX 1.2.5 software (Pettersen *et al.*, 2021; Vajda *et al.*, 2017; Kozakov *et al.*, 2017). B. Molecular surface model of the predicted ACP docking site of *AtALT4*. Residues within aa78–93 (blue), aa64–67 (green), and aa108–111 (purple) of the ALT thioesterase domain that form part of the interaction interface with docked ACP (yellow) are coloured and shown as stick models. Magnified views show inter-chain hydrogen bonds, represented as blue dashed lines, between residues of *AtALT4* and of docked *E. coli* ACP.
4.5 - DISCUSSION

4.5.1 - Chimeric enzymes inform our understanding of ALT substrate selectivity

The behaviour of chimeric ALT enzymes in *E. coli* provides much insight into what areas of the ALT hot dog-fold domain are involved in this process, and why certain activity patterns are more common among ALTs. Considering the product profiles of wild-type and chimeric ALTs together illustrates that residues 31–36 of the central α-helix primarily influence chain-length preference, rather than oxidation state specificity as initially hypothesized. It stands to reason that this region would affect chain-length preference given that residue 35 of the central ALT α-helix is structurally analogous to T137 of *Uc*FatB, a key cavity-forming residue that can alter chain-length specificity when mutated (Fig. 4.2C) (Feng *et al.*, 2017; Xue & Feng, 2017). Chain-length specificity is necessarily a function of acyl-binding cavity depth in acyl thioesterases that share the core hot-dog fold domain structure. Functional mutants with the 31-CQH[G/C]RH-36 motif consistently displayed strong affinity for 12–14 carbon acyl chains, such that ≤10-carbon substrates were heavily disfavoured (Fig. 4.5). However, the *MtALT1*-ABC2 and *AtALT4*-ABC mutants, which possessed the 31-CQH[G/C]RH-36 motif, showed no greater preference for partially reduced substrates than the *MtALT2*-C or *AtALT4*-BC mutants, which did not possess this motif (Fig. 4.5). Instead, interactions between regions of the ALT hot dog-fold domain β-sheet identified in this study (aa108–111, aa78–96, and aa64–67) dictate whether the enzyme is able to accept substrates of partially reduced oxidation states. For example, the *AtALT4*-BCDE chimera, where all of these regions have been mutated, displayed markedly increased affinity for β-keto FAs, while mutating aa108–111 or aa78–93 of *MtALT1* and *ZmALT1* allowed the resulting variants to act on partially reduced substrates that the wild-type enzymes could not accommodate (Fig. 4.5). It is likely that the sequence features required
for an ALT to accommodate the 31-CQH[G/C]RH-36 motif and retain thioesterase function are what grant an ALT the ability to cleave partially reduced acyl-ACP chains. This combined with the 31-CQH[G/C]RH-36 motif’s dramatic effect on chain-length specificity would, in part, account for why ALTs that naturally possess this motif consistently prefer 14:1 β-keto and 14:1 3-OH fatty acyl-ACPs in *E. coli*. The biological significance of this pattern has yet to be determined.

Computational modelling of the ALT thioesterase domain also provides a possible explanation for the natural co-occurrence of the 31-CQH[G/C]RH-36 and 108-KXXA-111 motifs, and the observed inactivity of ALT variants containing only one these motifs. The side-chain of Val111, which is conserved in most ALTs that do not possess 31-CQH[G/C]RH-36, is likely too large to accommodate 35-RH-36 within the catalytic domain core, leading to structural disturbances that would result in enzyme inactivity (Fig. 4.8). The small side-chain of Ala/Gly 111 in ALTs containing the 31-CQH[G/C]RH-36 motif would therefore be needed to leave space for 35-RH-36. This interaction is similar to what occurs in the N-terminal domain of FatB-type thioesterases. Residue 111 of the ALT catalytic domain is also structurally analogous to the cavity-forming residue S219 in *Uc*FatB, which occludes the acyl-binding cavity and greatly reduces enzyme activity when mutated to bulky tryptophan (Fig. 4.2C) (Feng et al., 2017; Xue & Feng, 2017). Although aa78–93 of the thioesterase domain also play a necessary role in accommodating the 31-CQH[G/C]RH-36 motif, as demonstrated by the AtALT4-ABC, MtALT1-ABC, and MtALT-ABC2 variants, the reason for this is not obvious from structural models. However, AtALT3 and AtALT4 only differ at R/M35 within aa31–36, suggesting that interactions between aa35 and other residues within aa108–111 and aa78–93 are critical to thioesterase function (Fig. 4.4).
The behaviour of ALT variants characterized in this work points toward Lys64 being necessary for 3-OH FA production. Out of all characterized ALT-type thioesterases from diverse plant species, only AtALT4 and ZmALT4 possess Asn64 in place of a conserved Lys64 residue in the hot dog-fold domain (Fig. 4.4, Appendix 1). Neither generate 3-OH FAs in E. coli (Kalinger et al., 2018). An AtALT4 variant could only produce 3-OH FAs upon introduction of Lys64, while the MtALT1-ABC and ZmALT-B chimeras, which already possessed this residue, were capable of producing substantial amounts of 14-carbon 3-OH FAs (Figs. 4.4, 4.5). Although fully reduced FAs and MK precursors from 6 to 16 carbons in chain-length are observed as products of non-recombinant and mutant ALT-type thioesterases, 3-OH FA production by ALTs appears to be restricted to the 12–14 carbon chain-length range (see Chapter 2) (Kalinger et al., 2018). It is possible that ALTs simply cannot accommodate ≤ C10 3-hydroxyacyl-ACP substrates without introducing unnatural mutations, or that ≤ C10 3-hydroxyacyl-ACPs are not abundant enough in E. coli acyl-ACP pools for these products to be detected.

Curiously, none of the chimeric ALT enzymes tested here produced MK precursors as their primary products, although these are the major products of numerous ALTs found in nature, such as ShMKS2, its characterized orthologues in the Solanaceae, and AtALT3, when they are expressed in E. coli (Pulsifer et al., 2014; Yu et al., 2010; Kalinger et al., 2018; Khuat et al., 2019). While MK precursors made up a significant proportion of the products of certain chimeric ALTs, such as AtALT4-ABCDE and MtALT1-ABC2, total productivity of these variants was usually compromised (Figs. 4.5, 4.7). This indicates that other determinants of ALT oxidation state specificity have yet to be elucidated, and that they likely form stabilizing interactions with aa31–36 of the hot dog-fold domain. In the case of the AtALT4-ABCDE
mutant, however, substrate availability could also be a major contributing factor to its low productivity. The saturated ≥C10 acyl chains that this variant prefers are much less readily available in *E. coli* acyl-ACP pools than the monounsaturated ≥C10 acyl chains favoured by heterologously expressed ALT-type thioesterases (Cronan & Thomas, 2009). Further targeted mutagenesis experiments involving *Sh*MKS2 and its close orthologues may assist in features of the ALT catalytic domain that contribute to specificity towards β-keto acyl-ACP substrates.

4.5.2 - Knowledge gained from the study of ALT-like thioesterases is applicable to other bacterial single hot dog-fold thioesterases that are widely used in metabolic engineering strategies

Bacterial single hot dog-fold acyl-CoA thioesterases belonging to the same structural clade as ALTs are used extensively in metabolic engineering pipelines to produce medium-chain β-keto, hydroxy, and fully reduced fatty acids in microbial systems (Yan *et al.*, 2020; Hanko *et al.*, 2018; Goh *et al.*, 2014; Guevara-Martínez *et al.*, 2019). However, little work has been done to understand the biochemical underpinnings of substrate specificity within this enzyme clade (Caswell *et al.*, 2022). Despite plant ALTs being acyl-ACP thioesterases, not acyl-CoA thioesterases like other enzymes that share the same core catalytic domain structure, information regarding the determinants of ALT substrate specificity could still be extended to related bacterial single thioesterases (Caswell *et al.*, 2022). Amino acids that modified substrate specificity in the chimeric ALTs and were predicted to be acyl binding cavity-forming residues in computational models align consistently to cavity-forming residues in the crystal structure of *E. coli* YbgC (Figs. 4.2B, 4.8). This supports the notion that structural features that dictate ALT substrate preferences and activity may also affect these properties in related bacterial single hot dog-fold thioesterases. Knowledge gained from the study of ALTs can be applied to
the construction of optimized mutant versions of related bacterial thioesterases, such as FadM, YciA, and YbgC-type thioesterases, which have already shown promise in metabolic engineering applications (Yan et al., 2020; Hanko et al., 2018; Caswell et al., 2021; Goh et al., 2014; Guevara-Martínez et al., 2019; Yan et al., 2022).

4.5.3 - ALT variants have unique potential as metabolic engineering tools in plant and microbial hosts

Given their unique substrate specificities compared to other plant acyl-ACP thioesterases (e.g. FAT-type), ALTs themselves are also attractive from a biotechnological standpoint. Heterologous expression of medium-chain specific FatB thioesterases originating from plants such as Cuphea spp. and Umbellularia californica, is common in metabolic engineering pipelines to produce medium-chain fatty acids, fatty alcohols, and methylketones in microbes, and in the development of transgenic oilseed crops that accumulate high medium-chain content in seed triacylglycerols (Dehesh et al., 1996; Voelker et al., 1996; Radakovits et al., 2011; Lin et al., 2018; Zhang et al., 2011; Kim et al., 2015; Wu & Sun, 2014; Yan et al., 2020, Hernández-Lozada et al., 2018; Yan et al., 2022; Hernández-Lozada et al., 2020). ALTs also naturally lend themselves to these purposes. The non-recombinant and chimeric ALTs described in this work overlap in their specificities with seed expressed medium-chain FatB-type thioesterases (Jing et al., 2011). Nonetheless, certain unique chain-length specificity patterns are observed among ALTs; for instance, certain ALTs possess high affinity for 6:0-ACP, while no FatB-type thioesterases with appreciable activity toward 6:0 chain-lengths have been identified (Jing et al., 2011). Developing ALT variants with strong specificity towards 6:0-ACP could therefore be especially useful for future biotechnological applications. The generally broader range of fatty acid chain-lengths output by ALT-type thioesterases may also be beneficial for certain
applications. For instance, medium-chain triglyceride mixtures containing C6-C12 chain-lengths are administered in obesity management and the treatment of metabolic disorders (Nagao & Yanagita; 2010).

Heterologous overexpression of acyl-ACP thioesterases is sufficient to produce fully reduced fatty acids of desired chain-lengths. However, chemical conversion of these free fatty acids is required to produce other medium-chain oleochemicals, such as methylketones and 3-OH fatty acids, which introduces a redundant cycle of acyl-ACP thioester cleavage and reactivation to acyl-CoA thioesters (Yan et al., 2020; Hanko et al., 2018). With their ability to produce β-keto and 3-OH fatty acids directly from intermediates of fatty acid biosynthesis, overexpression of ALT-type thioesterases would eliminate the need for this repetitive step in the production of valuable medium-chain compounds. Highly productive recombinant ALTs with specificity for medium-chain 3-hydroxyacyl-ACPs would be especially useful in metabolic engineering contexts involving bacterial strains, since often-used bacterial acyl-CoA thioesterases such as TesB, YigA, and FadM, only produce these compounds as minor byproducts (Yan et al., 2022). In addition to their value as surfactants and additives to cosmetics, biodiesel and other consumer products, medium-chain 3-OH FAs are building blocks of poly-3-hydroxyalkanoates, biodegradable polymers with desirable plasticizer and elastomer properties (Yan et al., 2022). Developing recombinant ALT enzymes with greatly increased 3-OH FA productivity could be achievable in the near future, as certain features of the ALT catalytic domain that can be manipulated to boost affinity for 3-OH fatty acyl-ACPs were uncovered in the present study. The capability of ALTs to act on 3-hydroxy-ACP intermediates of fatty acid biosynthesis could also prove useful in plant systems; however, while 3-hydroxy fatty acids are well-known components of bacterial endotoxins, they are rarely
documented as plant metabolites outside of fatty acid biosynthesis, especially in the medium-chain length range (Racovita et al., 2015; Clericuzio et al., 2017; Özen et al., 2004). It is unknown whether recombinant ALTs with preference for 3-hydroxyacyl-ACPs in *E. coli* would retain these preferences *in planta*.

If recombinant ALT enzymes are to be developed for use as metabolic engineering tools, the potential for heterologously expressed ALTs to behave in unexpected ways must also be considered. Overexpression of select ALTs in the seeds of *Arabidopsis thaliana* and *Camelina sativa*, as well as in *Nicotiana benthamiana* leaves demonstrated that their chain-length and oxidation state preferences are very similar in *E. coli* cells versus plant tissues (see Chapter 3). While this work suggested that ALTs behave similarly whether they are receiving substrates from ACPs of plant or bacterial origin, it is possible that alternate substrate specificity patterns will arise when ALTs interact with ACPs from other species. Recent work by Sztain et al. (2021) demonstrated that acyl carrier proteins communicate what chain length they are delivering to potential partner enzymes, such as thioesterases, through highly specific protein-protein interactions (Sztain et al., 2021). When a partner enzyme detects that an ACP is delivering a suitable substrate, the tethered acyl chain is translocated from the ACP core to the active site of the partner enzyme (Sztain et al., 2021). While ALTs would be equipped to reliably interpret substrate information communicated by plastidial acyl-ACPs in plant tissues where they are naturally expressed, structural differences in ACPs from other host species could create the possibility for substrate identity to be “misread”, resulting in changes to substrate specificity and enzyme activity (Sztain et al., 2021). Additional considerations must be made in microbial systems, where processes involving fatty acyl-ACPs and fatty acyl-CoAs are not
segregated to distinct cellular compartments as they are in plants (Ohlrogge & Jaworski, 1997; Cronan & Thomas, 2009).

4.5.4 - Decoding individual-residue contributions to ALT substrate specificity will allow for rational design of recombinant ALTs for distinct biotechnological applications

Features of the ALT hot dog-fold domain highlighted in this study likely affect substrate specificity in most ALTs, and these amino acids can be manipulated to broadly control the substrate preferences of ALT enzymes as demonstrated here. For instance, chain-length preferences can be directed towards C6–8 or C12–14 acyl-ACPs, products of partially reduced oxidation states can be reduced or eliminated, or affinity for 3-OH acyl-ACPs can be greatly increased by manipulating the regions of the ALT catalytic domain described in this work (Fig. 4.5). This is a promising step towards rational design of ALT enzymes with product profiles that are suited to specific industrial applications, however, fine-tuning recombinant ALTs to output a narrow range of products of interest will require a better understanding of per-residue contributions to ALT substrate specificity. Thioesterase-ACP docking simulations involving modelled ALT variants, and substrate-fitting studies with attached acyl chains, will allow for the generation of more detailed hypotheses regarding the roles of individual residues in shaping ALT function. Engineering ALT variants with chain-length specificities that deviate from natural trends (i.e. preference for C8 acyl-ACP, C12-C14 acyl-ACPs, or a broad mixture of both) may also require the introduction of residue substitutions that are uncommon or even absent from ALTs found in nature.

As industrial applications demand enzymes that generate products of interest in large quantities, improving the productivity of ALT-type thioesterases is another critical goal that must be pursued to harness their biotechnological potential. While certain regions of the ALT
hot-dog fold domain identified here are linked to increased production of certain compounds, such as 8:0 FA, C12–14 FAs, and C14 3-OH FAs, it is more difficult to comment on how they affect total thioesterase productivity due to differences in recombinant protein expression levels, and the sometimes inconsistent effects of mutating these regions on total FA + MK production by ALTs from different plant species (Fig. 4.5). Moving forward, absolute, rather than relative quantification of fatty acids and methylketone precursors generated by promising ALT variants with improved productivity and desired specificities will provide a more informative assessment of the contributions of individual residues and motifs toward thioesterase productivity.

Randomized mutagenesis within regions of the hot dog-fold domain identified in this study is likely the most effective means of obtaining useful, highly active ALT variants for further experimentation, and gaining insight into how the chemical properties of individual residues contribute to an ALT’s substrate preferences. However, such an approach depends on a quick and reliable mutant screening method. The development of a high-throughput screening method capable of detecting increases in the production of 3-OH FAs and MK precursors, or differentiating between different 3-OH FA and MK chain-lengths, would be an invaluable asset for the construction of a collection of mutant ALTs that could be of use in various biotechnological contexts.

4.6 – CONCLUSIONS

This work uncovered several structural regions that function as core determinants of substrate specificity in ALT-type thioesterases. By exchanging sequence fragments encoding these regions between pairs of ALT enzymes with different substrate preferences and profiling
the products of the resulting chimeric thioesterases in *E. coli*, we demonstrated that mutating these regions of the ALT catalytic domain results in substantial changes to chain-length and oxidation state specificity. Also, these changes are largely consistent among ALTs originating from diverse plant species. Chain-length and oxidation state preferences are found to be coupled among ALT-like thioesterases found in nature, with most ALTs that primarily act on β-keto and 3-hydroxy acyl-ACP intermediates in *E. coli* also preferring 12–14 carbon chain-lengths. We determined through motif-swapping and computational modelling that this is due to the presence of two short interacting motifs in ALTs that contribute to the catalytic domain core, and their interaction with a third, critical region predicted to also participate in ACP docking. The behaviour of chimeric ALTs also provided evidence that it is possible to tune ALT chain-length specificity and oxidation state preferences separately. By manipulating the regions of sequence identified in this study, it is possible to adjust an ALT’s specificity for 6–8 carbon versus 12–14 carbon acyl chain-lengths, and its selectivity for fully reduced or 3-hydroxyacyl-ACPs. The capability of ALT enzymes to directly produce both fatty acids and other industrially valuable medium-chain oleochemicals is advantageous in many biotechnological applications. The results of the targeted mutagenesis experiments described here ultimately show that engineering recombinant ALT enzymes that are highly active and specific for products of interest is a feasible goal that could be met by further investigating the underlying mechanisms of substrate specificity in these unique acyl-ACP thioesterases.
Chapter 5: Summary of findings and future research directions

5.1 – SUMMARY OF RESEARCH FINDINGS

5.1.1 – ALTs are a highly diverse family of unusual plant acyl-ACP thioesterases, and likely occupy multiple, highly specialized roles in plant defence

Profiling the substrate specificities of ALT enzymes from diverse plant species in E. coli revealed that ALTs can be categorized into four groups according to their general chain-length and oxidation state preferences. Group 1 ALTs exhibit broad specificity, generating a wide variety of products ranging from 8-16 carbons in chain-length, but show preference for 14-carbon β-ketoacyl- and 3-hydroxyacyl-ACPs. Group 2 ALTs mainly act on 6-10 carbon fully reduced acyl-ACP and β-keto acyl-ACP substrates. Group 3 ALTs prefer 12-14 carbon, fully reduced acyl chains. Group 4 ALTs display preference for 16-carbon acyl-ACPs. It must be emphasized that all ALTs within a specificity group still differ significantly from one other in the range of secondary products they generate and their productivity in E. coli, such that no two ALTs are exactly alike.

While medium-chain fatty acids (FAs) and β-keto fatty acids (β-keto FAs) had been previously reported as products of Arabidopsis and wild tomato ALT enzymes, this work demonstrated that ALTs also generate medium-chain 3-hydroxy fatty acids (3-OH FAs). This is highly novel, as medium-chain 3-OH FAs had not been previously recorded as products of plant acyl-ACP thioesterases, and are rarely documented in plants outside of their role as intermediates in de novo fatty acid synthesis (Racovita et al., 2015; Clericuzio et al., 2017; Özen et al., 2004). As a result of ALT activity, many more plant species than presently known
may produce medium-chain 3-OH FAs as secondary metabolites, with biological functions that are as of yet unexplored.

The functional characterization of ALTs from diverse plants also provides insight into their evolutionary history and potential biological roles. Early-diverging plant lineages (i.e. before the evolution of enclosed seeds) were overrepresented among ALTs with specificity for fully reduced 16-carbon chains, an activity pattern associated with the ubiquitous thioesterases central to primary lipid metabolism (i.e. Fat-type thioesterases). ALT enzymes possibly descended from a primary metabolic palmitoyl-ACP thioesterase that later evolved specificity for medium-chain fatty acyl-, β-ketoacyl-, and 3-hydroxyacyl-ACPs. Despite sharing high sequence identity, the paralogous ALTs characterized here usually belonged to different substrate specificity groups, suggesting that these evolved to occupy several different secondary metabolic functions. For example, ALTs that produce highly volatile 6-10 carbon FAs and β-keto FAs in E. coli are localized to floral organs according to transcriptomic data (Pulsifer et al., 2014; Sekhon et al., 2011; He et al., 2009), where their products might serve to protect floral parts against pathogen infection, or signal beneficial insects such as pollinators and predators of herbivorous species (Chen et al., 2003). In plant species with a single ALT gene, this encoded a Group 1 ALT that produced a wide range of medium-chain β-keto- and 3-OH FAs with notable insecticidal and antimicrobial properties. These genes are also typically expressed in most organs according to transcriptomic data, which suggests that when a plant has only one ALT, it participates in plant-wide pest and pathogen defence (Fasoli et al., 2012; Severin et al., 2010; Libault et al., 2010; Tran et al., 2019).

5.1.2 – ALT enzymes maintain their catalytic diversity in planta, and can be used to increase medium-chain fatty acid content in plant tissues
Although ALT enzymes display an unexpectedly broad substrate specificity range when expressed in *E. coli*, it was not known whether they would also exhibit this characteristic *in planta* (Feng & Cronan, 2009). Overexpression of *Arabidopsis* ALT1-4 in plant seeds and leaves demonstrated that ALTs do have a high level of catalytic diversity in plant tissues, and that their chain-length and oxidation state preferences in plant and bacterial cells generally agree. *AtALT1* and *AtALT4* display the same specificity in plant tissues and *E. coli*, giving rise to C12-C14 FAs and C6-C8 FAs, respectively, when expressed in *N. benthamiana* leaves. Their overexpression in *Arabidopsis* and *Camelina* seeds also resulted in accumulation of these MCFAs in the seed triacylglycerols of both species. Interestingly, overexpression of *AtALT4* also led to a large increase in 18:1 d9 fatty acid content in seed triacylglycerol, accompanied by a decrease in 18:2 and 18:3 fatty acids, which are the most abundant components of *Arabidopsis* and *Camelina* seed oil. While β-keto FAs produced by ALTs would not be incorporated into seed oils, small amounts of C10-14 FAs, which are trace products of *AtALT2* in *E. coli*, accumulated in *AtALT2*-overexpressing *Arabidopsis* seed triacylglycerols. Unexpectedly, seed-specific overexpression of *AtALT3* did not result in consistent changes to seed oil composition.

Co-expression of *AtALT2* or *AtALT3* with the *ShMKS1* β-ketoacid decarboxylase in *N. benthamiana* leaves resulted in the accumulation of 13:0 methylketone (MK) derived from 14:0 β-keto FAs produced by ALT activity. Significant amounts of 14:0 FA were also produced by *AtALT3* in *N. benthamiana*. Since *AtALT3* shows preference for 14-carbon β-keto fatty acyl-ACPs in *E. coli*, this indicates that the chain length and oxidation state preferences of *AtALT3 in planta* and in bacteria are similar. However, 14-carbon β-keto FAs are not made by *AtALT2* in bacteria, and the MK derivatives of 8:0 and 10:0 β-keto FAs, which are major products of *AtALT2* and secondary products of *AtALT3* in *E. coli*, were not detected in ALT-expressing *N.
*benthamiana* leaves. It is possible that the chain-length preferences of *AtALT2/3 in planta* and in *E. coli* do not overlap completely, but this could also have resulted from the low activity of *ShMKS1* towards <C12 chain-lengths (Yu et al., 2010). Additionally, despite *AtALT3* displaying similar substrate specificity to the wild tomato ALT enzyme *ShMKS2* in *E. coli*, MK accumulation in *N. benthamiana* leaves co-expressing *Arabidopsis* ALTs with *ShMKS1* was much lower than in leaves co-expressing *ShMKS1* and *ShMKS2*. These data suggest that a physical interaction critical for catalytic efficiency takes place between *ShMKS1* and *ShMKS2*, but cannot occur readily between *ShMKS1* and *Arabidopsis* ALTs.

5.1.3 - *Three short interacting amino acid motifs within the ALT hot dog-fold domain are core determinants of thioesterase chain-length and oxidation state specificity*

The work described in this thesis also provided substantial insight into the structural basis of the unusual substrate state specificities of ALT enzymes. Comparative motif searches of the hot dog-fold domain sequences of ALTs characterized in Chapter 2 and in previous literature identified two conserved amino acid motifs in ALTs that prefer 12-14 carbon β-keto and 3-hydroxy fatty acyl-ACP substrates in *E. coli*: a 31-CQH[G/C]RH-36 motif located on the central α-helix, and a 108-[K/R/Q][A/G][I/T][A/G]-111 motif (which typically reads 108-K[A/G][I/T]A-111) nearby on the underlying β-sheet. Exchanging sequence fragments between pairs of ALT proteins with different substrate specificities to create chimeric mutants, and profiling the products of these mutants in *E. coli*, confirmed that the identified motifs are key determinants of ALT substrate specificity alongside a third interacting region spanning residues 78-93 of the hot dog-fold β-sheet. Together, the 108-K[A/G][I/T]A-111 motif and aa78-93 grant ALTs the ability to act on partially reduced substrates, while the 31-CQH[G/C]RH-36 motif is largely responsible for a strong shift towards 12-14 carbon chains.
Characterization of additional chimeric mutants also demonstrated that aa64–67 of the ALT thioesterase domain influence selectivity for 3-hydroxy-acyl-ACP substrates, with a conserved Lys64 residue being required for activity towards 3-hydroxyacyl-ACPs. By manipulating the identified sequence features, the chain-length preferences of ALTs can be reliably directed towards C6-10 or C12-14 acyl-ACPs, and the products of partially reduced oxidation states can be eliminated or affinity for 3-hydroxyacyl-ACP substrates can be increased. Certain mutant ALTs also displayed increased productivity in E. coli as compared to their wild-type counterparts.

5.1.4. – Significance of findings and future research directions

The findings described in this thesis demonstrate that ALT enzymes can be useful tools for the sustainable production of medium-chain fatty acids and their chemical derivatives in plant or microbial systems. The ability of ALTs to produce such a wide range of high-value MCFAs, MK precursors, and, as determined here, 3-OH FAs, is particularly advantageous, since each of these compound classes lends itself to a distinct range of industrial applications. This work provides a foundation from which to begin exploring the use of recombinant ALT enzymes in biotechnological contexts, as core determinants of ALT chain-length and oxidation state specificity have been identified, and it is now clear that ALTs are amenable to heterologous expression and targeted mutagenesis in both plants and bacteria. Any changes in recombinant substrate specificity observed in bacteria would likely be carried over to a transgenic plant system, and vice-versa.

Also, by uncovering the vast array of substrate specificity profiles that have evolved among the ALT enzyme family, this work indicates that much can be learned about the roles of
medium-chain acyl lipids in plant secondary metabolism by characterizing the physiological functions of ALTs. Aside from the incorporation of C8-C14 FAs into seed oils of certain species as a result of FatB activity, and the insecticidal function of medium-chain MKs derived from ShMKS2 activity, little is known regarding what purposes MCFAs and their derivatives might serve as secondary, rather than primary, plant metabolites.

In summary, there are two logical directions for future investigation involving ALT-type thioesterases from plants: to engineer ALTs for efficient production of industrially valuable MCFAs and their derivatives in plants and microbes, and to elucidate the biological functions of these unusual acyl-ACP thioesterases. The following sections address both of these potential research currents.

5.2 – IMPROVING THE SPECIFICITY AND PRODUCTIVITY OF ALT ENZYMES FOR BIOTECHNOLOGICAL APPLICATIONS

5.2.1 – Developing mutagenesis and screening techniques to isolate ALT variants with improved activity towards desirable substrates

Although the results of targeted mutagenesis experiments described in Chapter 4 represent a significant advancement towards designing ALTs with product profiles tailored to specific industrial applications, some aspects of ALT substrate specificity, such as what drives affinity for β-ketoacyl-ACPs, remain unknown. Additionally, mutant ALT productivity was not always improved. Moving forward, it will be necessary to make simultaneous use of rational design methods, randomized mutagenesis and screening, and computational methods to engineer custom ALT variants that output high yields of desirable compounds.

The creation of ALT mutant libraries in microbial fatty acid auxotroph strains, like the 8:0
FA-dependent *ΔlipB* deletion strain of *E. coli* used by Hernández-Lozada *et al.* (2018) and the 14:0 FA-dependent YB218 strain of *Saccharomyces cerevisiae* (Duronio *et al.*, 2001) could enable screening for variants with increased specificity for certain fully reduced medium-chain fatty acids. Fluorometric screening with the lipophilic dye Nile Red could be used to quickly estimate and compare the productivity of ALT-expressing microbial strains, and eliminate less active variants from mutant libraries for directed evolution experiments (Huang *et al.*, 2009). As shown in Figure 5.1, treating spent culture media with Nile Red can distinguish between each of four K27(DE3) *E. coli* strains expressing ALT enzymes with varied total productivity (ng fatty acids / mL), as well as distinguish strains expressing active ALTs from a control strain harbouring an empty vector.

Efforts should also be directed towards the development of screening techniques capable of detecting increases in 3-OH FA and β-keto FA production. Such methods would greatly facilitate the creation of useful ALT variants for biotechnological applications, since a major advantage of ALT-type thioesterases is their unique ability to produce these compounds. Developing microbial strains dependent on supplementation with intermediates of fatty acid biosynthesis is probably not feasible; for instance, the instability of β-ketoacids precludes the creation of a useful β-keto FA auxotroph. Some colourimetric and fluorometric assays, though, may be adaptable for detection of 3-OH FAs and the MK derivatives of β-keto FAs. For example, the iodoform test can indicate the presence of MKs, and the 2,4-dinitrophenylhydrazine assay described by Zhou *et al.* (2016) is dependent on the reduction of ketones to alcohols. However, it may still be necessary to rely on GC-MS to screen mutant libraries for increased affinity towards β-keto- and 3-hydroxyacyl-ACPs.
Figure 5.1. Fluorescence intensity ($\lambda_{\text{ex}} = 550\text{nm}, \lambda_{\text{em}} = 625\text{nm}$) of Nile Red-treated culture media from K27(DE3) *E. coli* cultures expressing ALT-type thioesterases, with respect to total fatty acid content in culture media. Each strain was assayed in quadruplicate (n=4). Cultures of K27(DE3) *E. coli* strains harbouring *ALT* genes were induced and harvested as described in Chapter 2, Section 2.3, and total fatty acid content in culture media was quantified by GC-FID as described in Chapter 4, Section 4.3. 190 µL of culture supernatant was added to wells of a black, clear-bottomed microplate. 10 µL of a 0.10 mg/mL solution of Nile Red in DMSO was then added to each well and mixed. The microplate was incubated away from light at RT for 10 min., after which sample fluorescence intensity was measured with a Cytation 5 multi-mode microplate reader. Fluorescence intensity of a blank (190 µL liquid LB broth + 10 µL 0.10 mg/mL Nile Red in DMSO) was subtracted from all measurements. *BdALT1* = *B. distachyon* ALT1, *GbALT1* = *G. biloba* ALT1, *MtALT2* = *M. truncatula* ALT2, *ZmALT1* = *Z. mays* ALT1. Error bars represent standard deviation.

Additionally, taking advantage of rapid advancements in the sophistication and accuracy of machine learning models could simplify the process of identifying ALT mutants with desirable properties. Reliable machine learning models for substrate specificity prediction have already been applied to acyl-ACP thioesterases. EnZymClass, developed by Banerjee et al. (2022), can predict whether a Fat-type thioesterase will display short- and medium-chain, long-chain, or mixed chain-length specificity, and Jing et al. (2021) report a model capable of identifying sequence features that influence the specificity and productivity of FatB-type
thioesterases. Similar models could be trained on a dataset of characterized plant and bacterial single hot-dog fold acyl-ACP thioesterases, including the wild-type and mutant ALTs characterized in this work. However, the development of a model able to classify acyl-ACP thioesterases based on specificity for individual acyl chain-lengths (i.e. C8:0 versus C14:0) remains a challenge. There are too few characterized thioesterases to serve as a learnset for a task of this depth, which would likely require substrate specificity data from hundreds of enzymes. As a large enough database of characterized acyl-ACP thioesterases eventually becomes available, likely through directed evolution studies, generating such a model will be feasible. In time, it will be possible to simulate directed evolution computationally, and experimentally characterize only a small number of variants predicted to have desired substrate specificity or increased productivity.

While the work presented in this thesis provided qualitative information regarding the affinity of ALTs for substrates of various chain-lengths and oxidation states, the analysis of ALT enzyme kinetics will be necessary to fully understand how the properties of engineered ALTs compare to their natural counterparts, and thereby maximize their catalytic efficiency. Although the cost of procuring medium-chain acyl-ACP substrates directly is prohibitive, acyl-ACP synthesis methods outlined by Zhu et al. (2009) and Yang et al. (2023) could make assays to determine the kinetic parameters of ALTs feasible. These use the purified soluble acyl-ACP synthetase (AasS) from the marine bacterium *Vibrio harveyi* to generate acyl-ACPs *in vitro* using either fully reduced or 3-hydroxy MCFA as substrates. They could also potentially be adapted for synthesis of beta-ketoacyl-ACPs through the addition of a plant or bacterial KASI-type condensing enzyme to reactions, as long as care is taken to avoid conditions that promote spontaneous decarboxylation of β-ketoacyl chains.
5.2.2 – Examing protein-protein interactions (PPIs) involving ALTs to optimize the production of valuable compounds

i. ALTs with acyl carrier proteins

Since the hydrolysis of acyl-ACP thioesters is driven by interaction between the ACP and the thioesterase, both protein partners necessarily influence enzyme kinetics and specificity. Recent research by Yang et al. (2023) demonstrates that ACP structure exerts a prominent effect on acyl-ACP thioesterase specificity. This study showed that the chain-length preferences of FatB2 from Cuphea pulcherrima FatB2 vary considerably when interacting with either acyl-ACPs from the sister species Cuphea lanceolata, or acyl-ACPs from E. coli. This was attributed to structural differences between ACPs of bacterial and plant origin that influence both recognition of the ACP by the thioesterase, and the ability of ACP to adjust conformationally to accommodate various acyl chain lengths (Yang et al., 2023).

Given the great extent to which ACP conformation can alter thioesterase function, rational design of ACP isoforms is likely as critical for boosting the production of medium-chain oleochemicals in plants and microbes as engineering the acyl-ACP thioesterases themselves. Maximizing the output of desired FAs by heterologously expressed ALTs will likely require the expression of recombinant ACPs engineered to favour thioesterase selectivity towards certain chain-lengths and oxidation states. Simulating mutations to regions of plant and bacterial holo- and acyl-ACPs in silico that form part of the thioesterase-ACP interaction interface, and docking these modified ACPs with modelled ALT enzymes, could guide efforts to optimize ACP-ALT interactions.
ii. ALTs with one another, and other single hot-dog fold proteins

Computational modelling of the ALT thioesterase domain demonstrated that these proteins likely form tetrameric “dimer-of-dimers” active assemblies similar to their distant bacterial homologues (Figs. 4.1, 4.2). Overexpression of one ALT at a time in a heterologous system, as in the experiments described in this thesis, would lead to homo-oligomer formation. However, some proteins of the hot-dog fold family can form heterodimers, with activity that differs from homodimeric assemblies (Sacco et al., 2007; Pidugu et al., 2009). It is possible that ALTs with overlapping expression patterns (i.e. AtALT3 and other Arabidopsis ALTs, M. truncatula ALT1/2) naturally form heterodimers in planta. Since hot-dog fold domain structure is so well-conserved across kingdoms, heterologously expressed ALTs might also dimerize with endogenous ALT-like thioesterases in plant plastids or with native single hot-dog fold thioesterases in bacterial cells, which may attenuate enzyme productivity or alter specificity.

Affinity purification and size-exclusion chromatography experiments performed in the Rowland lab has shown that ALTs are capable of forming hetero-oligomers (Khalil et al., unpublished data), and it would be worthwhile to also assess the stability of homo- versus hetero-oligomers through techniques such as differential scanning calorimetry. Overexpressing pairs of ALTs in E. coli or N. benthamiana leaves to promote hetero-oligomer formation could give rise to unusual product profiles, perhaps even ones that fall outside the catalytic groupings described in Chapter 2. Depending on how ALTs influence one another, the expression of two engineered ALTs at a time could lead to much more efficient production of compounds of interest than overexpression of a single ALT.
iii. ALTs with other potential partner enzymes

Aside from the necessary interactions of ALT monomers with one another and with ACPs that deliver substrates, it is not known whether ALTs participate in other PPIs. ALTs might interact with other plastid-localized proteins that convert their products to other chemical forms, or transport their acyl chain products out of the plastid. Determining whether such interactions take place, and if so, optimizing them through protein engineering, could increase the yield of medium-chain fatty acids and their derivatives resulting from ALT activity. Techniques to identify potential PPIs that involve ALTs are discussed further in Section 5.3.1, in the context of elucidating the physiological roles of ALT enzymes.

To date, the ShMKS1 decarboxylase is the only enzyme proven to act directly on the products of an ALT-type thioesterase (ShMKS2) in vivo (Yu et al., 2010; 2014). Although data presented in Chapter 3 of this thesis does not provide direct evidence of physical interaction between ShMKS1 and ShMKS2, certain findings suggest they may form a metabolon that increases the catalytic efficiency of ShMKS2 and allows for conversion of its β-keto FA products to MKs. Co-expression of ShMKS1 with ShMKS2 in N. benthamiana leaves results in the production of high levels of 13:0 MK, while co-expression of ShMKS1 with Arabidopsis ALT3, which has similar substrate preferences to ShMKS2 in E. coli bacteria, does not (Fig. 3.7). Expression of ShMKS2 on its own leads to 14:0 FA accumulation in leaves, which approximately doubles when ShMKS1 is co-expressed (Fig. 3.7). Although ShMKS1 is also capable of hydrolyzing acyl-ACP thioesters, this occurs at an extremely slow rate that could not account for an increase of this magnitude (Fig. 3.7) (Yu et al., 2010). However, there is no published work investigating the possibility of physical interaction between ShMKS1 and ShMKS2. In vitro affinity purification methods such as a pull-down assay, and techniques for
visualizing PPIs in live cells, for instance, bimolecular fluorescence complementation, could be used to detect whether ShMKS1 and ShMKS2 interact (Maple & Møller, 2011). Even if the two proteins do not interact directly, ensuring that they remain in close proximity, for example, by expressing both as a fusion protein, may facilitate substrate channeling and improve yield of medium-chain MKs in transgenic plant tissues.

5.2.3 – Using ALT enzymes to engineer transgenic plants with novel seed oil composition

Hexanoic acid (6:0 FA) and its hexyl ester derivatives have widespread uses in the metalworking, cosmetic, and food industries as components of cleansers, emulsifiers, flavourings, and dyes (Cheon et al., 2014). Hexanoic acid is also a precursor for the synthesis of hexanol, which is often used in biodiesel blends (Cheon et al., 2014; Verser & Park, 2012). As discussed in Chapter 3.5, efforts to increase MCFA content in seed oil typically involve seed-specific overexpression of highly active specialized FatB-type thioesterases, however, a FatB enzyme with significant activity towards 6:0 acyl-ACP has not been discovered to date. Certain ALTs, on the other hand, such as AtALT4 and MtALT1, do have affinity for 6:0 acyl-ACP substrates, and ALT overexpression can lead to increased 6:0 FA content in seed triacylglycerols as shown in Chapter 3. The activity of 6:0-specific ALTs could therefore be leveraged to engineer highly novel transgenic crop varieties with hexanoate-rich seed oils.

Increasing the proportion of hexanoic acid in seed oil to industrially significant levels, while still maintaining seed viability, will necessitate the manipulation of several genes involved in de novo fatty acid synthesis, plastidial fatty acid export, and TAG biosynthesis. ALT variants could be engineered for greater affinity for 6:0 chains, and ACP isoforms that favour thioesterase selectivity for 6:0 chains could also be expressed in seeds alongside these
ALT variants, as described above in Section 5.2.1. Enzymes involved in the re-activation of prematurely cleaved acyl chains, such as *Arabidopsis* AAE15 and AAE16, would also need to be inactivated to prevent 6:0 acyl chains released by the seed-expressed ALT from immediately re-entering this cycle (Tjellström *et al.*, 2013). It would also be of interest to overexpress central positive regulators of plant oil biosynthesis, such as the *WRINKLED1* transcription factor, which directly increases the expression of key fatty acid biosynthetic genes (Kong *et al.*, 2019), and assess whether this increases MCFA yields resulting from ALT activity and leads to more efficient incorporation of these into storage lipids.

As shown in Chapter 3.4, overexpression of 6:0-producing *AtALT4* in *Arabidopsis* and *Camelina sativa* seeds leads to defects in seed morphology and decreased seed viability (Figs. 3.2, 3.5). This suggests that hexanoyl-CoA thioesters may be directed to membrane lipid biosynthesis in addition to seed TAG, but this could be overcome through co-expression of 6:0-specific ALTs with acyltransferases that efficiently incorporate this chain-length into TAG. Suitable acyltransferases may be difficult to find within the plant kingdom, however; among the many oilseed plants with known seed TAG composition, 6:0 FA, if present, is only a minor component. Plant acyltransferases with specificity for other medium-chain lengths, such as coconut LPAAT2 and palm kernel DGAT, could possibly be engineered for activity towards 6:0 chains (Davies *et al.*, 1995; Aymé *et al.*, 2015). Alternatively, enzymes of animal or bacterial origin could be expressed alongside 6:0-specific ALTs, since TAG species containing 4:0-6:0 FA occur more commonly in animals and prokaryotes than plants. Heterologous expression of the cow DGAT enzyme that efficiently incorporates these chain-lengths into cows’ milk, for example, could be attempted (Marshall & Knudsen, 1976).
In some plants such as *Physaria* spp. and castor (*Ricinus communis*), unusual long-chain and very-long chain hydroxy fatty acids bearing mid-chain hydroxy groups are major components of seed triacylglycerols (reviewed in Cahoon & Li-Beisson, 2020). An ω-hydroxy fatty acid, kamlolenic acid, is a component of the seed oil of *Mallotus* spp. (Smith *et al.*, 2013). Some oilseed plants may therefore be able to incorporate unusual 3-hydroxy fatty acids into their triacylglycerols as well. It may be of interest to investigate the effects of seed-specific overexpression of ALT-type thioesterases that primarily produce 3-hydroxy fatty acids, such as *MtALT2* or *VvALT1* characterized in this thesis, on the seed triacylglycerol composition of genetically amenable oilseed species.

5.3 – UNDERSTANDING THE BIOLOGICAL ROLES OF ALT ENZYMES

5.3.1 – Tracing how the products of ALTs are metabolized in planta through lipidomics, reverse genetics, and protein-protein interaction mapping

With the exception of *ShMKS2*, it is not known whether the products of ALT enzymes are stored directly in plant tissues, converted to other chemical forms, or incorporated into larger lipid molecules. Comparing the lipid composition of stable *alt* loss-of-function lines and wild-type plants could identify products that result from ALT activity in planta, and I have generated a series of *alt* knockout lines in *Arabidopsis* for this purpose. This mutant collection, created through CRISPR-Cas9 gene editing, includes single knockouts targeting each of the four endogenous *Arabidopsis* ALT genes, double and triple knockouts targeting combinations of genes with overlapping expression patterns, and a quadruple knockout deficient in ALT function. Some of these lines have already been used for experimentation in the Rowland Lab. As determined in Pulsifer *et al.* (2014), the expression patterns of *AtALT1* and *AtALT2* indicate
that they may be involved in the synthesis of cuticular wax and root suberin monomers, respectively (Pulsifer et al., 2014). However, chemical analysis of leaf and stem cuticular wax and root suberin in alt knockouts was not indicative of changes to the composition of these polymers from wild-type plants (Morewood et al., unpublished data).

Untargeted lipidomics analyses with highly accurate and sensitive mass spectrometry techniques, such as UHPLC-MS/MS, could be used to trace how the products of ALTs are stored in plant tissues (Kehelpannala et al., 2020). Components that are less abundant or missing in the lipidomes of alt-deficient plants compared to control plants would likely originate from ALT activity. While such techniques allow for the analysis of intact lipid macromolecules, it may be sufficient to simply compare the chain-length composition of acyl lipids from alt knockout and wild-type plants. Total lipid extracts can be separated into individual lipid classes through preparative thin-layer chromatography, and each lipid type can be transmethylated and analyzed by GC-MS to assess chain-length composition (Hölzl & Dörmann, 2021; Wang & Benning, 2011). Due to the highly specific expression patterns of certain ALTs (i.e. AtALT4 is localized exclusively to anthers, AtALT1 is largely restricted to stem epidermis), it might be necessary to analyze lipids from dissected tissues, rather than whole organs, in order to detect differences between sample groups.

Investigating the functions of genes that are co-expressed with ALT genes might also provide clues as to how the products of ALTs are transported and metabolized in planta. For example, Arabidopsis gene co-expression networks indicate that the expression pattern of AtALT2 strongly overlaps with several uncharacterized genes belonging to the GDSL-like lipase/esterase, α/β hydrolase, HXXXD-type acyltransferase, lipid-transfer protein, and S-adenosyl-L-methionine-dependent methyltransferase families, which could potentially be
involved in the transport or chemical modification of the β-keto FA products of AtALT2 (Obayashi et al., 2022). Interestingly, AtALT2 expression also overlaps with MLP-LIKE PROTEIN 165 (MLP165), a gene encoding a member of the multifunctional, labile major latex protein family that is known to bind acyl lipids (Sterlin et al., 2019). MLP165 could therefore be involved in the transport or metabolism of products that arise from AtALT2 activity (Sterlin et al., 2019). Performing small-molecule pull-down assays with protein extracts from transgenic plant tissues overexpressing both AtALT2 and MLP165, similar to the experiments described by Sterlin et al., to identify ligands interacting with MLP-like proteins, could determine whether this is the case. If so, MLP165 may be an interesting candidate for further investigation.

Searching for novel PPIs involving ALTs could also provide more insight into how their products are transported or modified. Yeast-two-hybrid screening of Arabidopsis cDNA libraries could be conducted using ALTs as baits to identify potential PPIs (Mattioli & Melotto; 2018). However, the structural similarity of AtALT1-4 could be a complicating factor, potentially causing certain prey proteins to interact with all four Arabidopsis ALTs when they would only interact with a single ALT or subset in planta. A technique to identify PPIs directly in plant samples would avoid this problem, but would need to be highly sensitive and suited for the detection of transient interactions in order to capture interactions between ALTs and proteins other than ACP. Proximity labelling methods such as TurboID, which is optimized for the capture of transient PPIs involving a protein of interest in living cells, and can also be performed in specific cell compartments (i.e. the plastid), may be suitable (Mair et al., 2019; Wurzinger et al., 2022).

5.3.2 – Chemical profiling of volatile compounds emitted from plants deficient in ALT activity
Volatile organic compounds (VOCs) produced by plants, which can be released into the atmosphere or stored in specialized secretory structures, mediate a vast array of plant-environment interactions. They can function as attractants for pollinators and organisms that spread seeds, as defensive compounds against herbivores either directly or by attracting the natural predators and parasites of herbivorous pests, and as signals for neighbouring plants to activate defensive responses or promote developmental processes such as fruit ripening (reviewed in Tholl et al., 2016). Several plants also emit certain VOCs under abiotic stress conditions, which are hypothesized to scavenge reactive oxygen species and mitigate thermal stresses (reviewed in Tholl et al., 2016; Loreto & Schnitzler, 2010). The enzymatic products of ALTs, particularly \(<\text{C10 MCFAs}\) and the MK derivatives of \(\beta\)-keto fatty acids, are volatile, and these compounds are known components of headspace volatiles from many plant species of commercial importance including maize, rice, citrus, grapevine, coffee, cinnamon, cloves, and several members of the rue family (Buttery & Ling, 1984; Wang & Tan, 2019; Prokopy et al., 1998; Lazazarra et al., 2018; Forney & Markovetz, 1971). The ability of ALT enzymes to generate these compounds suggest that they may be involved in the production of VOCs emitted by flowers, roots, and / or leaves, and, by extension, occupy roles in plant defence and / or signalling.

Comparing the composition of volatile emissions from \(alt\) loss of function mutants and wild-type plants could indicate whether ALT activity contributes to a plant’s VOC profile. Such experiments would require a setup for the capture of headspace volatiles from whole plants or plant parts onto an adsorbent matrix, so that individual compounds within the complex headspace volatile mixture could be subsequently identified and quantified by GC-MS. Since the products of ALTs would likely be minor components of plant volatile emissions, and a large
number of genetically amenable plant species, including Arabidopsis, are low-emitting plants, a highly sensitive technique for volatile capture and concentration is necessary. Either static headspace sampling through solid-phase microextraction, dynamic headspace sampling through closed-loop stripping, would be suitable, since closed capture systems such as these do not introduce contaminants from ambient air nor allow plant volatiles to escape (reviewed in Tholl et al., 2016). Both techniques have been used successfully to study the floral volatile profile of Arabidopsis flowers (Donath & Boland, 1995; Chen et al., 2003; Tholl et al., 2005; reviewed in Tholl et al., 2006). It may also be informative to subject plants to artificial wounding or insect feeding prior to volatile capture, as ALT genes may be responsive to wounding or herbivory. The promoters of multiple ALT genes contain W- and WB-box elements that suggest that they might be responsive to physical wounding (Tran et al., 2019; Dong et al., 2003).

Another approach could be to take ALT genes from non-model plants that are implicated in the production of floral VOCs and overexpress these Arabidopsis, which does not naturally have a strong floral scent profile. Szenteczki et al. (2022) identified two ALT genes that are up-regulated during anthesis in the monocotyledonous flowering plant Arum maculatum, alongside genes encoding distant homologues of the wild tomato ShMKS1 decarboxylase. 2-heptanone (7:0 MK), which can be synthesized from 8:0 β-keto fatty acid, is a major component of the unappealing floral scent of A. maculatum. Overexpressing A. maculatum ALT genes with their putative MKS1-like partners in Arabidopsis under the control of strong constitutive or flower-specific promoters, and analyzing the VOC profiles of these transgenic plants, could lead to intriguing results. Linking 2-methylketone production to ALT activity in non-Solanaceous
plants would indicate that this function arose either prior to divergence of the Solanaceae, or separately in multiple plant taxa, contrary to what is currently theorized (Auldridge et al., 2011).

5.3.3 – Exploring the potential roles of ALT genes in plant defence and other plant-environment interactions

In addition to their ability to produce compounds with insecticidal and antimicrobial properties in E. coli and transgenic plant tissues, there is transcriptome-based evidence that ALT enzymes from certain plant species may participate in defence against pathogenic microbes and insect feeding as hypothesized. However, taking into consideration the widespread activity profiles of ALTs from diverse plants, the evolution of multiple paralogues in many plant species, and other available transcriptomic data, members of the ALT family likely carry out many more many physiological functions. For example, the expression patterns of some ALTs characterized in this thesis implicate them in interactions with beneficial microbes, or in abiotic stress responses. As described in more detail below, ALT genes from the model legume Medicago truncatula are up-regulated during arbuscular mycorrhizal symbiosis. Several ALT genes from monocot plants, including BdALT1 from Brachypodium distachyon (Bradi1g28810), OsALT1 from Oryza sativa subsp. indica (Os04g46730), and ZmALT2 from Zea mays (GRMZM2g077811) are associated with submergence and drought stress, with significantly increased transcript abundance in submergence- and drought-tolerant cultivars (Rivera-Contreras et al., 2016; Jung et al., 2010; Muthurajan et al., 2018; Campbell et al., 2015; Luo et al., 2010). Further insight into the biological roles of ALTs can be obtained by generating transgenic lines of genetically amenable plant species that either overexpress ALT genes, or are deficient in ALT function, and comparing the response of transgenic and wild-type plants to the biotic and abiotic stimuli with which these ALT genes are associated. A few avenues to
investigate the potential roles of ALTs in plant-microbe and plant-insect interactions, using model plant species, are summarized in the following subsections. It must be emphasized, however, that characterizing the role of ALTs in a single biological process or plant taxon will not provide a complete understanding of the physiological roles of all ALTs throughout the plant kingdom.

i. Investigating the associations of root-expressed ALTs from the model legume Medicago truncatula with pathogen defence and arbuscular mycorrhizal symbiosis

Several broadly expressed ALT genes explored in this work, such as A. thaliana ALT3, M. truncatula ALT2 (Medtr5g056390), O. sativa ALT1, and G. max ALT1 (Glyma01g080400) have high transcript abundance in both roots and aerial plant parts according to publicly available data, and other ALT genes, including Selaginella moellendorffii ALT1 (SELMODRAFT_89469) and A. thaliana ALT2 (At1g35250), are mainly root-specific (Huang et al., 2015; Pulsifer et al., 2014). As shown in Chapter 2, when expressed in E. coli, these root-expressed ALTs all produce significant amounts of the β-keto FA precursors of medium-chain MKs, which have antagonistic activity towards several phytopathogens (Huang et al., 2011; Zheng et al., 2013; Neri et al., 2007; Yuan et al., 2012; Wu et al., 2019; Riu et al., 2022). This leads to the hypothesis that ALTs localized to roots may generally function in pathogen defence. Transcriptomic data from M. truncatula shows a particularly strong association between MtALT2 and response to root pathogen infection. MtALT2 expression increases significantly in roots in the early stages of infection by the phytopathogenic bacterium Ralstonia solanacearum and fungus Macrophomina phaseolina (Carrère et al., 2021; Damiani et al., 2012; Mah et al., 2012). MtALT2 is also much more highly expressed in the mutant line Mtnf-ya1-1, which is
highly resistant to the root pathogen oomycete *Aphanomyces euteiches*, compared to susceptible plants (Rey *et al.*, 2016).

Given the established connection between increased *MtALT2* expression and resistance to bacterial and fungal root pathogens, it may be of interest to generate *Mtalt2* knockout mutants (either *Tnt1* retrotransposon insertion mutants or CRISPR-Cas9 edited) in both pathogen-susceptible and pathogen-resistant backgrounds, and experiment with these to determine whether loss of *MtALT2* function impacts infection susceptibility (Kaur *et al.*, 2021; Tadege *et al.*, 2008; Confalonieri *et al.*, 2021). Inoculating the roots of *Mtalt2* mutants and wild-type plants with *Ralstonia solanacearum*, *Macrophomina phaseolina*, or *Aphanomyces euteiches*, and assessing infection spread through both microscopy and macroscopic symptom-scoring techniques (Turner *et al.*, 2009; Gaige *et al.*, 2010; Rey *et al.*, 2016), could reveal whether *MtALT2* actively participates in defence against pathogenic microbes.

Interestingly, *MtALT1* (Medtr5g056230), which is expressed at low basal levels in roots, is strongly up-regulated in non-colonized mycorrhizal root cortical cells directly adjacent to cells that have been colonized by arbuscular mycorrhizal (AM) fungi (Carrère *et al.*, 2021). This suggests that the 6-10 carbon FA and β-keto FA products of *MtALT1* (or their derivatives) may signal AM fungi to promote root cell colonization, or support the fungi nutritionally, such that the cell begins accumulating these products in preparation for fungal colonization. *MtALT2* expression is also significantly lower in mutant plants where the RAM1 transcription factor, the central regulator of AM symbiosis-related genes, has been knocked out (data generously provided by Dr. Allyson MacLean). It would be worthwhile to examine whether colonization of the root cortex by AM fungi is limited in *alt* loss-of-function mutants, or whether overexpression of endogenous *ALT* genes leads to changes in periarbuscular membrane
structure, alterations in root system architecture, or other growth-related phenotypes. It would also be of interest to investigate whether ALT genes in *M. truncatula* are directly regulated by RAM1, or by one or more transcription factors that are downstream targets of RAM1, such as WRINKLED 5a, a master regulator of fatty acid biosynthesis during AM symbiosis (Lugenbuehl et al., 2017). Since little is known about the regulation of ALT genes in general, any efforts to identify associated transcription factors would yield valuable information.

**ii. Examining the effects of ALT activity on root microbiome composition**

Certain volatile compounds produced by ALTs, and their derivatives, are also emitted by plant-associated bacteria to signal one another, elicit physiological responses from plants that may be beneficial or detrimental to plant health, and exert antagonistic effects against microbes that harm the plant host (reviewed in Weisskopf et al., 2021). For instance, the methylketone derivatives of 6:0 and 8:0 β-keto fatty acids, which are produced by certain ALTs including root-expressed *Arabidopsis* ALT2, are also emitted by beneficial phytobacteria that colonize the plant rhizosphere and root (Riu et al., 2022). The roles of VOCs emitted by plant-associated *Bacillus* species are particularly well-studied. 2-nonanone emitted by root-associated *Bacillus amyloliquifaciens* subsp. *plantarum* and *Streptomyces griseus* has been shown to elicit tomato plant immunity against the pathogenic bacterium *Pseudomonas syringae* var. *tomato* (Riu et al., 2022). Also, 2-nonanone is present in volatile mixtures produced by *Bacillus* spp. isolates from avocado rhizosphere capable of inhibiting the fungal pathogen *Fusarium kuroshium*, by VOC emission, whereas 2-nonanone is not produced by isolates that cannot inhibit *F. kuroshium* growth through VOC emission alone (Guevara-Avendaño et al., 2020). 2-heptanone is also a major component of the VOC mixture emitted by the rhizobacterium *Bacillus* sp. JC03, which promotes plant root biomass accumulation by regulating synthesis of the phytohormones
auxin and strigolactone (Jiang et al., 2019). Methylketones ranging from 7-11 carbons in chain-length have also been shown to modulate the expression of bacterial genes involved in quorum sensing (Plyuta et al., 2014).

Secondary metabolites exuded by plant roots shape the composition and function of the root and rhizosphere microbiome (reviewed in Koprivova & Kopriva, 2022). Given the overlap between the product profiles of ALTs and the chemical composition of phytobacterial VOC mixtures, it would be of interest to investigate whether ALT gene activity influences microbiome composition and function. Just as bacterial VOCs can elicit physiological responses from plants, these same compounds, when emitted by a plant, could be involved in communication with soil bacteria. Loss of ALT gene function might impede a plant’s ability to associate with certain species that are normally present in the root and / or rhizosphere microbiome, thereby leading to changes in plant growth physiology and immunity to disease. Arabidopsis alt2 alt3 double knockout mutant lines deficient in root-expressed ALT activity could be used for experimentation in this area. To compare the root and rhizosphere microbiome composition of wild-type and alt knockout Arabidopsis plants, root endophytes and rhizosphere can be isolated from plants grown in natural soils through methods such as those described by Lundberg et al. (2012) or Bodenhausen et al. (2013). This would be followed by microbial DNA extraction, PCR amplification of the 16S ribosomal RNA gene, and high-throughput sequencing of amplified PCR product to determine the taxonomic composition of the microbiome (Bodenhausen et al., 2013).

iii. Determining whether a loss of ALT gene function influences plant susceptibility to insect herbivory
While ShMKS2 is the only ALT-type thioesterase with a confirmed role in defence against insect feeding, several other ALT enzymes (the Group 1 ALTs) can also produce the β-keto FA precursors of 11-15 carbon MKs, which are toxic to herbivorous insects and mites (Fridman et al., 2005). According to available transcriptomic data, some genes encoding these ALTs are either up-regulated in response to attack by insect herbivores, or expressed more highly in cultivars that are less susceptible to infestation. For instance, basal OsALT1 transcript abundance is significantly in higher in cultivars that are resistant to feeding by the brown planthopper and lepidopteran pests than in susceptible cultivars, and OsALT1 expression also increases following insect attack (Li et al., 2017; Qi et al., 2011). SmMKS2-2, a Group 1 ALT-encoding gene from Solanum melongena that was characterized following publication of the work summarized in Chapter 2, is also induced by insect herbivory and artificial wounding (Khuat et al., 2019). Group 1 ALTs from species other than S. hirsutum subsp. glabratum could therefore function in the biosynthesis of insecticidal compounds. Arabidopsis thaliana could be used as a system to investigate this, since AtALT3 is expressed in all aerial plant parts and has an activity profile similar to ShMKS2, generating the 12-16 β-keto fatty acid precursors of insecticidal 11-15 carbon MKs in E. coli (Pulsifer et al., 2014; Klepikova et al., 2016).

Insect feeding assays on alt3 knockout mutants and wild-type control plants could be conducted with, for example, Pieris rapae or Plutella xylostella moth larvae, which are specialists of the Brassicaceae. Susceptibility to herbivory could be assessed by measuring larval weight gain, larval mortality, and the extent of leaf defoliation by larvae placed on either alt knockout or wild-type plants (no-choice tests), and by scoring larval preference in addition to the above parameters when larvae are offered plants from both groups (two-choice tests) (Loivamäki et al., 2008). Since the products generated by ALT enzymes in E. coli and
transgenic plant tissues are volatile, as discussed in Section 5.2.2, it could also be informative to study whether loss of ALT function interferes with the plant’s ability to attract beneficial insect bodyguards in response to insect herbivory. Parasitoids and predators of insect pests will often respond to volatile cues emitted by plants in response to herbivore-induced tissue damage, and studies of their foraging behaviour around damaged transgenic and non-transgenic plants can indicate whether a plant gene of interest is involved in the biosynthesis of these chemical cues. *Cotesia vestalis* and *Cotesia rubecula* wasps, which are frequently used in such studies, are endoparasitoids of *P. rapae* and *P. xylostella* larvae (Tanaka *et al.*, 2018; McDowell *et al.*, 2010), and foraging-choice experiments could therefore also be conducted using *Arabidopsis* *alt* knockout mutants.

**Concluding Remarks**

The work described in this thesis characterizes the ALT enzyme family as an unusual, catalytically diverse class of plant acyl-ACP thioesterases. The findings presented here show that ALTs hold promise as metabolic engineering tools to produce valuable medium-chain oleochemicals in biological systems, and also provide insight into their evolution and potential biological roles. Homology models of ALT proteins reflect their ancient origins, demonstrating that they are more structurally similar to prokaryotic single hot-dog fold thioesterases than other plant acyl-ACP thioesterases. Screening the substrate specificities of ALTs from species across the plant kingdom, including a green microalga, indicated that ALTs likely descend from a prokaryotic, primary metabolic palmitoyl-ACP thioesterase, which later evolved the capability to act on intermediates of the *de novo* fatty acid biosynthesis cycle. ALT enzymes cleave acyl-
ACP thioesters prematurely to release medium-chain fatty acids, β-keto fatty acids, and, as first demonstrated here, 3-hydroxy fatty acids. The ALTs characterized herein could be categorized into four groups based on their general substrate chain-length and oxidation state preferences when expressed in *E. coli*. However, each ALT still gave rise to a unique product profile, which is indicative of an exceptionally high level of naturally occurring catalytic diversity among these thioesterases. Given the tendency for many plant species to possess paralogous ALTs with vastly different substrate preferences, it is likely that ALTs occupy several distinct roles in specialized metabolism. Future experimentation with transgenic plants to elucidate the biological functions of ALTs will likely reveal that several of their functions are taxon-specific, and perhaps even species-specific.

As shown in this work, the overexpression of ALT enzymes in plant tissues leads to increased production of medium-chain fatty acids and the β-keto fatty acid precursors of methylketones, demonstrating that ALTs can be used to increase medium-chain fatty acid content and manipulate acyl chain-length composition in either transgenic plants or microbes. Key structure-function relationships in ALT enzymes were also uncovered through targeted mutagenesis experiments. Three short interacting regions of amino acid sequence within the thioesterase domain core were found to influence ALT substrate specificity, and by exchanging these sequences between ALTs with different substrate preferences, chain-length and oxidation state specificity can be predictably altered. Certain features of ALTs give them a unique advantage over other plant thioesterases as potential sustainable sources of medium-chain oleochemicals. ALTs can generate the β-keto fatty acid and 3-hydroxy fatty acid precursors of industrially valuable methylketones and polyhydroxyalkanoates, which must normally be obtained through chemical conversion of free fatty acids, in a single catalytic step. Also, some
ALTs naturally show preference for acyl chain-lengths, such as 6:0, that are not favoured by other plant acyl-ACP thioesterases characterized to date. By providing a foundation from which to engineer ALT variants with increased productivity and desirable substrate specificity profiles, this research is a step towards harnessing the biotechnological potential of the ALT enzyme family.
APPENDIX 1. Alignment of the complete protein sequences of ALT-type thioesterases characterized in this thesis or in other literature. The red line indicates the end of a predicted N-terminal plastid targeting sequence and the start of the hot-dog fold thioesterase domain. Sequences were aligned using ClustalW (Larkin et al., 2007). Identical residues are highlighted in black, while chemically similar residues are highlighted in grey. At = Arabidopsis thaliana, Bd = Brachypodium distachyon, Cs = Cannabis sativa, Cr = Chlamydomonas reinhardtii, Gb = Ginkgo biloba, Gm = Glycine max, Mt = Medicago truncatula, Os = Oryza sativa subsp. indica, Sh = Solanum habrochaites subsp. glabratum, Sm = Solanum melongena, Sem = Selaginella moellendorffii, Vv = Vitis vinifera, Zm = Zea mays.
**APPENDIX 2.** PCR primers used to assemble genetic constructs described in this thesis.

**Table S2.A** – PCR primers used to assemble genetic constructs for overexpression of ALT-like thioesterases in *Arabidopsis thaliana* seeds and *Nicotiana benthamiana* leaves

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'–3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtALT1-pBinGly-F</td>
<td>5’-AAAGAATTCATGCTTAAAGCTACCGGCACAG-3’</td>
</tr>
<tr>
<td>AtALT1-pBinGly-R</td>
<td>5’-AAAGCTAGCTCAATATTCGACAACGTGTTGACA-3’</td>
</tr>
<tr>
<td>AtALT2-pBinGly-F</td>
<td>5’-AAAGAATTCATGCTTAAAGCTACCGGCACAG-3’</td>
</tr>
<tr>
<td>AtALT2-pBinGly-R</td>
<td>5’-AAAGCTAGCTCAATATTCGACAACGTGTTGACA-3’</td>
</tr>
<tr>
<td>AtALT3-pBinGly-F</td>
<td>5’-AAAGAATTCATGCTTAAAGCTACCGGCACAG-3’</td>
</tr>
<tr>
<td>AtALT3-pBinGly-R</td>
<td>5’-AAAGCTAGCTCAATATTCGACAACGTGTTGACA-3’</td>
</tr>
<tr>
<td>AtALT4-pBinGly-F</td>
<td>5’-AAAGAATTCATGCTTAAAGCTACCGGCACAG-3’</td>
</tr>
<tr>
<td>AtALT4-pBinGly-R</td>
<td>5’-AAAGCTAGCTCAATATTCGACAACGTGTTGACA-3’</td>
</tr>
<tr>
<td>AtALT1-TOPO-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT1-TOPO-R</td>
<td>5’-TCAATATTCGACAACGTGTTG-3’</td>
</tr>
<tr>
<td>AtALT2-TOPO-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT2-TOPO-R</td>
<td>5’-TCAATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT3-TOPO-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT3-TOPO-R</td>
<td>5’-TCAATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT4-TOPO-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>AtALT4-TOPO-R</td>
<td>5’-TCAATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>ShMKS1-ORF-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>ShMKS1-ORF-R</td>
<td>5’-TTATTTTATACAGCTGCTTAG-3’</td>
</tr>
<tr>
<td>ShMKS2-ORF-F</td>
<td>5’-CACCATGCTTAAAGCTACCG-3’</td>
</tr>
<tr>
<td>ShMKS2-ORF-R</td>
<td>5’-CTAGAGATGATGTTGATACCGC-3’</td>
</tr>
</tbody>
</table>

**Table S2.B** – PCR primers used to assemble genetic constructs encoding chimeric ALT enzymes for heterologous expression in *E. coli* bacteria

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtALT34-InFusion-For</td>
<td>5’-AATGGGTCGAGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT34-R35M-Rev</td>
<td>5’-TCATGCTGAGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT34-R35M-For</td>
<td>5’-CAACACCGGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT34-InFusion-Rev</td>
<td>5’-GACGGGAGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT34-M35R-Rev</td>
<td>5’-ATGGAGATGAGTCCATGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT4-M35R-For</td>
<td>5’-AAACGGGAGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT4-InFusion-Rev</td>
<td>5’-GACGGGAGATCCTCATGAGGTTGTTGAGGTTGGA-3’</td>
</tr>
<tr>
<td>AtALT4-CDS-For</td>
<td>5’-ATGGAGATGAGTCCATGAGGTTGGA-3’</td>
</tr>
<tr>
<td>Genotype</td>
<td>Primer Sequence</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td><em>AtALT4-B-Overlap-Rev</em></td>
<td>5'-AGCTATTCTTTTGCCTCAATATAACCT-3'</td>
</tr>
<tr>
<td><em>AtALT4-B-Overlap-For</em></td>
<td>5'-GAATAGCTGTATGCGTTTGAACAAGCACC-3'</td>
</tr>
<tr>
<td><em>AtALT4-CDS-Rev</em></td>
<td>5'-TCAACTGTCGTGTTTTGCC-3'</td>
</tr>
<tr>
<td><em>MtALT1-CDS-For</em></td>
<td>5'-ATGTTGAGTTTCTATGACGTG-3'</td>
</tr>
<tr>
<td><em>MtALT1-B-Overlap-Rev</em></td>
<td>5'-TGCTGTGGCCCTGGCTCCAAAAACAGGC-3'</td>
</tr>
<tr>
<td><em>MtALT1-B-Overlap-For</em></td>
<td>5'-AGGCCACAGCGATGGAGTTTTGATATAAAAACTAC-3'</td>
</tr>
<tr>
<td><em>MtALT1-CDS-Rev</em></td>
<td>5'-TTAAGAGCTCGTCTCCACCAATAC-3'</td>
</tr>
<tr>
<td><em>MtALT1-InFusion-For</em></td>
<td>5'-AATGGGTCGCAAATTAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>MtALT1-InFusion-Rev</em></td>
<td>5'-GACGGAGCTCGAATTCTAAGAGTCGTCTCAGACCCACTC-3'</td>
</tr>
<tr>
<td><em>ZmALT3-InFusion-For</em></td>
<td>5'-AATGGGTCGCAAATTAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT3-InFusion-Rev</em></td>
<td>5'-GACGGAGCTCGAATTCTAAGAGTCGTCTCCACCAATAC-3'</td>
</tr>
<tr>
<td><em>ZmALT3-CDS-For</em></td>
<td>5'-GCCGCAAGTTTTTGAG-3'</td>
</tr>
<tr>
<td><em>ZmALT3-CDS-Rev</em></td>
<td>5'-CTAAATAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT3-T108R-For</em></td>
<td>5'-CTAGAAGCGAGCCACTTCG-3'</td>
</tr>
<tr>
<td><em>ZmALT3-T108R-Rev</em></td>
<td>5'-AATGGGTCGCAAATTAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-InFusion-For</em></td>
<td>5'-GACGGAGCTCGAATTCTAAGAGTCGTCTCCACCAATAC-3'</td>
</tr>
<tr>
<td><em>ZmALT1-InFusion-Rev</em></td>
<td>5'-GCCGCAAGTTTTTGAG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-CDS-For</em></td>
<td>5'-AATGGGTCGCAAATTAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-CDS-Rev</em></td>
<td>5'-CTAAATAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-T108R-For</em></td>
<td>5'-CTAGAAGCGAGCCACTTCG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-T108R-Rev</em></td>
<td>5'-AATGGGTCGCAAATTAGGGTGAGAGAAGG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-R108T-For</em></td>
<td>5'-ATGGAAGCCACGGCG-3'</td>
</tr>
<tr>
<td><em>ZmALT1-R108T-Rev</em></td>
<td>5'-ATGGAAGCCACGGCG-3'</td>
</tr>
</tbody>
</table>
APPENDIX 3. Insert sequences of genetic constructs for heterologous expression of ALTs in K27(DE3) *E. coli*. Regions of vector homology required for InFusion cloning are highlighted in red. Restriction enzyme recognition sites used for cloning are underlined. In chimeric thioesterase sequences, sequence fragments that originate from another ALT protein are highlighted in blue.

**ZmALT1**

```
AATGGGTCGGGGGATCCGACGCAAGTTCACGGTGAGATGAGTGGATTCCATGACGTTGAACTGAAGGTGCGCGACTATGAGTTGGATCAGTACGGTGATGGTTAACAATGCAGTTTATGCTAGTTATTGCCAGCACGGTCGTCATGAACTCTTGCAAAACATTGGTATTAA
```

**ZmALT2**

```
AATGGGTCGGGGGATCCGACGCAAGTTCACGGTGAGATGAGTGGATTCCATGACGTTGAACTGAAGGTGCGCGACTATGAGTTGGATCAGTACGGTGATGGTTAACAATGCAGTTTATGCTAGTTATTGCCAGCACGGTCGTCATGAACTCTTGCAAAACATTGGTATTAA
```

**ZmALT3**

```
AATGGGTCGGGGGATCCGACGCAAGTTCACGGTGAGATGAGTGGATTCCATGACGTTGAACTGAAGGTGCGCGACTATGAGTTGGATCAGTACGGTGATGGTTAACAATGCAGTTTATGCTAGTTATTGCCAGCACGGTCGTCATGAACTCTTGCAAAACATTGGTATTAA
```

**ZmALT4**

```
AATGGGTCGGGGGATCCGACGCAAGTTCACGGTGAGATGAGTGGATTCCATGACGTTGAACTGAAGGTGCGCGACTATGAGTTGGATCAGTACGGTGATGGTTAACAATGCAGTTTATGCTAGTTATTGCCAGCACGGTCGTCATGAACTCTTGCAAAACATTGGTATTAA
```

**GmALT1**

```
AATGGGTCGGGGGATCCGACGCAAGTTCACGGTGAGATGAGTGGATTCCATGACGTTGAACTGAAGGTGCGCGACTATGAGTTGGATCAGTACGGTGATGGTTAACAATGCAGTTTATGCTAGTTATTGCCAGCACGGTCGTCATGAACTCTTGCAAAACATTGGTATTAA
```

GmALT1
SmALT1

AATGGGTCGGATCCATGAGCGAAGCATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
CrALT1

AATGGGTCGGATCCCCGGTGCGACGTGTTTATGCTACTGCAACATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
MtALT2

AATGGGTCGGATCCATGAGCGAAGCATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
MtALT1

AATGGGTCGGATCCATGAGCGAAGCATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
GbALT1

AATGGGTCGGATCCATGAGCGAAGCATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
BdALT2

AATGGGTCGGATCCATGAGCGAAGCATGAAATTTGAGCTCAAGTGCGCGATTACGAGTTAGACAACTATGG
TCTCTGTCAGAAACACGGTGTACGCGAGCTACTGCCAAGATTGAAGGTGGAATCTCT
TCTCGTACAGAACACGGTGTACGCGAGCTACTGCGACCATGCAAGACACGAGTTATTTGAAGAGCTGGGAATCAA
TTCTGAGGCTCTTGCCGGACTGGTGCTGTGCTTGCGCTTGCGGAGCTCAACATCAAGTACTTGAGCCCTTTGAA
ATCAGGGGATGTTTTCCTTGTTACGGCTAGGATAACTGGATCGTCGG
CTGTGAGACTGTTTTTCGACCACAACAT
TTACAAGCTTCCTCAACAAAGACAGTGTGAGGTAGCTAAGGAAATATCAAGGATAA
GAATTC
GAGCTCCG
AAGTGGTGACAAATTTTGTGATGTTTGTGACGTTGATGATTGAGTTGACGTTGATGATTGAGCTGTTGCAAGCATCAAAGGTGTAAGGATGATTTTCGAGCACTACAT
TGAGAAGCTGCCTAACCATGAGTTGATGGTGGAAGCAAAAGCAACAGCTGTTTGTCTCAACAAAGACTACCGGCC
AACTCGTTTACCTCCAGAGTTCTCTGTGCTCAAAGCTACAGTTTGGTCTAAG

CsALT1
AATGGGTCGGGATCCATGAAGCAAGTTTTATTAGAATTTGAGCTTAAGTCTGTGATTATGAGCTGATCAGTATGG
TGTTGGAACAAATATGTGTTATAGCAAGTAATTTGCGTGACAGCTCTTGTTGGAAAAGTTGAGCTACAA
TGCTGATGCTGGTTGCATCTTTGAGCTACATTTACAAGCTGTACAGCTACAGCTACATTTACAGCTACTGAG
AAGTTGAGCAGGTTTGGGAAGGTAGGCTGTTGATGATTGAGCTGCTTCTGATGCCTACAGTTTTCGAGCTATGG
CTGAAACTGACTTTAACTAGAGCAGTTATGGTGAGCTGAGCTTCTCTCTTCTTCTGACATTTGCTACCTTCT

VvALT1
AATGGGTCGGGATCCATGAAGCAAGTTTTATTAGAATTTGAGCTTAAGTCTGTGATTATGAGCTGATCAGTATGG
TGTTGGAACAAATATGTGTTATAGCAAGTAATTTGCGTGACAGCTCTTGTTGGAAAAGTTGAGCTACAA
TGCTGATGCTGGTTGCATCTTTGAGCTACATTTACAAGCTGTACAGCTACAGCTACATTTACAGCTACTGAG
AAGTTGAGCAGGTTTGGGAAGGTAGGCTGTTGATGATTGAGCTGCTTCTGATGCCTACAGTTTTCGAGCTATGG
CTGAAACTGACTTTAACTAGAGCAGTTATGGTGAGCTGAGCTTCTCTCTTCTTCTGACATTTGCTACCTTCT

OsALT1
AATGGGTCGGGATCCATGAAGCAAGTTTTATTAGAATTTGAGCTTAAGTCTGTGATTATGAGCTGATCAGTATGG
TGTTGGAACAAATATGTGTTATAGCAAGTAATTTGCGTGACAGCTCTTGTTGGAAAAGTTGAGCTACAA
TGCTGATGCTGGTTGCATCTTTGAGCTACATTTACAAGCTGTACAGCTACAGCTACATTTACAGCTACTGAG
AAGTTGAGCAGGTTTGGGAAGGTAGGCTGTTGATGATTGAGCTGCTTCTGATGCCTACAGTTTTCGAGCTATGG
CTGAAACTGACTTTAACTAGAGCAGTTATGGTGAGCTGAGCTTCTCTCTTCTTCTGACATTTGCTACCTTCT

BdALT1
AATGGGTCGGGATCCATGAAGCAAGTTTTATTAGAATTTGAGCTTAAGTCTGTGATTATGAGCTGATCAGTATGG
TGTTGGAACAAATATGTGTTATAGCAAGTAATTTGCGTGACAGCTCTTGTTGGAAAAGTTGAGCTACAA
TGCTGATGCTGGTTGCATCTTTGAGCTACATTTACAAGCTGTACAGCTACAGCTACATTTACAGCTACTGAG
AAGTTGAGCAGGTTTGGGAAGGTAGGCTGTTGATGATTGAGCTGCTTCTGATGCCTACAGTTTTCGAGCTATGG
CTGAAACTGACTTTAACTAGAGCAGTTATGGTGAGCTGAGCTTCTCTCTTCTTCTGACATTTGCTACCTTCT

AtALT3
AATGGGTCGGGATCCATGAAGCAAGTTTTATTAGAATTTGAGCTTAAGTCTGTGATTATGAGCTGATCAGTATGG
TGTTGGAACAAATATGTGTTATAGCAAGTAATTTGCGTGACAGCTCTTGTTGGAAAAGTTGAGCTACAA
TGCTGATGCTGGTTGCATCTTTGAGCTACATTTACAAGCTGTACAGCTACAGCTACATTTACAGCTACTGAG
AAGTTGAGCAGGTTTGGGAAGGTAGGCTGTTGATGATTGAGCTGCTTCTGATGCCTACAGTTTTCGAGCTATGG
CTGAAACTGACTTTAACTAGAGCAGTTATGGTGAGCTGAGCTTCTCTCTTCTTCTGACATTTGCTACCTTCT

201
AtALT4

AATGGGTCGCGGATCCATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACCACAGTTTGAA
AAGCTTGC
GGCGGCAC

AtALT3-A

AATGGGTCGCGGATCACATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACCACAGTTTGAA
AAGCTTGC
GGCGGCAC

AtALT4-A

AATGGGTCGCGGATCCATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACACAGTTTGAA
AAGCTTGC
GGCGGCAC

AtALT4-B

AATGGGTCGCGGATCCATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACACAGTTTGAA
AAGCTTGC
GGCGGCAC

AtALT4-AB

AATGGGTCGCGGATCCATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACACAGTTTGAA
AAGCTTGC
GGCGGCAC

AtALT4-C

AATGGGTCGCGGATCCATGAGTGAGTTCATAGGTTGAGCTTAAGTAGTGATTAGAATTGGATCAAGTTTG
TGTTGTGAAACATCGCTTTACGCAACACTCGCCAAACAGCGCATGCAAGAATTTCTAGAGTAGTATTTGCA
CTGTGATGATGATGCTGGCTGTAAGGCTTATCGCAATATCAAGTGTGCAAACTATTGACGTTCTGAAAC
TAGCGGACACAGTTTCGTTAGGAAAGTGACATATCTTCGACATCTCCAT
CTGGAAACTTCAAATACAGAGTTTATTTTTGGGCAACACAGTTTATGTGTTGCTGACAAACAGACGCGTCC
TGTTGTGTTATTCATTCATACGCTCTAAATTTGTTCACTTCTTACGCCAAAACGACACAGTTTGAA
AAGCTTGC
GGCGGCAC

\[\text{CGCGGATCC}\]
AtALT4-BCD

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG

AtALT4-BCDE

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG

MtALT2-A

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG

MtALT1-A

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG

MtALT1-B

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG

MtALT1-AB

AATGGGTCCGGGATCCATGAGTGAGTTCATAGGTTGAGCTAATAGTCGATTATGATATGGGACATTTGGAATTCGGATCAGTTTGG
TTGTTGTAACAATCGTGTTTCGACAAAATCTGACCAAACACGTCGATGCAATATTTCTAGAGGATATGGATACAA
CTGTGATGAAGTTCGGCTGGTGAAGCCTTAGCAATATCAGAGTTGACAATG
TTTTGCTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC

MtALT1–C

AATGGGTCGCGGATCCCATGTGAGTCTTATGAGCTGGAACCTAATAAAGGTTGGAACCTATATCGGATCGTGATG
TGTTGGTCAACAATGCAATTTATGCTAAATTGACAAAATATGTGGTAGCAGATGTGGATAGG
TTTGGTGTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC

MtALT1–ABC

AATGGGTCGCGGATCCATGTGAGTCTTATGAGCTGGAACCTAATAAAGGTTGGAACCTATATCGGATCGTGATG
TGTTGGTCAACAATGCAATTTATGCTAAATTGACAAAATATGTGGTAGCAGATGTGGATAGG
TTTGGTGTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC

MtALT1–ABC2

AATGGGTCGCGGATCCATGTGAGTCTTATGAGCTGGAACCTAATAAAGGTTGGAACCTATATCGGATCGTGATG
TGTTGGTCAACAATGCAATTTATGCTAAATTGACAAAATATGTGGTAGCAGATGTGGATAGG
TTTGGTGTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC

MtALT1–ABC + aa78–82

AATGGGTCGCGGATCCATGTGAGTCTTATGAGCTGGAACCTAATAAAGGTTGGAACCTATATCGGATCGTGATG
TGTTGGTCAACAATGCAATTTATGCTAAATTGACAAAATATGTGGTAGCAGATGTGGATAGG
TTTGGTGTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC

ZmALT1–A

AATGGGTCGCGGATCCAGGAGCAAGATTTTTCAGAGATCGGAGAGATTTTGGGAGGCAGACGACCTGGAAGAGTACGG
CGTGGGTCAACAATGCAATTTATGCTAAATTGACAAAATATGTGGTAGCAGATGTGGATAGG
TTTGGTGTGATGTGATTCGGAGTGGTGATGCAATGGCACTGTCAAATTTATCCCTCAAATTCCTTGCACCATTAAG
AAGTGGAGACAAATTTGTTGTAAGGTTTAGAATTTCCGGCATATCGGCAGCTCGTTTATACCTTGATCAGTTCAT
CTATAAGCTACCAAATCACAAGCCTGTTTTGGAAGCC
AAGGCCACAGCA
GTGAGGCTTGATAAAAACTATCGTCC
TCTTCGAATTTCAGAAGATATGAAGTCTAAAATTTTTAAATGTATTGGTGGAGACGACTCTTAA
AAGCTTGGCGGC
CGCAC
ZmALT3-A

AATGGGTGTCGGATCCCGCGCAAGTTTTTCGGGAGATGACGCTCGTACCCCGACGTGACGGGTTGTG
GGTACGCTACTGTCATATGGTCATGGTACGGTGATGGTCGTCGTGAAATAGCATGGTACGG
GCATGGTTATGGGAGATGGCGTTGTCGTCGTAAGCTGGGCATCAAAGGTGAAGCCTGTCCAATACGG
CAACGAGCTGCGGTACGTGGTCACTTCTCCACCTAAACCAAGATACCGGCAACCTCGCATGTTCCCGG

ZmALT1-B

AATGGGTGTCGGATCCCGCGCAAGTTTTTCGGGAGATGACGCTCGTACCCCGACGTGACGGGTTGTG
GGTACGCTACTGTCATATGGTCATGGTACGGTGATGGTCGTCGTGAAATAGCATGGTACGG
GCATGGTTATGGGAGATGGCGTTGTCGTCGTAAGCTGGGCATCAAAGGTGAAGCCTGTCCAATACGG
CAACGAGCTGCGGTACGTGGTCACTTCTCCACCTAAACCAAGATACCGGCAACCTCGCATGTTCCCGG

ZmALT3-B

AATGGGTGTCGGATCCCGCGCAAGTTTTTCGGGAGATGACGCTCGTACCCCGACGTGACGGGTTGTG
GGTACGCTACTGTCATATGGTCATGGTACGGTGATGGTCGTCGTGAAATAGCATGGTACGG
GCATGGTTATGGGAGATGGCGTTGTCGTCGTAAGCTGGGCATCAAAGGTGAAGCCTGTCCAATACGG
CAACGAGCTGCGGTACGTGGTCACTTCTCCACCTAAACCAAGATACCGGCAACCTCGCATGTTCCCGG

ZmALT1-C

AATGGGTGTCGGATCCCGCGCAAGTTTTTCGGGAGATGACGCTCGTACCCCGACGTGACGGGTTGTG
GGTACGCTACTGTCATATGGTCATGGTACGGTGATGGTCGTCGTGAAATAGCATGGTACGG
GCATGGTTATGGGAGATGGCGTTGTCGTCGTAAGCTGGGCATCAAAGGTGAAGCCTGTCCAATACGG
CAACGAGCTGCGGTACGTGGTCACTTCTCCACCTAAACCAAGATACCGGCAACCTCGCATGTTCCCGG

ZmALT3-C

AATGGGTGTCGGATCCCGCGCAAGTTTTTCGGGAGATGACGCTCGTACCCCGACGTGACGGGTTGTG
GGTACGCTACTGTCATATGGTCATGGTACGGTGATGGTCGTCGTGAAATAGCATGGTACGG
GCATGGTTATGGGAGATGGCGTTGTCGTCGTAAGCTGGGCATCAAAGGTGAAGCCTGTCCAATACGG
CAACGAGCTGCGGTACGTGGTCACTTCTCCACCTAAACCAAGATACCGGCAACCTCGCATGTTCCCGG

206
APPENDIX 4. Electron impact (EI) mass spectra used to identify the products of heterologously ALT-type thioesterases.

S4.A – Electron impact mass spectra (70eV ionization energy) of fatty acids and methylketones extracted from spent media of K27(DE3) E. coli cultures expressing ALT-type thioesterases. 3-hydroxy fatty acids were conclusively identified based mass spectra of their trimethylsilyl derivatives in silylated lipid extracts, as described in Chapter 2, Section 2.2.5. Sample spectra (left) are compared to those of authentic standards (right). Standard spectra of monounsaturated methylketones were obtained from Pulsifer et al. (2014); all other standard spectra were obtained from the National Institute of Standards and Technology’s NIST17 library.
Sample Spectra

- 2-tridecanone
- 6(Z)-tridecan-2-one (Goh et al., 2012)
- (Z)-6-tridecan-2-one (Pulsfer et al., 2014)
- (Z)-5(Z)-dodecenoic acid

Standard Spectra

- Dodecanoic acid
Sample Spectra

2-pentadecanone

8(Z)-pentadecan-2-one (Goh et al., 2012)

(Z)-8-pentadecan-2-one (Puisifer et al., 2014)

7(Z)-tetradecenoic acid

Tetradecanoic acid
Sample Spectra

7(Z)-3-hydroxytetradec-7-enoic acid, 2TMS

Standard Spectra

3-hydroxytetradecanoic acid, 2TMS

none available
S4.B – Electron impact mass spectra (70eV ionization energy) of fatty acid methyl esters identified in *Arabidopsis thaliana* and *Camelina sativa* seed oil samples, and of medium-chain fatty acid methyl esters and methylketones in lipid extracts from *N. benthamiana* leaves overexpressing ALT-like thioesterases. All sample spectra are compared to standard spectra obtained from the National Institute of Standards and Technology NIST17 spectral library.
Sample Spectra

![Sample Spectra](image1)

Standard Spectra

![Standard Spectra](image2)

- Eicosanoic acid, methyl ester
- (Z)-11 eicosenoic acid, methyl ester
- (Z)-13 eicosenoic acid, methyl ester
- (Z,Z)-11,14 eicosadienoic acid, methyl ester
REFERENCES


