Functional analysis of ketoacyl synthase and dehydratase domains from a PUFA synthase of *Thraustochytrium* in *Escherichia coli* and *Arabidopsis thaliana*

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

in Partial Fulfilment of the Requirements

for the Degree of Doctor of Philosophy

in the Department of Food and Bioproduct Sciences

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

By

Xi Xie

2019

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head

Department of Food and Bioproduct Sciences

University of Saskatchewan

Office 3E08, Agriculture Building

51 Campus Drive

Saskatoon, Saskatchewan,

Canada

or

Dean College of Graduate and Postdoctoral Studies University of Saskatchewan 116

Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan S7N 5C9

Canada

ABSTRACT

Polyunsaturated fatty acid (PUFA) synthase in *Thrasustochytrium* comprises three subunits, each with multiple catalytic domains. Among these domains, ketoacyl synthase (KS) and dehydratase (DH) domains play critical roles in retaining and introducing double bonds during the biosynthesis of very long chain polyunsaturated fatty acids (VLCPUFAs). To functionally analyze these domains, two putative KS domains (KS-A: KS domain from subunit-A and KS-B: KS domain from subunit-B) of the PUFA synthase were dissected and expressed as standalone enzymes in *E. coli*. The results showed that both KS-A and KS-B domains could complement defective phenotypes of *E. coli FabB* and *FabF* mutants, but the mutagenized ones where the active site cysteine residue was replaced by lysine could not. Overexpression of these domains individually in a wild type *E. coli* strain increased fatty acid production. Particularly, expression of KS-B domain produced a higher ratio of unsaturated fatty acids (UFAs) to saturated fatty acids (SFAs), while expression of KS-A improved the overall production of SFAs more effectively. These results imply that KS-A is more comparable to *E. coli* FabF, while KS-B is more similar to *E. coli* FabB in catalytic function.

To further characterize the KS domain from the subunit-B of the PUFA synthase, plastidial expression of this domain in *Arabidopsis thaliana* was conducted. The results showed that it could functionally complement the defective growth phenotypes of *At*KASI knockout mutant generated using CRISPR/Cas9. Expression of the KS domain under a seed specific promoter in wild type *Arabidopsis* could significantly enhance seed weight and seed oil, and promote seed germination and seedling growth. These results indicate that the main condensation process of fatty acid biosynthesis in plants might be a limiting step and overexpression of the KS domain from a PUFA synthase of microbial origin may offer a new strategy to increase oil production in oilseed crops.

To functionally characterize the DH domains of the PUFA synthase, one DH domain from subunit-A (DH-A) and two DH domains from subunit-C (DH1-C and DH2-C) were dissected from the synthase and expressed as standalone proteins in both mutant and wild type *E. coli* strains. Expression of DH domains in a temperature sensitive mutant (*FabA*) defective in 3-hydroxyacyl-ACP dehydratase activity showed that all these DH domains could complement the defective growth phenotype. Overexpression of DH domains drastically altered the fatty acid production and

the ratio of unsaturated to saturated fatty acids. These results indicate that the DH domain from subunit-A is comparable to DH domains of polyketide synthases, while the DH domains from subunit-C are more similar to *E. coli* FabA in catalytic function. Successful complementation and functional expression of the embedded KS and DH domains from the PUFA synthase in *E.coli* and *Arabidopsis* is an important step forwards for elucidating the molecular mechanism in the biosynthesis of VLCPUFAs catalyzed by the PUFA synthase in *Thraustochytrium* and offers strategies for metabolic engineering of the anaerobic pathway in heterologous systems to produce specialty fatty acids for nutritional and industrial uses.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

℃	Degrees celsius
ACP	Acyl carrier protein
ALA	α-linolenic acid
ARA	Arachidonic acid
AT	Acyltransferase
CE	Cholesterol ester
CFA	Cyclopropane fatty acid synthase
CL	Complementation line
CLF	Chain length factor
CoA	Coenzyme A
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein-9 nuclease
DAF	Days after flowering
DAG	Diacylglycerol
DAG	Day after growing
DGLA	Dihomo-γ-linolenic acid
DH	Dehydratase
DHA	Docosahexaenic acid
dNTP	Deoxynucleotide triphosphate
DPA	Docosapentaenoic acid
DsRED	Red fluorescent protein
E.coli	*Escherichia coli*
eGFP	Green fluorescent protein
EPA	Eicosapentaenoic acid
ER	Enoyl reductase
ETA	Eicosatetraenoic acid
FA	Fatty acid
FabA	3-hydroxyacyl-ACP dehydratase
FabB	3-ketoacyl-ACP synthase I

FabF	3-ketoacyl-ACP synthase II
FabH	3-ketoacyl-ACP synthase III
FabI	Enoyl-ACP reductase
FabZ	3-hydroxyacyl-ACP dehydratase
FAME	Fatty acid methyl ester
FAS	Fatty acids synthase
g	Gravity
GC	Gas chromatography
h	Hour
HRP	Horseradish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAS	Ketoacyl-ACP synthases
KR	Ketoacyl reductase
KS	Ketoacyl synthase
L	Liter
LA	Linoleic acid
LB broth	Luria-Bertani broth
LCFA	Long chain fatty acid
LPAAT	Lysophosphatidic acid acyltransferase
MAT	Malonyl-acetyl transferase
MS medium	Murashige and Skoog medium
NADPH	Nicotinamide adenine dinucleotide phosphate
NBRP	National BioResource Project
NCBI	National Center for Biotechnology Information
ng	Nanogram
NPD1	Neuroprotectin D1
NTG	N-methyl-N′-nitro-N-nitrosoguanidine
OA	Oleic acid
OE	Overexpression
ORF	Open reading frame
P	P-value

PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PI	Phosphatidylinositol
PKS	Polyketide synthase
Ppant	Phosphopantetheinyl
PPTase	Phosphopantetheinyl transferase
PS	Phosphatidylserine
PUFA	Polyunsaturated fatty acid
PUFA synthase	Polyunsaturated fatty acid synthase
rpm	Revolutions per minute
RT-PCR	Reverse transcription PCR
s	Second
SDA	Stearidonic acid
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SFA	Saturated fatty acid
TAG	Triacylglycerol
THA	Tetracosahexaenoic acid
TPA	Tetracosapentaenoic acid
Tris-HCl	Tris(hydroxymethyl)aminomethane
Ts	Temperature sensitive
UFA	Unsaturated fatty acid
μg	Microgram(s)
μL	Microliter(s)
μM	Micromolar
UV	Ultraviolet
VLCPUFA	Very long chain polyunsaturated fatty acid
WRI1	WRINKLED1

CHAPTER 1. INTRODUCTION

1.1 Overview

Very long chain polyunsaturated fatty acids (VLCPUFAs) such as docosahexaenoic acid (DHA, 22:6n-3), arachidonic acid (ARA, 20:4n-6) and eicosapentaenoic acid (EPA, 20:5n-3) are well established as nutraceutical compounds for human health and nutrition. Clinical studies have shown that these fatty acids play important roles in retinal and neuronic development, chronic disease prevention, and cell homeostasis maintenance. For example, DHA, one of the most abundant fatty acids in the rod outer segment of retina, is involved in rhodopsin activation and rod and cone development (Giusto *et al*., 2000, German *et al.*, 2015). DHA and ARA, the components of cell membrane of nervous and immune systems, have important roles in maintaining normal neurological activity and immune function (Crawford *et al*., 1997; Sprecher, 2002). In addition, DHA and ARA, precursors for bioactive compounds such as docosanoids and eicosanoids, are involved in mediating various physiological activities (Serhan, 2005; Ridi *et al*., 2010). Moreover, DHA, a ligand for transcriptional factors, can regulate the expression of numerous genes in human cells (Diep *et al.*, 2000).

In nature, VLCPUFAs are synthesized through two distinct pathways. In the aerobic pathway that mainly occurs in eukaryotes, double bonds of VLCPUFAs are introduced by oxygenic desaturations on existing long chain fatty acids (LCFAs) catalyzed by front-end desaturases such as Δre synthesized through two distinct in animals and microorganisms, and methyl-end desaturases such as ω3-, Δ 12-, and Δ 15-desaturases in plants and microorganisms. In the anaerobic pathway that mainly occurs in oceanic microbes, VLCPUFAs are *de novo* synthesized by a polyunsaturated fatty acid (PUFA) synthase and double bonds are introduced during the acyl chain-extending process without the involvement of any aerobic desaturations. PUFA synthase with multiple domains such as ketoacyl synthase (KS), ketoacyl reductase (KR), dehydratase (DH) and enoyl reductase (ER) as well as malonyl-acetyl transferase (MAT) and acyl carrier protein (ACP) catalyzes a series of chemical reactions such as condensation (by KS domain), ketoacyl reduction (by KR domain), dehydration (by DH domain) and enoyl reduction (by ER domain) of an acyl chain repeatedly during synthesis. Among these catalytic domains, KS and DH domains play key roles for positioning double bonds in VLCPUFAs.

Thraustochytrium sp. 26185, a marine unicellular protist, can produce a high level of nutritionally important VLCPUFAs such as docosapentaenoic acid (DPA, 22:5n-6) and DHA. Our earlier research identified a few front-end desaturases such as Δ4- and Δ5-desaturases and a number of ELO type elongases in the aerobic pathway in this protist (Qiu *et al*., 2001; Wu *et al*., 2005), Our recent genomic survey, heterologous expression and *in vivo* feeding experiments showed that two critical desaturation steps (Δ9 and Δ12 desaturations) in the pathway were absent and the anaerobic pathway catalyzed by a PUFA synthase was solely responsible for DHA biosynthesis in *Thraustochytrium* (Meesapyodsuk & Qiu, 2016). Similar to those from other marine microbes (Gong *et al*., 2014; Okuyama *et al*., 2007; Ruiz-López *et al*., 2012), the PUFA synthase from *Thraustochytrium* is a mega-enzyme which comprises multiple catalytic domains. However, the exact functions of these domains in the biosynthesis of VLCPUFAs are unknown. This thesis research aimed to functionally analyze two types of domains (KS and DH) of the PUFA synthase by heterologous expressions in *Escherichia coli* and *Arabidopsis thaliana*.

1.2 Hypothesis

The DHA biosynthesis by a PUFA synthase in *Thraustochytrium* involves the coordinated activity of multiple catalytic domains. Among these domains, KS and DH domains are important for positioning double bonds during the biosynthetic process. Particularly, the KS domains are able to catalyze the condensation of unsaturated acyl-ACP with malonyl-ACP, resulting in retaining a double bond in the acyl chain. On the other hand, the DH domains can catalyze the specific dehydration of hydroxyacyl-ACP to introduce a *cis* double bond in the acyl chain. Coordination of these two types of domains can position multiple double bonds in VLCPUFAs intricately. Moreover, KS and DH domains of the PUFA synthase can cooperate with other components of type II fatty acid synthase for fatty acid synthesis in bacteria and plants.

1.3 Objective

The overall objective of this research was to elucidate the catalytic mechanism of a PUFA synthase for the DHA biosynthesis in *Thraustochytrium*. The specific objectives were to functionally characterize KS domains and DH domains of the PUFA synthase from *Thraustochytrium* in *E. coli* and *Arabidopsis* by complementation test and overexpression.

2

CHAPTER 2. LITERATURE SURVEY

2.1 Introduction of fatty acids

2.1.1 Chemical structure

Fatty acid consists of a hydrocarbon chain with a carboxyl group at one end and a methyl group at the other. The acidic carboxyl group reacts readily with a hydroxyl group to form an ester linkage (Arild & Drevon, 2006). Fatty acids can be esterified to glycerol backbone, becoming the important part of glycerolipids, such as triacylglycerols (TAGs) and phospholipids. Phospholipids are the main constituents of biological membranes. Common phospholipids found in membranes include phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), diphosphatidylglycerol and phosphatidylinositol (PI). Sphingolipids are another class of lipids derived from aliphatic amino alcohol sphingosine, a product from the condensation of serine and palmitic acid (16:0). Sphingosine can be bonded to a fatty acid (FA) through an amide-linkage and to a head group such as phosphocholine through ester-linkage, giving a sphingomyelin (Spener *et al*., 2013).

2.1.2 Nomenclature

Fatty acid can have a trivial name, a systematic name, or be named according to 'Δ' and 'ω' notation (Figure 2.1). In the trivial nomenclature, each fatty acid has a name without detailed structural information, e.g. arachidonic acid (20:4n-6). However, trivial names may cause confusion between two fatty acids sharing the same or similar name with different structures. For example, arachidonic acid and arachidic acid (20:0) have similar names, but their structures are very different. Different from the trivial nomenclature, the systematic nomenclature names fatty acids according to the number of carbon atoms, and the number of double bonds and their positions relative to the carboxyl group of an acyl chain. For instance, a 16-carbon saturated fatty acid palmitic acid (trivial name) is hexadecanoic acid (systematic name), and a 22-carbon VLCPUFA cervonic acid (trivial name) is written as *cis*-4,7,10,13,16,19-docosahexaenoic acid (systematic name). The 'ω' or 'Δ' notation is regularly used for naming unsaturated fatty acids based on the location of double bonds. The 'ω' naming system marks the first double bond location counting from the methyl end, while the 'Δ' naming system marks the double bond locations from the

carboxyl end. For example, arachidonic acid is written as 20:4n-6. This means it has the chain length of 20 carbons, and four double bonds with the first double bond at the sixth carbon counting from the methyl end (n-6). The 'n' or 'ω' notation for naming fatty acids is useful when considering fatty acid biosynthesis, as fatty acids are extended at the carboxyl end and the location of the last double bond from the methyl end remains unchanged during the synthetic process.

	Trivial Name	Systematic Name	'n' or 'ω' notation	Δ notation
20:3n-6 (DGLA)	Dihomo-γ-linolenic acid	all cis-8,11,14-eicosatrienoic acid	n-6/ω-6	Δ8,11,14
20:4n-6 (ARA)	Arachidonic acid	all cis-5,8,11,14-eicosatetraenoic acid	n-6/ω-6	Δ5,8,11,14
20:5n-3 (EPA)	Timnodonic acid	all-cis-5,8,11,14,17-eicosapentaenoic acid	n-3/ω-3	Δ5,8,11,14,17
22:6n-3 (DHA)	Cervonic acid	all-cis-4,7,10,13,16,19-docosahexaenoic acid	n-3/ω-3	Δ4,7,10,13,16,19

Figure 2.1 Nomenclature of VLCPUFAs. Δ notation, carbon numbering starts from carboxyl end; 'n' or 'ω' notation, carbon numbering starts from methyl end.

2.2 Very long chain polyunsaturated fatty acids

Very long chain polyunsaturated fatty acids (VLCPUFAs) are the fatty acids with the chain length longer than 18 carbons and two or more double bonds. According to the last double bond location, VLCPUFAs can be classified into two groups. The first group is named ω-3 (or alternatively n-3) fatty acids that have the last double bond at the third carbon away from the

methyl (CH₃) end, while the second group is named ω-6 (or alternatively n-6) fatty acids that have the last double bond at the sixth carbon away from the methyl end.

2.2.1 Omega-6 VLCPUFAs

Omega-6 VLCPUFAs are a group of polyunsaturated fatty acids (PUFAs) longer than 18C with the last double bond at the sixth carbon counting from the methyl end. As compared to ω-3 fatty acids, ω-6 fatty acids are more widely found in our diet with refined vegetable oils and foods cooked in vegetable oils. The most abundant ω-6 fatty acids from food are linoleic acid (LA) and gamma-linolenic acid (GLA). After consumption, these PUFAs can be easily metabolized to ω-6 VLCPUFAs by elongases and desaturases in humans following Δ6 desaturation → Δ 6 elongation → Δ5 desaturation pathways, yielding dihomo-gamma-linolenic acid (DGLA) and ARA. DGLA and ARA are also found in a number of lower living species including bacteria, fungi, microalgae, macroalgae, lichens and some lower plants (Shanab *et al.*, 2018).

2.2.2 Omega-6 VLCPUFAs and human health

In recent years, ω-6 VLCPUFAs, especially ARA, have drawn public attention, as some clinical results indicate too much of these fatty acids might have undesirable effects on human health. ARA is the most abundant ω-6 VLCPUFA in humans with multiple physiological activities (Tallima & El, 2018). Firstly, ARA affects membrane fluidity which in turn influences the functions of membrane proteins, and lots of those are involved in cellular signaling. In addition, ARA can be metabolized to eicosanoids (prostanoids and leukotrienes) with pro-inflammatory activity through oxygenated processes that can regulate many physiologic processes (Sprecher, 2002). Several reports indicate that ω-6 VLCPUFAs, especially free ARA, may possess tumoricidal activity, and promotes the production of free radicals through a lipid peroxidation process (Ridi *et al.*, 2010). Prostaglandins derived from ARA contribute to neuroinflammation, and facilitate the development of *Alzheimer's* disease (Kulmacz, 1998; Rink & Khanna, 2010).

2.2.3 Omega-3 VLCPUFAs

Omega-3 VLCPUFAs are another group of VLCPUFAs with the last double bond at the third carbon counting from the methyl end. Some of ω-3 VLCPUFAs, such as EPA and DHA, have received growing attention due to their health benefits for humans and animals. Two sources of EPA and DHA exist in nature: one is oceanic fish and the other is marine microorganisms including bacteria, microalga and fungi. Currently, fish population in oceans is declining due to

the uneven frequency of *El Nino*, environmental contamination and overfishing. Therefore, marine microorganisms are an increasingly important source for ω-3 VLCPUFAs. Microbial single-cell oils including oil from protist Thraustochytrid have become one of the significant segments within ω-3 ingredient market in recent years. Compared to the fish source of ω-3 VLCPUFAs, the microbial source of ω-3 VLCPUFAs has several advantages. The fish oil possesses poor oxidative stability, while the oil from microorganisms is more stable. The fish oil is usually accompanied with unpleasant odor and taste, while the oil from microorganisms does not have the problem. Additionally, ω-3 VLCPUFAs-producing microorganisms are more readily grown than fish under a controlled environment. More importantly, unlike microorganisms, fish generally do not *de novo* synthesize DHA and rely on DHA-containing feeds for growth and development. The microbial source of ω-3 VLCPUFAs is more sustainable (Sprague *et al.*, 2017).

2.2.4 Omega-3 VLCPUFAs and human health

2.2.4.1 Docosahexaenoic Acid

Among ω-VLCPUFAs, DHA may be the most important ω-3 fatty acid for human health. DHA is a VLCPUFA with 22-carbon atoms and six cis-double bonds positioning at 4, 7, 10, 13, 16, 19 carbons counting from the carboxyl end. It belongs to ω-3 PUFAs as it has the last double bond located at the third carbon from the methyl end (Figure 2.2).
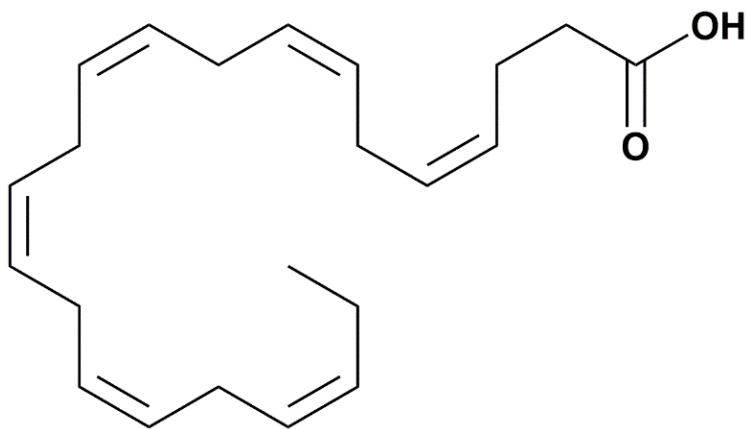
Figure 2.2 Structure of DHA.

DHA is an essential fatty acid for human health. Firstly, DHA is an imperative component of neuron, retina and immune cells. In retina, DHA is the most abundant fatty acid in rod outer segment, the photosensitive part of photoreceptor cells. It is involved in signal transduction, rhodopsin activation, and rod and cone development. In nervous system, DHA is the structural component of synaptic membranes, playing an important role in the maintenance of normal neural function (Crawford *et al*., 1997; Sprecher, 2002). In brains, DHA is the major fatty acid accumulated in cortical gray matter, the functional part for executing physical and physiological activities. The growth of cortical gray matter has positive effects on sensorimotor integration, attention-executive function and memory (McNamara, 2014). In addition, DHA is the precursor for bioactive compounds docosanoid neuroprotectin D1 (NPD1) that is involved in mediating various physiological activities. Synthesis of NPD1 occurs mainly in response to oxidative stress and/or neurotrophin activation for the homeostatic maintenance of cellular integrity. NPD1 promotes neuronal and/or photoreceptor cell survival, modulates inflammatory signaling cascades and prevents brain damage (Niemoller & Bazan, 2010). Thirdly, DHA is a key biological modulator for gene expression. The expression regulation by DHA occurs mainly at the transcriptional level by activating a few nuclear transcriptional factors. Therefore, maintaining an optimal level of DHA in human body is critical and the deficiency would have serious consequences on human health such as hypertension, arthritis, arteriosclerosis, thrombosis, *Alzheimer's* disease, type 2 diabetes, dementia and attention deficit-hyperactivity disorder (Ajith & Jayakumar, 2019; Daiello *et al*., 2015; Newell *et al*., 2017; Vandongen *et al*., 1993; Wu *et al*., 2012).

2.2.4.2 Other ω-3 VLCPUFAs

Other ω-3 VLCPUFAs such as EPA and DPA (22:5n-3) are considered to be important for health and wellbeing of humans and animals as well. Some clinical and epidemiological results indicate these ω3-VLCPUFAs have potential in preventing or treating diseases such as myocardial infarction, bronchial asthma, bowel inflammation, depression and certain types of cancers (Niemoller & Bazan, 2010; Freedman *et al.*, 2004; Vandongen *et al*., 1993; Kobayashi *et al*., 2004). For example, EPA can be converted to beneficial bioactive eicosanoids for modulating blood pressure and blood coagulation and inflammatory and immunological reactions. Some ω3-VLCPUFAs have also been shown to reduce *de novo* lipogenesis in rat livers, and decrease oxidative stress by lowering the activity of superoxide dismutase and glutathione peroxidase

(Neschen *et al*., 2002; Ruiz-Gutierrez *et al*., 1999; Rustan *et al*., 1988). It is generally believed that the ω3-VLCPUFA concentration in brains declines along with aging, which is accompanied by a decrease in neuronal survival and synaptic density, and loss of both gray and white matter volume as well as reduction of cognition (Masliah *et al*., 1993; Moretti *et al*., 2012).

2.3 VLCPUFA biosynthesis

In nature, the biosynthesis of VLCPUFAs, including DHA, goes through two distinct pathways, aerobic pathway and anaerobic pathway according to whether it requires molecular oxygen during the biosynthetic process (Qiu, 2003).

2.3.1 Aerobic pathway

DHA biosynthesis in the aerobic pathway usually occurs in cytosol and endoplasmic reticulum on long chain fatty acid precursors through alternating desaturation and elongation. Double bonds are introduced by an oxygenic desaturation process (Figure 2.3). The biosynthetic process involves sequential additions of double bonds to a saturated fatty acid, mainly 18:0, by $\Delta 9$ and $\Delta 12$ desaturases to produce LA (18:2n-6), which can then be further desaturated by $\Delta 15$ desaturase to give an ALA (18:3n-3). ALA is then metabolized to DHA following $\Delta 6$ desaturation → $\Delta 6$ elongation → $\Delta 5$ desaturation → $\Delta 5$ elongation → $\Delta 4$ desaturation in some marine microorganisms (Abe *et al.* 2006; Kajikawa *et al*. 2004; Meesapyodsuk & Qiu 2014; Qiu *et al*. 2001; Shanab *et al*. 2018; Tavares *et al*. 2011; Ye *et al*. 2015). However, in mammals, DHA cannot be synthesized from saturated and monounsaturated long chain fatty acids due to the absence of $\Delta 12$ and $\Delta 15$ desaturases. Therefore, the precursor ALA must be obtained from diets for DHA synthesis. Biosynthesis of EPA, a precursor for DHA in mammals, can be achieved by $\Delta 6$ desaturation of ALA to produce stearidonic acid (SDA, 18:4n-3) that is then elongated to eicosatetraenoic acid (ETA, 20:4n-3), followed by $\Delta 5$ desaturation. After that, EPA is elongated twice to tetracosapentaenoic acid (TPA, 24:5n-3). This fatty acid is then desaturated by $\Delta 6$ desaturase to tetracosahexaenoic acid (THA, 24:6n-3) which undergoes one circle of β-oxidation, giving rise to DHA (Bazinet & Layé, 2014; Sprecher, 2002; Uauy *et al*. 2000).

FAS
18:0 (SA)

Δ9 Desaturase

18:1-9 (OA)

Δ12 Desaturase

18:2-9, 12 (LA)

Δ15 Desaturase

18:3-9,12,15 (ALA)

Δ6 Desaturase

18:4-6,9,12,15 (SDA)

Fatty acid elongase

20:4-8,11,14,17 (ARA)

Δ5 Desaturase

20:5-5,8,11,14,17 (EPA)

+2C   Fatty acid elongase   +2C
+2C

22:5-7,10,13,16,19 (DPA)   24:5-9,12,15,18,21 (TPA)

Δ4 Desaturase   Δ6 Desaturase

22:6-4,7,10,13,16,19 (DHA)   24:6 -6,9,12,15,18,21 (THA)

B-oxidation 2C shortening

22:6-4,7,10,13,16,19 (DHA)

Thraustochytrium   **Mammals**

Figure 2.3 The aerobic (desaturation and elongation) pathway for DHA biosynthesis in *Thraustochytrium* and mammals (Adapted from (Qiu, 2003)). SA, Stearic acid; OA, Oleic acid; LA, Linoleic acid; ALA, α-Linolenic acid; SDA, Stearidonic acid; ARA, Arachidonic acid; EPA, Eicosapentaenoic acid; DPA, Docosapentaenoic acid; DHA, Docosahexaenoic acid; TPA, Tetracosapentaenoic acid; THA, Tetracosahexaenoic acid.

2.3.1.1 Key enzymes in the aerobic pathway

In the aerobic pathway, desaturases and elongases are two key types of enzymes in producing VLCPUFAs.

Fatty acid desaturase

Fatty acid desaturase is the enzyme that can remove two hydrogen atoms from a fatty acid chain by using molecular oxygen as a cofactor, introducing a carbon-carbon double bond (Shanklin

& Cahoon, 1998). Desaturases are categorized as Δ-x desaturases that introduce a double bond at carbon position x counting from the carboxyl end of a fatty acid; or ω-y desaturases that introduce a double bond at carbon position y counting from the methyl end (Hitz *et al*., 1994; Yadav *et al*., 1993). Thus, each desaturase introduces a double bond at specific positions in a fatty acid and named accordingly. For example, Δ9 desaturase can introduce a double bond at the 9th position from the carboxyl end. In mammal, there are two Δ-x desaturases, Δ5 and Δ6 desaturases, in the biosynthesis of DHA, while in microorganisms, both Δ-x and ω-y types of desaturases including Δ4, Δ5, Δ6, Δ9, Δ12 and Δ15 desaturases participate in the biosynthesis of VLCPUFAs in the aerobic pathway.

Fatty acid elongases

Fatty acid elongases are a kind of enzymes that produce very long chain fatty acids by elongating fatty acid precursors. Elongases cooperate with fatty acid desaturases to generate very long chain polyunsaturated fatty acids (Jump, 2009). Catalytic functions of elongases and fatty acids synthases (FAS) are similar in extending the fatty acid chain by two carbons each cycle. However, two differences exist between them: (1) the intermediates of elongases are bound to coenzyme A (CoA), while FAS substrates are acyl-ACP. (2) FAS synthesizes long chain fatty acids from acetate, while elongases extend long chain fatty acids that are synthesized by FAS (Campbell & Nikolau, 2011).

2.3.2 Anaerobic pathway

Anaerobic pathway is an alternative pathway for VLCPUFA synthesis, which only occurs in microorganisms. In this pathway, double bonds are introduced during an acyl chain-extending process, rather than by aerobic desaturation. DHA biosynthesis in the anaerobic pathway is catalyzed by PUFA synthase, a polyketide synthase (PKS)-like mega-enzyme. PUFA synthase in microorganisms comprises one or more units and each unit may contain multiple domains such as KS, KR, DH, ER, AT, MAT and ACP. These domains work coordinately for catalyzing VLCPUFA synthesis through a repetitive cycle of reactions including: condensation, ketoreduction, dehydration and enoyl-reduction (Figure 2.4). PUFA synthases have been found in marine bacteria and unicellular eukaryotes such as *Schizochytrium* and *Thraustochytrium.*
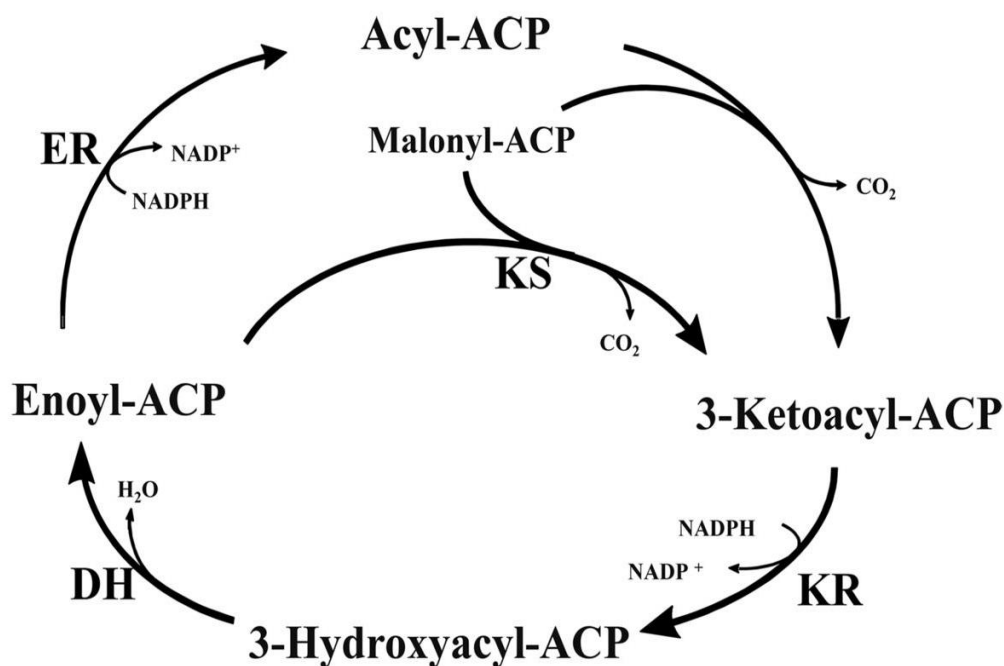
Figure 2.4 The anaerobic pathway for DHA biosynthesis (Meesapyodsuk & Qiu, 2016). ACP, acyl carrier protein; KS, 3-ketoacyl-ACP synthase; KR, 3-ketoacyl-ACP reductase; DH, 3-hydroxyacyl-ACP dehydratase; ER, enoyl-ACP reductase.

2.3.2.1 Principal domains in PUFA synthase

PUFA synthase is a mega-enzyme comprising multiple catalytic domains with specific functions. Understanding the role of each domain is important for elucidating the mechanism for the biosynthesis of VLCPUFAs by the synthase.

ACP domain

ACP domains function in shuttling substrates among the active sites of the synthase domains. ACP carries an acyl chain through a thioester linkage to a phosphopantetheinyl group (Ppant). The Ppant group functions as an arm to deliver an acyl-chain to the active sites of individual catalytic domains such as KS domain for elongation, KR domain for ketoreduction, DH domain dehydration, and ER domain for enoyl reduction. Therefore, ACP domain participates in all the catalytic reactions (Quadri *et al.*, 1998).

KS domain

KS domains are the key domains of a PUFA synthase catalyzing the condensation between a growing acyl chain in the form of an acyl-S-KS complex and an extender unit malonyl-S-ACP. Some of the KS domains can form a heterodimer (KS/CLF) with chain length factor (CLF) that is a catalytically inactive version of a KS domain. The interface of the KS-CLF dimer may define the size of a substrate-binding channel (Witkowski *et al.*, 2002; Xu *et al.*, 2013).

KR domain

A KR domain in a PUFA synthase is able to stereospecifically reduce the keto group of β-ketoacyl-ACP intermediate. This reaction is catalyzed by the activity of $NAD(P)^+/NAD(P)H$-dependent oxidoreduction. Interestingly, the activity displays stereospecificity towards substrates. The active site fixes the β-carbon relative to NADPH geometrically and facilitates a stereospecific attack by the hydride. Some X-ray crystal structure analysis reveals that specific residues of KR domain may control stereospecificity by guiding the entry of substrates into the active site groove from different sides (Kavanagh *et al*., 2008; Keatinge-Clay, 2007).

DH domain

DH domains are categorized into two types, one is PKS-like DH, and the other is FabA-like DH according to sequence and function (Leesong *et al.*,1996; Li *et al.*, 2015.). PKS-like DHs catalyze the dehydration of a hydroxyacyl intermediate, producing a trans double bond between α and β-carbon, which will be reduced by enoyl-reductase activity. However, FabA-like DH functions as a bi-functional enzyme. It, likes PKS-like DH, can catalyze the dehydration of a hydroxyacyl intermediate introducing a double bond between α and β-carbon, meanwhile, it can also switch configuration from trans to cis configuration and change position of the double bond. A DH domain possesses a double hotdog fold, mimicking a homodimer structure formed by FabA or FabZ from *E. coli*.

ER Domain

ER domain in PUFA synthase is able to catalyze the reduction of enoyl-ACP to saturated acyl-ACP. It employs NADPH to stereoselectively reduce trans-α, β double bond generated by DH. Compared to the other domains, ER has the least structural and functional information (Yin *et al.*, 2001; Ames *et al.*, 2012).

Other domains

Except for these key domains, there are some other domains in PUFA synthase including MAT domain and AT domain. The MAT domain catalyzes the transfer of malonyl-CoA to malonyl-ACP as an extender unit for condensation reactions (Maier *et al.*, 2006). The AT domain may participate in the last step of fatty acid production by cleaving the thioester bond of final acyl-ACP to release the free fatty acid (Lomakin *et al.*, 2007).

2.4 VLCPUFA biosynthesis in *Thraustochytrium*

2.4.1 *Thraustochytrium*

Thraustochytrium is a unicellular protist producing a high level of DHA in storage lipids. It is a good model for us to learn more about DHA biosynthesis for several reasons: (1) there are a large number of *Thraustochytrium* species in nature; (2) many species of *Thraustochytrium* produce a high level of lipids as part of their biomass; (3) some species of *Thraustochytrium* produce a large amount of VLCPUFAs, particularly DHA.

2.4.2 Aerobic pathway for DHA biosynthesis in *Thraustochytrium*

Our earlier research indicates that an aerobic pathway exists in this species for the biosynthesis of VLCPUFAs, and DHA can be synthesized through a final desaturation by Δ4 desaturase (Qiu *et al*., 2001). Our recent research indicates although Δ4, Δ5, Δ6 and ω3-desaturases as well as a number of ELO type elongases are found in *Thraustochytrium*, two critical desaturation steps, Δ9 and Δ12 desaturations are absent in the protist (Zhao, *et al.*, 2016). Therefore, the aerobic biosynthetic pathway of VLCPUFAs is believed to be incomplete and inefficient in *Thraustochytrium* (Meesapyodsuk & Qiu, 2016).

2.4.3 Anaerobic pathway for DHA biosynthesis in *Thraustochytrium*

The anaerobic pathway for the biosynthesis of VLCPUFAs in the protist was identified through genome sequencing, *in vitro* and *in vivo* assays. A series of evidence shows that it is the dominant and complete pathway for the biosynthesis of VLCPUFAs catalyzed by a PUFA synthase in *Thraustochytrium* (Meesapyodsuk & Qiu, 2016). Similar to those from other marine microbes (Gong *et al*., 2014; Okuyama *et al*., 2007; Ruiz-López, *et al*., 2012), the PUFA synthase from *Thraustochytrium sp.* 26185 functions as a mega-enzyme which comprises three subunits each with multiple catalytic domains and the biosynthesis of VLCPUFAs takes place through the

13

cooperation of these catalytic domains. However, the detailed mechanism for VLCPUFA biosynthesis by the PUFA synthase and intricate positioning of multiple double bonds in VLCPUFAs is unknown.

2.5 Ketoacyl-ACP synthase

Ketoacyl-ACP synthase is an enzyme catalyzing the Claisen condensation during the fatty acid biosynthesis (Janine *et al*., 2014). The first condensation occurs between acetyl-CoA/ACP and malonyl-ACP giving β-ketobutyryl-ACP with two carbon extension on the starting precursor. According to the existing form of these enzymes, ketoacyl-ACP synthase can be categorized into two types, one works as an individual enzyme in type II FAS complex, and the other functions as an embedded domain in mega-enzymes such as type I FAS, PKS and PUFA synthase.

2.5.1 Ketoacyl-ACP synthase in type II FAS

Ketoacyl-ACP synthase is a soluble and enzymatic component in type II FAS system that mainly occurs in bacteria and plants.

2.5.1.1 Ketoacyl-ACP synthase in bacteria

In bacteria, type II FAS is located in cytoplasm and comprises a few components such as ketoacyl-ACP synthase, hydroxyacyl-ACP dehydratase, ketoacyl-ACP reductase, enoyl-ACP reductase and ACP. Ketoacyl-ACP synthase comes with three types: ketoacyl-ACP synthase III (FabH), ketoacyl-ACP synthase I (FabB) and ketoacyl-ACP synthase II (FabF). FabH catalyzes the initial condensation using an acetyl-CoA/ACP as a primer and malonyl-ACP as an extender. FabB catalyzes elongation of fatty acids with chain length from 4 to 14 carbons (White *et al*., 2005). FabF has elongation activity with C4 to C14 acyl-ACPs as well. However, it possesses the preference to C16 substrates (C16:0 or C16:1) and exhibits higher activity at lower temperature (Lee *et al*., 2013). Different from FabF, FabB plays an essential role in UFA biosynthesis in bacteria. As desaturases are absent in most of bacteria, UFAs are synthesized by type II FAS system. A double bond is introduced into the growing acyl chain with 10 carbons by FabA, forming a 3-cis-decenoyl-ACP. FabB retains the double bond by elongating cis-3-decenoyl-ACP to 3-keto, 5-cis-dodecenoyl-ACP. Thus, the condensation step catalyzed by FabB is a rate-limiting step in UFA synthesis, and determines cellular UFA content in bacteria (Lu *et al*., 2004). Deletion of *FabB* gene leads to auxotroph, as unsaturated fatty acids are essential for *E. coli* (Rosenfeld *et al*.,

1973). Overexpression of FabB alters the ratio of UFAs to SFAs, and co-expression of FabB, FabA and thioesterase significantly enhances UFA production in *E.coli* (Cao *et al*., 2010).

2.5.1.2 Ketoacyl-ACP synthase in plant

In plant, components of type II FAS locate in plastids and they share amino acids sequence similarity to those present in bacterial species. Similarly, ketoacyl-ACP synthases are categorized into three types: ketoacyl-ACP synthase I (KASI), ketoacyl-ACP synthase II (KASII) and ketoacyl-ACP synthase III (KASIII). The first condensation of acetyl-CoA with malonyl-ACP to 3-ketobutyryl-ACP is catalyzed by KASIII. Subsequent condensations from C4 to C16 are catalyzed by KASI and the final condensation from C16 to C18 catalyzed by KASII.

KASI

KASI plays a central role in fatty acid biosynthesis in plant. It catalyzes the main condensation steps from C4 to C16 and deficiency of KASI activity affects the final fatty acid production and results in a significant change in membrane lipid composition and multiple morphological defects including chlorotic and curly leaves, reduced fertility, semi-dwarfism and abnormal growth of mesophyll cells (Wu & Xue, 2010). Thus, KASI gene is a target of several transcriptional factors regulating lipogenesis in plants. For example, the positive transcription factors such as WRINKLED1 (WRI1), FUSCA3 (FUS3) and LEC1-like proteins can bind to the promoter region of *KASI* to activate the expression of KASI for increasing fatty acid and TAG biosynthesis (Manan *et al*., 2017; To *et al*., 2012). The repressor such as MYB89 can bind to the promoter regions to down-regulate *KASI* gene expression (Li *et al*., 2017).

2.5.2 KS domain

KS domains are the embedded domains in mega-enzymes such as type I FAS, PUFA synthase and PKS exhibiting keto-acyl-ACP synthase activity. KS domain normally functions as a dimer to catalyze a condensation process (Xu *et al*., 2013). They are highly conserved and share sequence similarity to those discrete ketoacyl-ACP synthases. Each KS domain possesses a thiolase fold with α/β/α/β/α architecture and the condensation activity is conferred within the dimerized KS domains. The substrate binding tunnels extend from the active site along the dimer interface. The catalytic machinery is comprised of the catalytic triad with a cysteine residue and two histidine residues (White *et al*., 2005). The reactive cysteine is positioned at a "nucleophilic elbow", such that the positive dipole of the α-helix decreases its pKa and increases its

15

nucleophilicity. Two histidine residues take part in decarboxylation reactions. One of the histidine residues participates in binding/activating the carboxylate of an extender unit, while the other histidine residue serves as a general base, activating a water molecule for attack at the activated carboxylate (Keatinge-Clay, 2012). Despite KS domains from type I FAS, PUFA synthase and PKS exhibit the similar thiolase fold and catalytic mechanism, their substrate uses vary significantly.

2.5.2.1 KS domain in type I FAS and PKS

Type I FASs are commonly found in animals, fungi and some bacteria such as Corynebacteria, Mycobacteria and Nocardia. The products produced by type I FASs are saturated long chain fatty acids with 16 or 18 carbons. Similar to that by type II FAS, fatty acid synthesis catalyzed by type I FAS goes through reiterative reactions of condensation, keto-reduction, dehydration and enoyl-reduction. Unlike type II FAS where three different types of KS enzymes catalyze condensation reactions from C2 to C16 coordinately, type I FAS contains only a single KS domain catalyzing all the condensations. In addition, unlike discrete ketoacyl-ACP synthases in type II FAS, type I FAS KS domain is highly specific for saturated acyl chains, thus final products are SFAs synthesized by type I FAS (Smith *et al*., 2003).

Type I PKS KS domains possess similar structures to human type I FAS KS domain, but with a much wider range of substrates that vary in acyl chain length and structure (Khosla *et al*., 2002). The catalytic machinery is comprised of three residues: a cysteine in TA**C**SSS motif and two histidine residues in EA**H**GTG and KSNIG**H**T motifs. In addition, PKS contains a CLF domain, a catalytically inactive version of a KS domain that can form a heterodimer KS-CLF with a KS domain. The interface of the KS-CLF dimer defines the size of the substrate-binding channel (Cox & Simpson, 2009). Mutations of four residues of the KS-CLF duplex at the bottom of the acyl pocket change the pocket size and shape, thereby altering substrates and products (Dutta *et al*., 2014).

2.5.2.2 KS domain in PUFA synthase

PUFA synthase is a mega-enzyme, and comprises multiple subunits each with multiple catalytic domains. The number, structure, and organization of domains in a PUFA synthase are very different from a type I FAS (Figure 2.5). For instance, PUFA synthase in *Thraustochytrium* is comprised of three subunits with two KS domains, one located in subunit-A, and the other in

subunit-B. In addition, it also contains a pseudo-KS domain called CLF adjacent to a KS domain in subunit-B that shares similar structure to KS domains but without active site residues (Tang *et al.*, 2003). Multiple KS domains and KS-CLF organization in a PUFA synthase might be required for double bond positioning during VLCPUFA synthesis.

| *Thraustochytrium* (DHA) | KS | MAT | 8× ACP | KR | DH | | KS | CLF | AT | ER | | | DH | DH | ER |
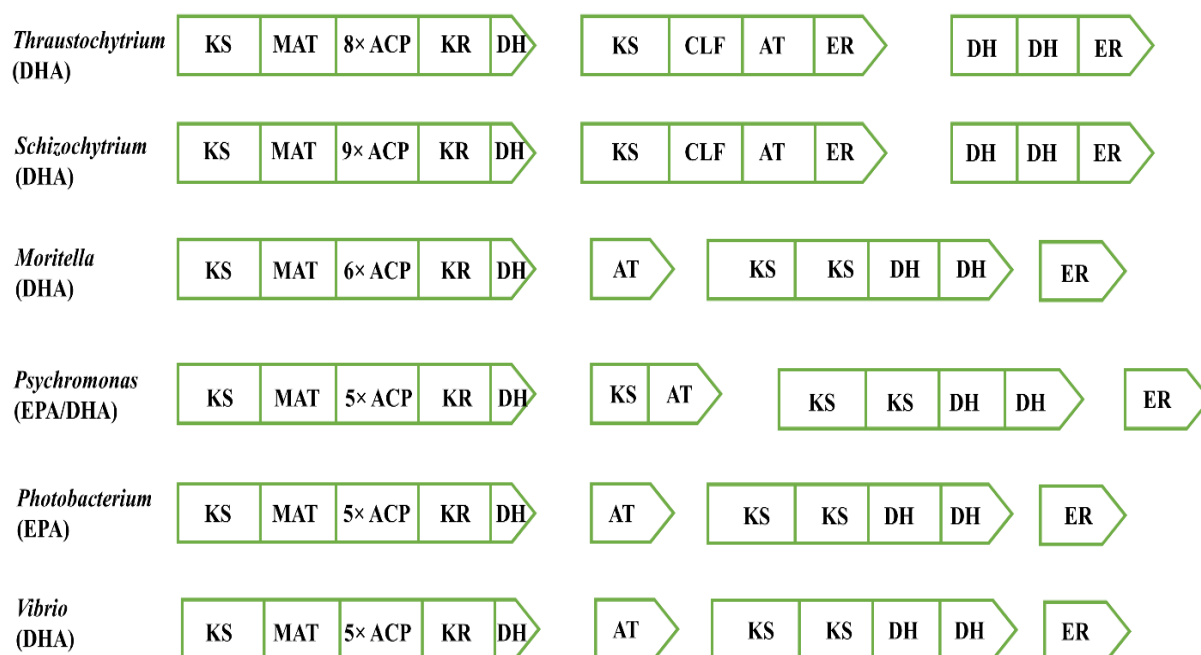
Figure 2.5 Structure of PUFA synthases from *Thraustochytrium* and other microbial species (Adapted from (Zhao *et al.*, 2016; Ji *et al.*, 2015; Kautharapu & Jarboe, 2012; Lauro *et al.*, 2013; Okuyama *et al.*, 2007; Soh *et al.*, 2018).

2.6 Hydroxyacyl-ACP dehydratase

Hydroxyacyl-ACP dehydratase (DH) catalyzes the dehydration of β-hydroxyacyl-ACP to enoyl-ACP during fatty acid biosynthesis. Removing a water molecule from β-hydroxyacyl-ACP generated by ketoacyl-ACP reductase results in introducing a double bond in the acyl chain. Hydroxyacyl-ACP dehydratase activity can be conferred by two forms of proteins, one is a discrete

enzyme in type II FAS and the other is an embedded domain in mega-enzymes such as type I FAS, PKS and PUFA synthase.

2.6.1 Hydroxyacyl-ACP dehydratase in type II FAS

Discrete hydroxyacyl-ACP dehydratase is a soluble enzyme, a component of type II FAS mainly occurring in bacteria and plants (Brown *et al*., 2009; Yoshida *et al*., 2016). In bacteria, hydroxyacyl-ACP dehydratase, also called FabA/Z in *E.coli*, is able to introduce a double bond by removing a water molecule from β-hydroxyacyl-ACP. Although FabA and FabZ share some similarity in their primary sequences, but their functions are very different. In *E.coli*, FabZ catalyzes the dehydration of a wide range of saturated and monounsaturated β-hydroxyacyl-ACPs from short chains to long chains, giving 2-trans-enoyl-ACPs that are usually reduced by enoyl-CP reductase (FabI) to saturated acyl-ACPs (Lu *et al*., 2004). However, FabA is a bifunctional enzyme catalyzing both the dehydration of β-hydroxydecanoyl-ACP and the isomerization of 2-trans-decenoyl-ACP to 3-cis-decenoyl-ACP, an essential step in UFA biosynthesis (Chirala & Wakil, 2004). Structurally, FabA and FabZ are highly similar, and both form the "hot-dog" fold with a six-stranded antiparallel β-sheet with topology 1/2/4/5/6/3 wrapping around a long central α-helix. During the catalytic process, two monomers bind with each other to form a double "hot-dog" structure. The active site is formed along the dimer interface with two critical histidine residues (His) and one aspartic acid (Asp, in FabA)/glutamic acid (Glu, in FabZ) residue contributed by individual monomers. Unlike FabZ, FabA possesses a few unique residues that play a role in widening the catalytic tunnel, thereby supporting the isomerization reaction (Leesong *et al*., 1996). Therefore, FabZ, without these residues, is incapable of accommodating 3-cis-decenoyl-ACP due to the spatial restriction in the substrate channel, thus preventing it from the isomerization of 2-trans-decenoyl-ACP (Kimber *et al*., 2004).

2.6.2 DH domain in type I FAS and PKS

In type I FAS and PKS systems, DH domains are fused with other catalytic domains in a polypeptide. Although the sequence similarity of DH domains from type I FAS and PKS are very low, the structures of these domains are similar, and all can form a double "hot-dog" structure.

Type I FAS DH

Type I FAS contains a DH domain with two non-identical pseudo sub-DH domains arranged side by side. It is likely that the two "pseudo sub-DH" domains form a dimer, a functional DH for catalysis (John *et al*., 2016; Lomakin *et al*., 2007). Therefore, structurally Type I FAS DH domains across species is highly conserved and can form a double "hot-dog" fold structure, each "hot-dog" fold contains a five-turn alpha helix acting as a sausage covered by seven-stranded antiparallel beta-sheets. The catalytic active site residues His and Asp are also highly conserved in the DH domains.

PKS DH

Dehydration catalyzed by DH domains of PKS is hypothesized to be similar to those by type II FAS dehydratase FabA and FabZ, each can dimerize to resemble a double "hot-dog" structure. However, the reactions performed by PKS DHs are more complicated in substrate selectivity. PKS DHs often operate on acyl chains with methyl, ethyl, or methoxy α-substituents, although how these DHs accommodate these substrates is unclear. In some cases, PKS DHs can catalyze the formation of cis double bonds in acyl products, indicating that these DHs can accommodate products with different geometry (Keatinge-Clay, 2012). For example, there are more than four DH domains in erythromycin PKS, but the products produced by these DH domains are diverse, partially due to the upstream products before the immediate dehydration process catalyzed by each DH domain (Keatinge-clay, 2008). The active site motifs of PKS DHs are less conserved. For instance, two DH domains of *Microcystins* PKS contain an active site aspartic acid residue in $H(X)_3D(X)_4P$ motif, while the same catalytic residue is located in HPALL**D** motif of PKS DH from *Saccharopolyspora spinose* (Keatinge-Clay, 2012).

2.6.3 DH domain in PUFA synthase

On the basis of the sequence survey, a PUFA synthase from microorganisms usually comprises three DH domains (Figure 2.5). One DH domain locates in the C-terminal region of a subunit consisting of KS-MAT-ACP-KR-DH, and the other two are found in a subunit with an arrangement of DH-DH-ER in eukaryotic PUFA synthases or of KS-KS-DH-DH in bacterial PUFA synthases. The DH domain in KS-MAT-ACP-KR-DH subunit is believed to function as a PKS-like dehydratase, while the DH domains in the tandem arrangement are more similar to FabA

dehydratase (Xie *et al*., 2017). PKS-like DH domains contains catalytic residues Asp in H(X)$_3$**D**(X)$_4$P motif and His in **H**(X)$_3$**G**(X)$_4$P motif (Rouhiainen *et al*., 2004) catalyzing the removal of a water molecule from 3-hydroxyacyl-ACP thereby introducing a trans double bond in enoyl-ACP which is subsequently reduced to saturated acyl-ACP by an ER domain. FabA-like DH domains have higher sequence similarity including conserved catalytic motifs to FabA which is essential for UFA biosynthesis in *E. coli*.

CHAPTER 3. FUNCTIONAL ANALYSIS OF KS DOMAINS FROM PUFA SYNTHASE OF *THRAUSTOCHYTRIUM* IN *E.COLI*.

3.1 Abstract

Thraustochytrium sp. 26185 accumulates a high level of DHA, a nutritionally important ω-3 VLCPUFA synthesized primarily by the polyunsaturated fatty acid (PUFA) synthase, a type I PKS-like mega-enzyme. However, the molecular mechanism of the PUFA synthase for positioning multiple cis-double bonds in the acyl chain remains elusive. The PUFA synthase in this species comprises three large subunits each with multiple catalytic domains. It was hypothesized that among these domains, KS domains might be critical for catalyzing the condensation of specific unsaturated acyl-ACPs with malonyl-ACP, whereby retaining double bonds in the extended acyl chain. To investigate the function of these putative KS domains, two segment sequences from subunit A (KS-A) and from subunit B (KS-B) of the PUFA synthase were dissected and then expressed as standalone enzymes in *Escherichia coli*. The results showed that both KS-A and KS-B domains, but not mutagenized ones where the cysteine residue at the predicted active sites was substituted by a lysine residue, could complement defective phenotypes of both *E. coli FabB* and *FabF* mutants. Overexpression of these domains in a wild type *E. coli* showed that 1.4-fold, and a 1.45-fold increase in total fatty acid production, respectively, relative to the control. KS-B contributed a higher ratio of UFAs to SFAs than KS-A, while KS-A could improve overall production of the fatty acids more effectively, particularly for production of SFAs, implying that KS-A might be more comparable to FabF while KS-B might be more similar to FabB in the catalytic function. Successful complementation and functional expression of the embedded KS domains from the PUHA synthase in *E. coli* is the first step forward to the future mechanistic study through *in vitro* assays with purified domain proteins incubated with other components for the final elucidation of the molecular mechanism of the PUFA synthase for the biosynthesis of VLCPUFAs.

3.2 Introduction

VLCPUFAs are fatty acids with the chain length equal to or greater than 20 carbons and with two or more double bonds. They are classified into two categories, ω-3 and ω-6 fatty acids, depending on the location of the last double bond from the methyl end in an acyl chain (Scorletti & Byrne, 2013). VLCPUFAs, particularly DHA, are important for human health and wellbeing (Qi *et al*., 2004). DHA is an essential membrane component of neuron, retina and immune cells. In retina, DHA is a major fatty acid in rod outer segment, the photosensitive part of the photoreceptor cell (Qiu, 2003), it involves the signal transduction, rhodopsin activation, and rod and cone development in the photoreceptor. In the nervous system, DHA influences synaptic function through modulation of the endocannabinoid, a regulator of the synaptic system. In the adult brain, DHA is mainly accumulated in the cortical gray matter, it has positive effects on sensorimotor integration, attention-executive function and memory (Diau *et al*., 2005; Hadders-Algra, 2010). In addition, DHA is a precursor for docosanoid mediating various physiological activities in mammals. For example, neuroprotectin D1 (docosanoid) is a neuroprotective derivative of DHA participating in protection against retina and brain damage (Organisciak *et al*., 1996; Mukherjee *et al*., 2007; Reich *et al*., 2001). Furthermore, DHA is a ligand of nuclear transcriptional factors regulating the expression (Deckelbaum *et al*., 2006) of numerous genes at the transcriptional level by activating these transacting factors (Uauy *et al*. 2000). Therefore, maintaining an optimal level of DHA is crucial for human to stay healthy, and appropriate dietary supplementation of DHA is encouraged for protection against chronic diseases, such as hypertension, arthritis, arteriosclerosis, depression and thrombosis (Horrocks & Yeo, 1999).

There are two distinct pathways for the *de novo* biosynthesis of VLCPUFAs in nature according to whether it requires molecular oxygen during the biosynthetic process. The aerobic pathway usually occurs in cytosolic endoplasmic reticulum of eukaryotes and double bonds are introduced by an oxygenic desaturation process while fatty acid chains are elongated through the elongation process. Therefore, in the aerobic pathway, desaturases and elongases are two types of key enzymes for synthesizing VLCPUFAs (Qiu *et al*., 2001). So far, many desaturases and elongases in the pathway have been identified from a wide range of eukaryotes including animals, plants, fungi, microalga and other microorganisms. The anaerobic pathway, also called a polyketide synthase pathway, usually occurs only in prokaryotic and eukaryotic microbes. In this pathway, double bonds are introduced during the acyl chain-extending process without any

desaturation process. The biosynthesis is catalyzed by a PUFA synthase, a PKS-like enzyme. A PUFA synthase in microorganisms usually comprises one or more subunits, each subunit may contain multiple domains such as KS, KR, DH, ER, AT, MAT and ACP. These domains work coordinately for catalyzing VLCPUFA synthesis through reiterative reactions including condensation, ketoreduction, dehydration and enoylreduction.

Thraustochytrium sp. 26185 is a unicellular protist producing a high level of DHA in storage lipids. Our previous research indicates that there is an aerobic pathway in this species for the biosynthesis of VLCPUFAs, and DHA could be synthesized through the final step of $\Delta 4$ desaturation (Qiu *et al*., 2001). Our recent genome sequencing uncovered an alternative anaerobic pathway for the biosynthesis of VLCPUFAs in this species. However, the *in vitro* and *in vivo* assays showed that the latter pathway was primarily responsible for the biosynthesis of DHA, while the aerobic pathway might be the progenitor and relic system with missing components. Like similar enzymes identified from other marine microbes (Gong *et al*., 2014; Okuyama *et al*., 2007; Ruiz-López *et al*., 2012), the PUFA synthase from *Thraustochytrium sp.* 26185 is comprised of three subunits each with multiple catalytic domains. Among these domains, two putative KS domains from subunit A and subunit B might be critical for retaining double bonds by selectively condensating unsaturated acyl chain primers with an extender unit malonyl-ACP (Cui *et al*., 2015; Metz *et al*., 2001; Tang *et al*., 2003). In this work, the two putative embedded KS domains of the PUFA synthase from *Thraustochytrium* sp. 26185 were delimited by multiple alignments and homology modeling using protein sequences of ketoacyl synthase from *E.coli*, and KS domains from microbial Type I PKS and other PUFA synthases. The KS domains were then expressed as standalone enzymes for functional analysis through complementation test, site-directed mutagenesis and overexpression in *E. coli*. The results from these studies provide insight into functions of these KS domains of the PUFA synthase from the *Thraustochytrium*.

3.3 Experimental approach

3.3.1 Experimental flow

The experimental approach for this study shown as follows (Figure 3.1).
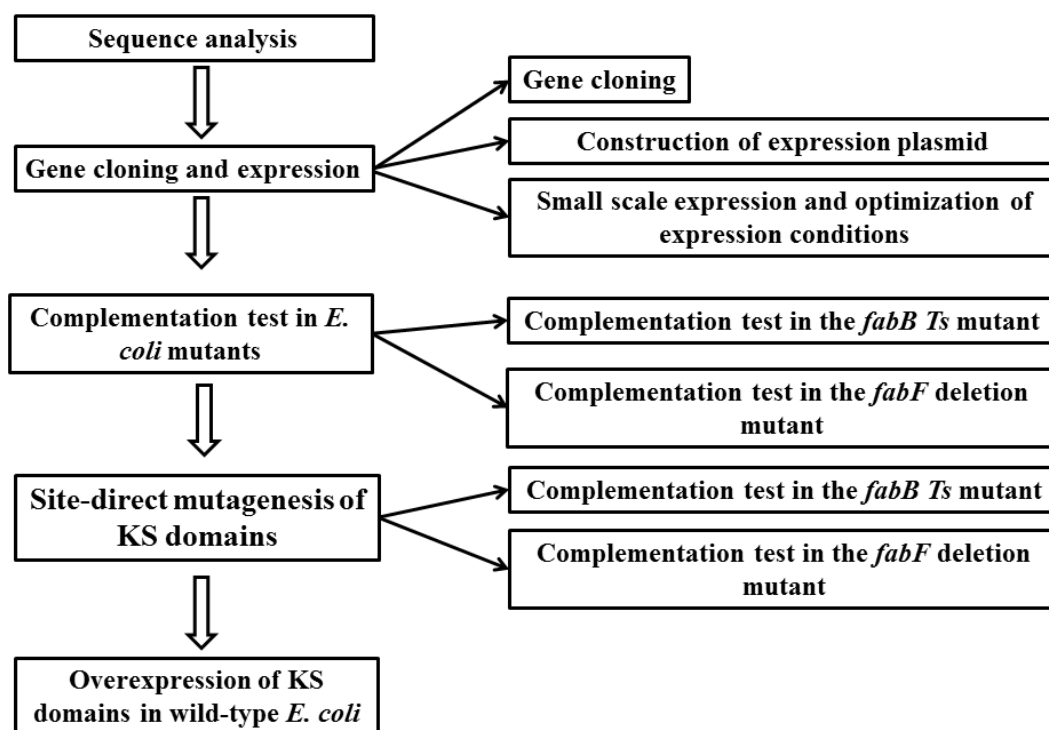
Figure 3.1 Experimental flow of Chapter 3.

3.3.2 Strains, Plasmid, media, and reagents

E. coli strains *Top*10 and *BL*21 (DE3) were purchased from Invitrogen Biotechnology Co. (Grand Island, USA) grown in the Luria-Bertani medium (1% peptone, 0.5% yeast extract, and 1% sodium chloride). The *E.*coli strain overexpressing the *E.*coli *FabB* was provided by Dr. Charles O. Rock, Department of Biochemistry, St. Jude Children's Research Hospital, Tennessee, USA (Heath & Rock, 1996). The temperature sensitive (*Ts*) *FabB* and knockout *FabF* mutants were purchased from National BioResource Project (NBRP *E. coli*, Microbial Genetics Laboratory, National Institute of Genetics, Japan)(Yamazaki & Sugawara, 2009). Expression vectors, such as pET28a, pET15b and pGEM-T, were purchased from Promega Co. (CA, USA). The plasmid extraction, DNA purification and gel extraction kits were obtained from Bio Basic Inc. (New York, Canada). HP Taq DNA polymerase was purchased from Bio Basic Inc. Q5 polymerase and

restriction enzymes, and dNTP were purchased from New England Biolabs (Ipswich, MA, USA). Primers were synthesized through Sigma-Aldrich (St. Louis, MO, USA).

3.3.3 Sequence analysis of the PUFA synthase in *Thraustochytrium*

The sequence analysis of KS domains from the PUFA synthase in *Thraustochytrium sp. 26185* was performed on the genome sequence. The three subunits (subunit-A, subunit-B and subunit-C) were used as input database and tools for cataloging FASs/PKSs including BLAST (NCBI), ClusterMine360 (Conway & Boddy, 2013), antiSMASH (Medema *et al*., 2011) and NRPS-PKS (Ansari *et al*., 2004) which were used to identify putative domains and active sites. The multi-sequence alignment of homologous sequences from *Aurantiochytrium*, *Schizochytrium*, *Emiliania huxleyi* and *Shewanella fidelis* by DNAMAN was used to delimit the borderlines of each domain, which was further confirmed by on-line homology modeling tools provided by the SWISS-MODEL (Biasini *et al*., 2014) and Phyre server (Kelley & Sternberg, 2009).

3.3.4 Cloning and expression of KS domains from PUFA synthase in *E. coli.*

To express the KS domains in *E. coli*, five primers were designed for amplifying them from subunit-A (KS-A), and subunit-B (KS-B and CLF) with His-tags at the N-terminus and a stop codon at the C-terminus. The KS domain from subunit-A was amplified with the two primers (F KS-A and R KS-A). The KS domain and the KS-CLF domain from ORF-B were amplified with the primers (F KS-B and R KS-B) and primers (F KS-B and R CLF-B), respectively (Table 3.1). Amplicons with the expected size were first cloned into an intermediate pGEM-T vector for sequencing confirmation.

Table 3.1 Primers used for KS domain amplification.

Name	Sequence (5'-3')	Annotation
F KS-A	<u>GAATTC</u>ATGGCGGCCCGCG	*Eco*R I site (underline)
R KS-A	<u>AAGCTT</u>CTACGCGTGCTCCGGCTC	*Hin*d III site (underline)
F KS-B	<u>GAATTC</u>ATGGCCGCGCGCAAC	*Eco*R I site (underline)
R KS-B	<u>AAGCTT</u>CTAGTGCTCCTCGAAGACGGCG	*Hin*d III site (underline)
R CLF-B	<u>AAGCTT</u>CTAGCGCTCGTGGTGGCC	*Hin*d III site (underline)

For the expression plasmid construction, the KS sequences from pGEM-T vectors and a *FabB* gene from pET15b vector were sub-cloned into a His-tagged vector pET28a(+) (Figure 3.2). The purified KS fragment from pGEM-T and pET28a(+) vector digested with *Eco*RI and *Hin*d III were ligated according to the standard procedure. Similarly, purified *FabB* fragment from pET15b and pET28a(+) digested with *Xba*I and *Bam*HI were ligated. The digested DNAs were run on an agarose gel with wide-toothed comb wells. After electrophoresis, target bands were cut from the gel and purified using BioBasic's miniprep kit (2015). Purified KS and FabB fragments and corresponding digested plasmids were ligated as follows: 2µL of 5× ligation buffer, 1.5µL of pET28a(+) linearized vector, 4.5µL of KS domain fragments and 0.5µL Invitrogen T4 DNA ligase at 4°C overnight. The ligations, KS/pET28a and *fabB*/pET28a, were transformed into competent *E. coli* Top10 cells. The transformation was carried out using a chemical transformation approach. The sample was incubated at 37°C for 45 minutes and shaked vigorously (250 rpm), and placed on LB agar plate containing 50µg/mL kanamycin for selection. The selection plates were incubated at 37°C overnight, and putative transformants were selected, sub-cultured, and screened by Polymerase chain reaction (PCR) using a universal primer (T7 promoter) and a gene-specific primer. Positive transformants were inoculated into 5mL of liquid LB medium plus 50µg/mL kanamycin and incubated at 37°C overnight. The plasmids were extracted using BioBasic's miniprep kit (2015), and digested with restriction enzymes to confirm the recombinant plasmids.
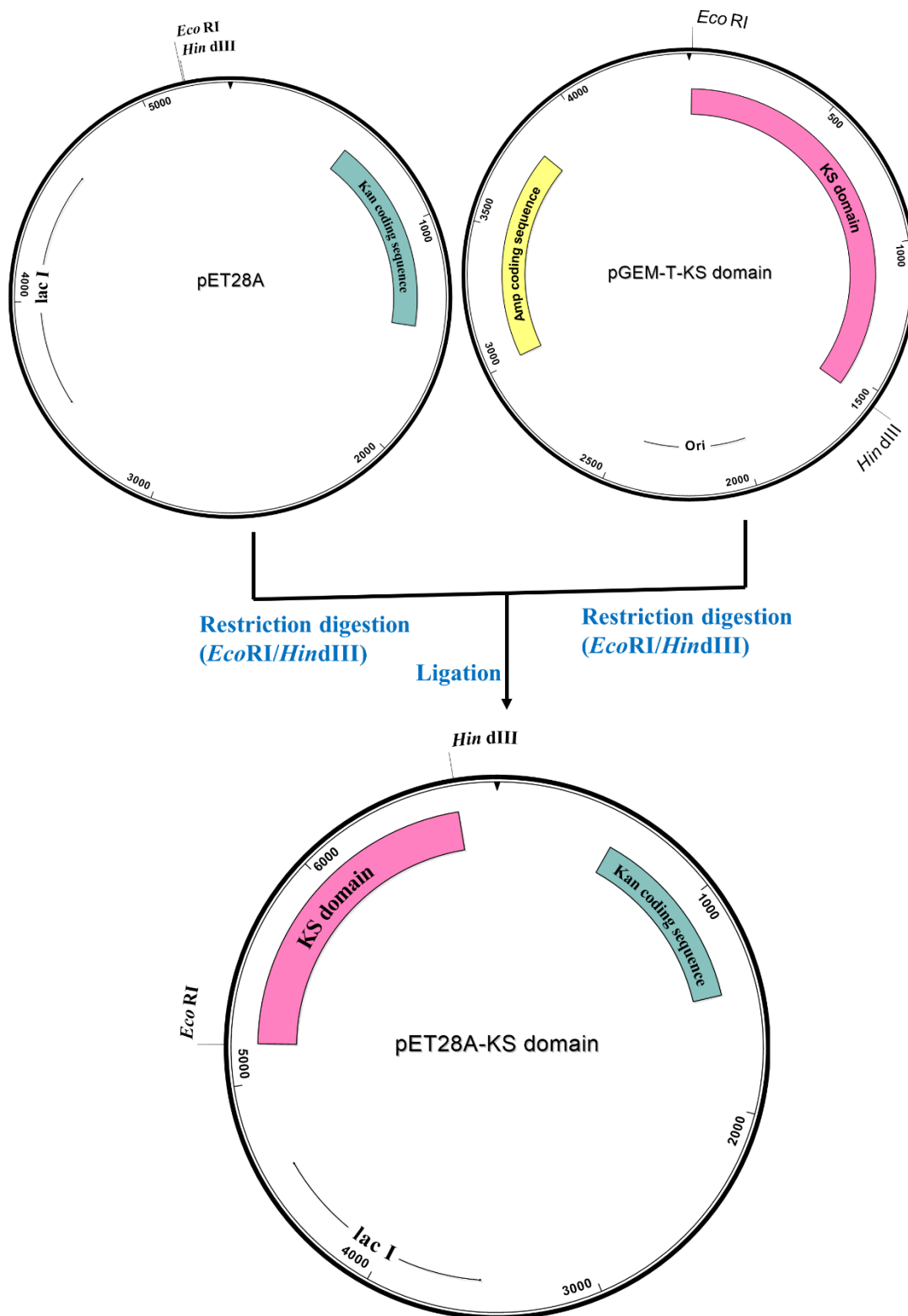
Figure 3.2 Construction of the recombinant plasmid expressing KS domains with pET28a.

The confirmed plasmids pET28a-KS-A, pET28a-KS-B, pET28a-KS-CLF-B, and pET28a-*fabB*, were transformed into competent *E. coli BL*21 (DE3) cells for expression of the recombinant plasmids. Small scale expression of target proteins was conducted before scaling up the expression procedure. A single colony from the host *E. coli BL*21 (DE3) harboring an expression plasmid was grown overnight in 10 mL of LB medium with 50 μg kanamycin/mL at 37 °C. A 200 μL aliquot of the resulting culture was inoculated into 20 mL of fresh LB medium containing kanamycin and incubated aerobically at 37 °C until an OD_{600} of 0.4–0.6 reached. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the final concentration of 1 mM for inducing the expression. After 12-hour induction, the cell was harvested using 6,000 rpm centrifugation (Eppendorf centrifuge 5804 R) at 4°C for 10 min, and resuspended in 5 volumes of ice-cold buffer (20 mM Tris-HCl, pH7.0). The cells were broken down through beadbeater (MINI BEAD BEATER 8, Biospec Products), and the mixture was centrifuged at 15,000 rpm at 4°C. The proteins in cell pellets and supernatant in 200 μL of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer (60 mM Tris-HCl, pH 6.8, 2% SDS, 0.1% bromophenol blue, 25% glycerine and 14.4 mM β-mercaptoethanol) were separated by electrophoresis. To improve protein expression, the transformants were cultivated under different temperature (20°C, 25°C, 37°C), or with different concentrations of IPTG (1 mM, 0.5 mM, 0.01 mM).

3.3.5 Complementation of *E. coli* mutants defective in the ketoacyl-ACP synthase activity with the KS domains

The complementation tests were conducted in two *E. coli* mutants defective in ketoacyl-ACP synthase I (FabB) and II (FabF) activity, respectively. The *FabB* (*Ts*) temperature sensitive mutant was derived from *E.coli* strain K-12, W3110 mutagenized by N-methyl-N′-nitro-N-nitrosoguanidine (NTG). This mutant could not grow at 42°C (but could survive at 30°C). The temperature sensitive phenotype could be complemented with infection of a Kohara phage carrying a wild type *FabB* gene. The *FabF* deletion mutant is a knockout mutant from the Keio collection where the *FabF* gene was deleted (Baba *et al*., 2006).

Complementation test in the *FabB Ts* mutant

The competent cells of the *FabB Ts* mutant were prepared according to Chung (1989). The recombinant plasmids pET28a-KS-A, pET28a-KS-B, pET28a-KS-CLF-B, pET28a-*fabB* and

empty vector pET28a, were transformed into competent *Ts FabB* mutant cells as described above and the transformants were spread on the plates containing 50 µg/mL kanamycin and grown at 30°C overnight for selection. The putative transformant colonies were selected and screened by PCR with a universal primer (T7 promoter primer) and a gene-specific primer. Positive colonies were inoculated into 5 mL of liquid LB medium plus 50 µg/mL kanamycin and incubated at 30°C overnight. For the complement plate assay, 2 µL cell cultures were dropped in duplicate on a LB-agar plate with 50 µg/mL kanamycin and 1mM IPTG. The plates were incubated overnight at 30°C and 42°C, respectively.

The fatty acid profiles of the cell cultures were obtained through Gas chromatography (GC) analysis. A single colony from transformants harboring the expression plasmid was grown overnight in 10 mL of LB medium with 50 µg/mL kanamycin at 30°C. A 200 µL aliquot of the resulting culture was inoculated into 20 mL of fresh LB medium containing kanamycin and grown aerobically at 30 °C until an OD_{600} of 0.4–0.6 reached. IPTG was added to the final concentration at 0.5 mM for inducing the expression. After 12-hour expression, the cells were harvested using 6,000 rpm centrifugation at 4°C for 10 min, and washed three times with 5 volumes of PBS buffer (20 mM PBS pH7.0). The samples were measured using a NanoDrop 2000c spectrophotometer and adjusted to OD_{600}=1.0 with PBS buffer. An equal amount of cells (10 mL) was used for fatty acid analysis with heptadecanoic acid (17:0) as an internal standard. For the analysis, 2 mL of 1% sulfuric methanol was added to cell pellets, and the mixture was incubated at 80°C for 1 hour. After cooling down on ice, the sample was added with 2 mL hexane and 1mL 0.9% NaCl. The mixture was centrifuged for 10 minutes at 2,200×g. The upper hexane phase was transferred to a new glass tube and dried with N_2 gas. After drying, 200 µL of hexane was added to resolubilize the lipid, transferred into insert-containing vials for the GC analysis using Agilent 7890A system with an auto-injector G4513A.

Complementation test in the *FabF* deletion mutant

For preparing the expression plasmids for complementation test in *FabF* deletion mutant, KS domain from pET28a was sub-cloned into the expression vector pET15b (Figure 3.3). The confirmed recombinant plasmids were then transformed into the competent *fabF* deletion mutant and selected on the plates containing 50 µg/mL kanamycin and 100 µg/mL ampicillin. For screening the positive colonies, a universal forward primer (T7 promoter) and a gene-specific

reverse primer were used for PCR reaction. The positive colonies were inoculated into 5 mL of liquid LB medium plus 50 µg/mL kanamycin and 100 µg/mL ampicillin and grown at 37°C overnight. For the fatty acid analysis, a single colony the mutant transformants harboring the expression plasmid were grown overnight in 10 mL of LB medium with 50 µg/mL kanamycin and 100 µg/mL ampicillin at 37 °C. A 200 µL aliquot of the overnight culture was inoculated into 20 mL of fresh LB medium containing kanamycin and ampicillin and grown aerobically at 37 °C until an A_{600} of 0.4 reached. IPTG was added to the final concentration at 0.5 mM for inducing the expression. Fatty acid analysis of the *FabF* deletion mutant carrying the recombinant plasmids was conducted as described above.
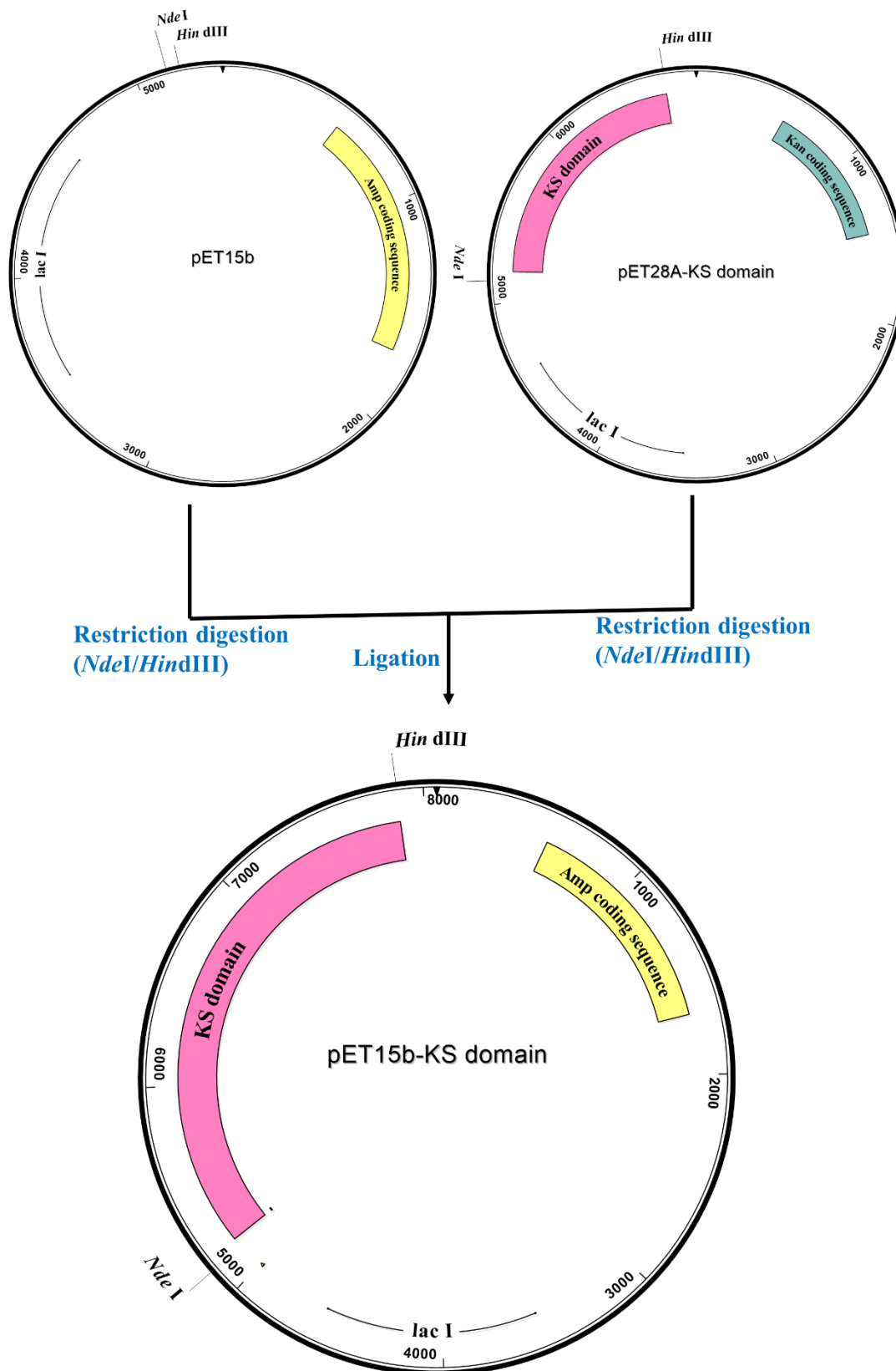
Figure 3.3 Construction of the recombinant plasmid expressing KS domains with pET15b.

3.3.6 Site-directed mutagenesis of the KS domains

Site-direct mutagenesis was carried using overlapping PCR according to Ho (1989). Two pairs of internal primers with complementary ends and desired mutation sites were designed for PCR amplification to create mutants of the KS domains (Table 3.2). The recombinant plasmids pET28a-KS-A, pET28a-KS-B, and pET28a-KS-CLF-B carrying wild type KS-A, KS-B and KS-B-CLF fragments were used as templates for the mutagenesis. For each mutagenesis, the first round of PCR was performed with one flanking primer and one internal primer and the second round of PCR was conducted with another flanking primer and the other internal primer. For mutagenized KS-A, two DNA fragments were amplified by overlapping PCRs with two sets of primers (T7 promoter primer and R KS-M-A, F KS-M-A and T7 terminator primer) using pET28a-KS-A as a template. For mutagenized KS-B and KS-CLF-B, DNA fragments were amplified by overlapping PCRs with two sets of primers (T7 promoter primer and R KS-M-B, F KS-M-B and T7 terminator primer) using pET28a-KS-B and pET28a-KS-CLF-B as templates, respectively (Table 3.2). After PCR amplification, the mutagenized KS domains (named KS-M-A, KS-M-B, and KS-M-CLF-B) were subcloned into pET28a or pET15b and transformed into *E.coli Top10* cells for propagation of recombinant plasmids. The recombinant plasmids pET28a-KS-M-A, pET28a-KS-M-B, and pET28a-KS-M-CLF-B, were confirmed by restriction digestion and DNA sequencing. The confirmed plasmids were transformed into *E. coli* strains for functional analysis.

Table 3.2 Primers used for site-directed mutagenesis

Name	Sequence (5'-3')	Annotation
F KS-M-A	GCCTCGAAATCGCTCATCGCCGTC	mutation site (underline, Cys to Lys)
R KS-M-A	GGCGGTTTCGTCGACGACGC	mutation site (underline, Cys to Lys)
F KS-M-B	GCGACCAAAGCGCTCTACGTGCTC	mutation site (underline, Cys to Lys)
R KS-M-B	AGCGGCTTTGTCGACCGAGTAGTG	mutation site (underline, Cys to Lys)

Lipid extraction and GC analysis

A single colony from *E. coli* transformants harboring the expression plasmid was grown overnight in 10 mL of LB medium with 50 µg/mL kanamycin, or 50 µg/mL kanamycin and 100 µg/mL ampicillin at 30°C. A 200 µL aliquot of the overnight culture was inoculated into 20mL of fresh LB medium containing kanamycin, or kanamycin and ampicillin, and incubated aerobically at 30°C until an OD_{600} of 0.4 reached. IPTG was added to a 0.5 mM final concentration for inducing the expression. Fatty acids profiles of transformants carrying original or mutagenized KS domains were analyzed by GC as described above.

3.3.7 Overexpression of the KS domains in wild type *E. coli*

Overexpression was performed in the wild type strain *E.coli BL*21 (DE3). The expression plasmids pET28a-KS-A, pET28a-KS-B, pET28a-KS-CLF-B, pET28a-*fabB,* and control vector pET28a were transformed into the competent cells. The transformants were then spread on the plates containing 50 µg/mL kanamycin. A single colony of the positive transformant harboring the expression plasmid was grown overnight in 10 mL of LB medium with 50 µg/mL kanamycin at 37°C. A 500 µL aliquot of the overnight culture was inoculated into 50 mL of fresh LB medium containing kanamycin and incubated aerobically at 37 °C until an OD_{600} of 0.4 reached. After adding IPTG, the cell culture was divided into two equal volumes and incubated at 20°C and 37°C respectively for inducing the expression. Fatty acids profiles of *E. coli BL*21 (DE3) transformants expressing different KS domains at different temperatures were analyzed by GC as described above.

3.4. Results

3.4.1 Sequence analysis of KS domains of a PUFA synthase in *Thraustochytrium sp.* 28165

To study the biosynthesis of VLCPUFAs in *Thraustochytrium sp.* 28165, we recently sequenced the whole genome of the protist. Analysis of the genome sequence revealed co-existence of a standard aerobic pathway and an alternative anaerobic pathway for the biosynthesis of DHA in this species. Our *in vitro* and *in vivo* assays showed that DHA was primarily synthesized by the PUFA synthase, a mega-enzyme comprising three large subunits encoded by three open reading frames (ORFs), each with multiple catalytic domains that were strongly predicted by the presence of characteristic active site sequence motifs (Figure 3.4). One KS domain (KS-A) was

identified in subunit-A while the other KS domain (KS-B) was found in subunit-B. The two KS domains had amino acid sequence identity with each other at about 20% and both shared sequence similarity with FabB and FabF, two condensing enzymes involved in the biosynthesis of fatty acids in *E. coli*. In addition, a CLF-like domain was also identified adjacently to KS-B in subunit-B that was considered as a catalytically inactive version of a KS domain in a PKS system. The side-by-side arrangement of KS and CLF domains was believed to be critical for biosynthesis of polyketides (Keatinge-Clay, 2012). Boundary sites of these domains were determined through multiple sequence alignments and homology modeling using protein sequences of prokaryotic ketoacyl synthases, and KS domains from microbial Type I PKS and other PUFA synthases. The KS-A domain ranged from residue 1 to 492 in subunit-A, while the KS-B and CLF domains were from residue 1 to 469 and 470 to 1023 in subunit-B, respectively. For the functional characterization of these domains, the delimited DNA regions were amplified and expressed as standalone proteins in *E. coli*.
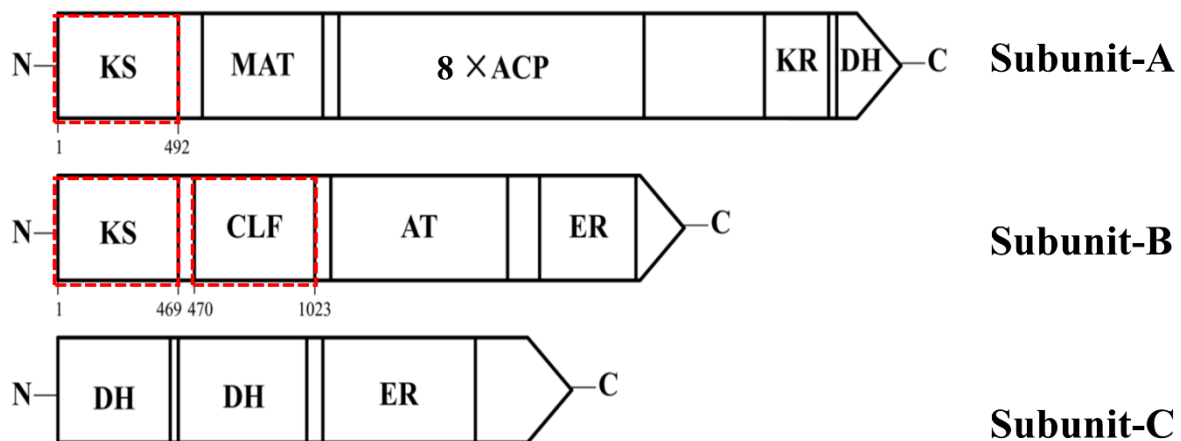
Figure 3.4 Structure of PUFA synthases from *Thraustochytrium*

3.4.2 Expression of the KS domains in an *E. coli FabB* mutant

E. coli possesses three condensing enzymes, FabB, FabF and FabH for the fatty acid biosynthesis. FabH catalyzes the initial condensation of acetyl-CoA with malonyl-ACP to ketobutyryl-ACP. FabB and FabF with high sequence similarity with each other catalyze subsequent condensation reactions. However, FabF could specifically elongate long chain fatty acids such as palmitic acid (16:0) and palmitoleic acid (16:1-9) while FabB could catalyze elongation of fatty acids with chain length from 4 to 14 carbons (Wang & Ma, 2013). As both KS domains from the PUFA synthase showed sequence homology to FabB and FabF, the embedded domains of the PUFA synthase were excised and then expressed as standalone enzymes in the *FabB* and *FabF* mutants for complementation assays.

FabB was an essential gene for the biosynthesis of unsaturated fatty acids for *E. coli* to maintain the homeostasis of cell membrane (Broekman & Steenbakkers, 1973; Campbell & Cronan, 2001; Rosenfeld *et al*., 1973), therefore, no knockout strain of this gene was readily available for the complementation test. Thus, complementation assays of the KS domains were carried out in a *FabB* (*Ts*) mutant where the chromosomal *FabB* gene was chemically mutagenized, as such the *FabB* mutant became temperature sensitive and could grow at 30°C, but not at a non-permissive temperature (42°C). However, the *FabB* mutant carrying KS-A, KS-B and KS-CLF-B domains as well as the wild type *FabB* were all capable of growing at both temperatures (30°C and 42°C), in contrast to the mutant with the empty vector that ceased growth at 42°C on the plate (Figure 3.5). This result clearly showed that both KS domains from subunit-A and subunit-B could complement the defective phenotype of *FabB* and restore the function for the biosynthesis of unsaturated fatty acids in *E. coli*.

Figure 3.5 Complementation assays of the KS domain proteins in the *E. coli FabB Ts* mutant on the plate. KS domains and *E. coli FabB* were inserted behind the T7 promoter of the expression vector pET28a individually. The recombinant plasmids were transformed into the *FabB Ts* mutant where the chromosomal *FabB* gene is mutagenized. The plates with kanamycin and IPTG were incubated overnight at 30°C and 42°C, respectively.

When grown at the permissive temperature (30°C), the *FabB* mutant, like its wild type, produced six major fatty acids including myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), cis-vaccenic acid (18:1-11), and 11,12-methylene-octadecanoic acid (19:0-cyc). The two cyclopropane fatty acids (17:0-cyc and 19:0-cyc) were synthesized by cyclopropane fatty acid synthase (CFA) using unsaturated fatty acids 16:1-9 and 18:1-11 as substrates, respectively (Grogan & Cronan, 1997). However, compared to the wild type, the amount of these fatty acids in the mutant was substantially reduced, particularly 16:0, 16:1-9, 17:0-cyc and 18:1-11, which led to drastic reduction in the amount of the total fatty acids in the mutant strain. On the other hand, expression of the KS domains from the PUFA synthase in the mutant, like expression of the wild type *FabB* in the mutant, could clearly reverse

the defective phenotype in the fatty acid production. In particular, when KS-A and KS-B were expressed in the mutant, the amounts of fatty acids such as 16:0, 16:1-9, 17:0-cyc and 18:1-11 were comparable (KS-B) to or even higher (KS-A) than WT control, which were all significantly higher than that in the mutant control with the empty vector (Figure 3.6).
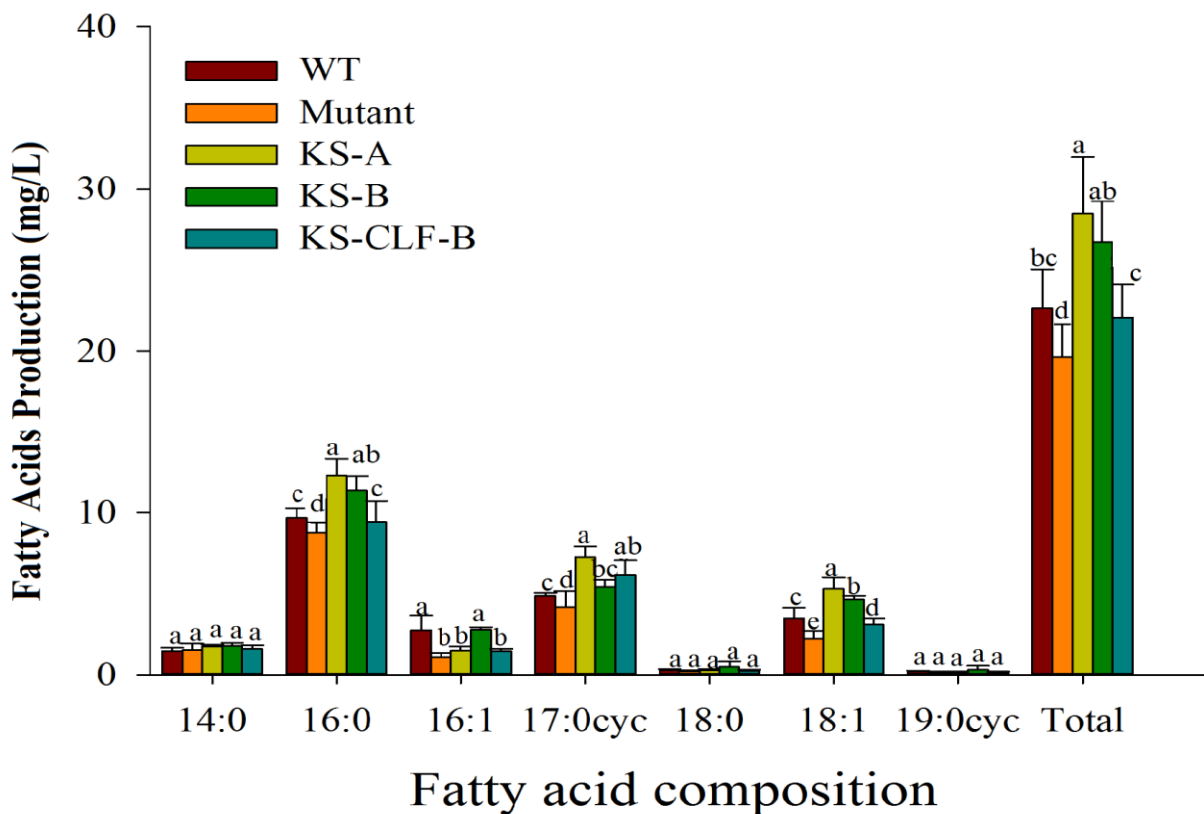
Figure 3.6 The amount of fatty acids produced in the *E. coli FabB* mutant expressing the KS domains from the PUFA synthase individually. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significant different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12- methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

37

3.4.3 Expression of the KS domains in the *E. coli FabF* mutant

Unlike *FabB, FabF* is a non-essential gene for *E. coli,* thus the plate assay could not be employed for the complementation test of the KS domains in the mutant. Instead, complementation assays of the KS domains in a *FabF* mutant (knockout) were conducted only through fatty acid analysis of the mutant and the mutant transformants expressing the KS domains. In comparison with the wild type control, nearly all fatty acids were significantly reduced in the mutant, leading to about 4-fold decrease in the production of the total fatty acids (Figure 3.7). However, when the KS domains were expressed in the mutant, the transformants restored the fatty acid production to the level similar to that in the wild type. In particular, levels of 16:1-9 and 17:0-cyc were significantly increased in the mutant transformant expressing the KS-A domain, which resulted in the amount of the total fatty acids being even slightly higher than that in the wild type. This result indicated that these KS domains could also functionally complement the defective phenotype of the *E. coli FabF* mutant.
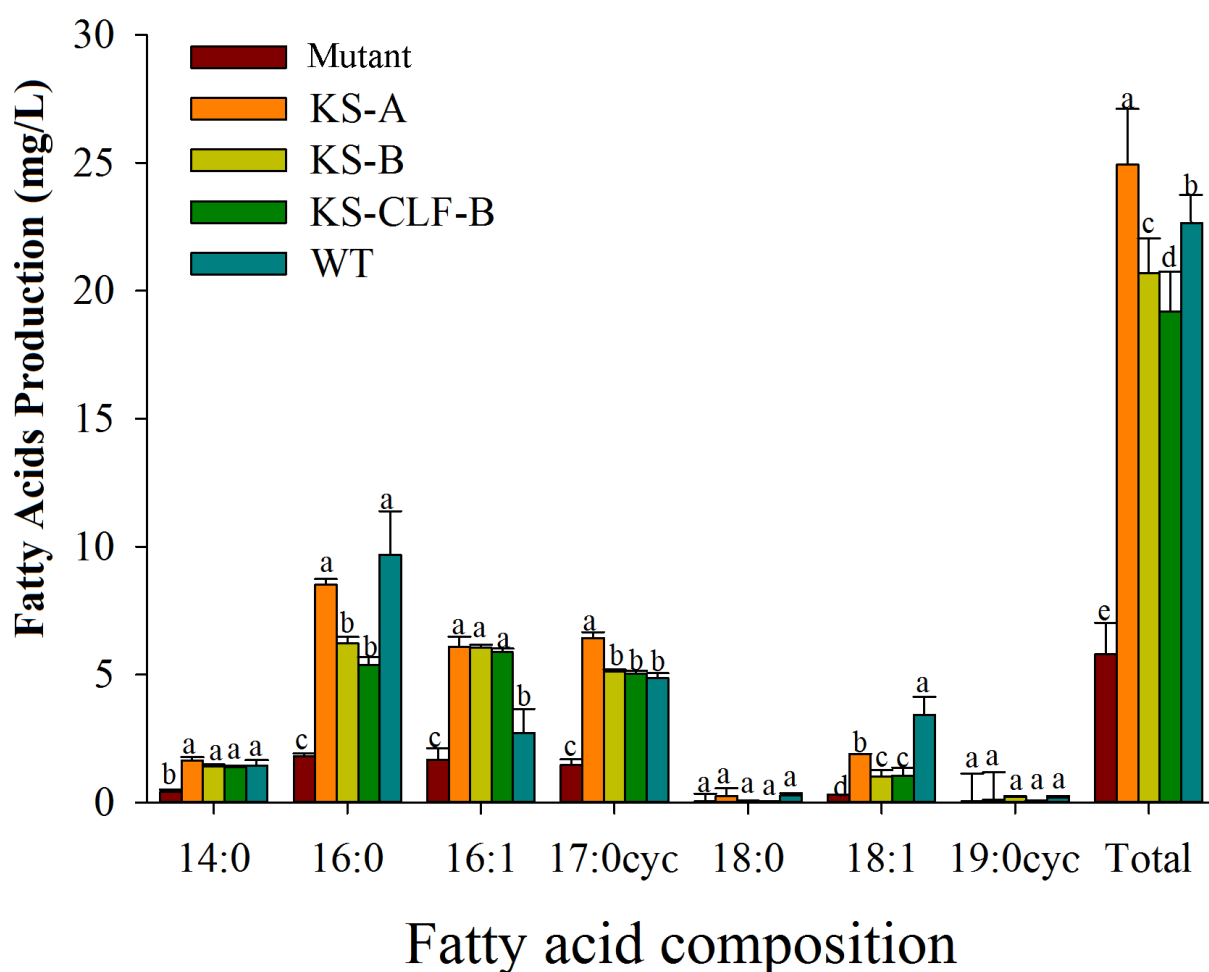
Figure 3.7 The amount of fatty acids produced in the *E. coli FabF* mutant expressing the KS domains from the PUFA synthase individually. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12- methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

3.4.4 Functional analysis of the mutagenized KS domains

Identity of the KS domains in the PUFA synthase was predicted based on the presence of characteristic residues (Cys-His-His) at their putative active sites. To examine the authenticity of these residues for the function of the KS domains, the cysteine residue, which is necessary for nucleophilic attack to substrate, at the predicted active sites of both KS domains was individually replaced by a lysine residue using site-directed mutagenesis and the mutant domains were then expressed as standalone enzymes for the complementation analysis. Complementation assays of the two mutagenized KS domains in the *FabB* mutant showed that, like the original mutant, the mutant transformants expressing the corresponding mutagenized KS domains could grow only at 30°C, but not at the non-permissive temperature 42°C, which was in contrast to the mutant transformants expressing the wild type KS domains that could grow at both temperatures (Figure 3.8). This result reaffirmed function of the KS domains and essentiality of the cysteine residue of the domains for the catalytic function as FabB for the biosynthesis of unsaturated fatty acid in *E. coli*.
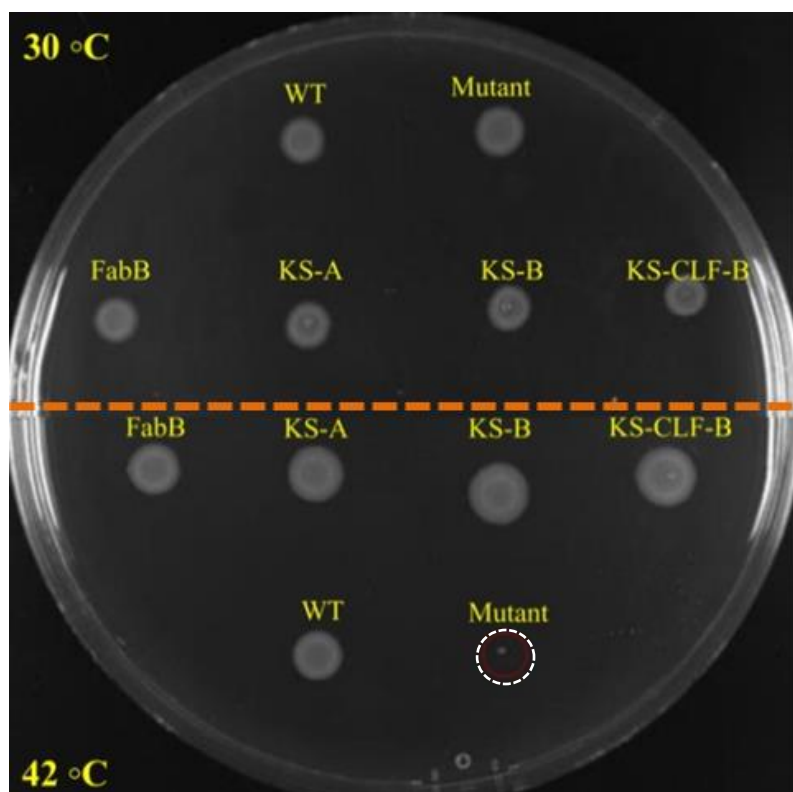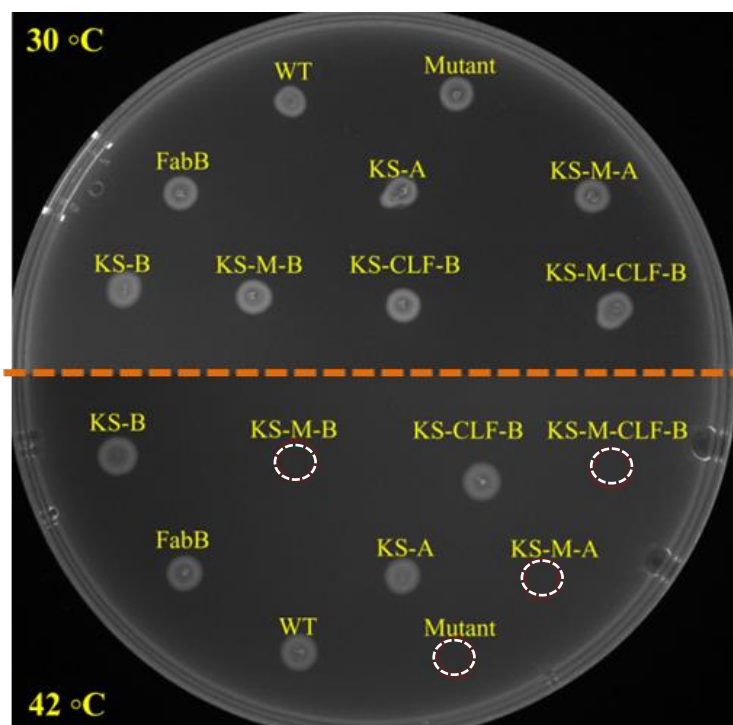
Figure 3.8 Complementation assay of the mutagenized KS domain proteins in the *E. coli FabB Ts* mutant on the plate. The KS domains and *E. coli FabB* were inserted behind the T7 promoter of the expression vector pET28a, individually. The plates were incubated overnight at 30°C and 42°C, respectively.

Complementation assays of the mutagenized KS domains in the *FabF* knockout mutant showed that the fatty acid production in the mutant expressing both mutagenized KS domains was comparable to that in the original mutant where the amount of almost all the fatty acids were substantially reduced in comparison to that in the wild type control, resulting in the drastic reduction of the total fatty acids in the cells. This result further confirmed that the cysteine residue was also critical for the function of the KS domains to complement activity of FabF in *E. coli* (Figure 3.9).
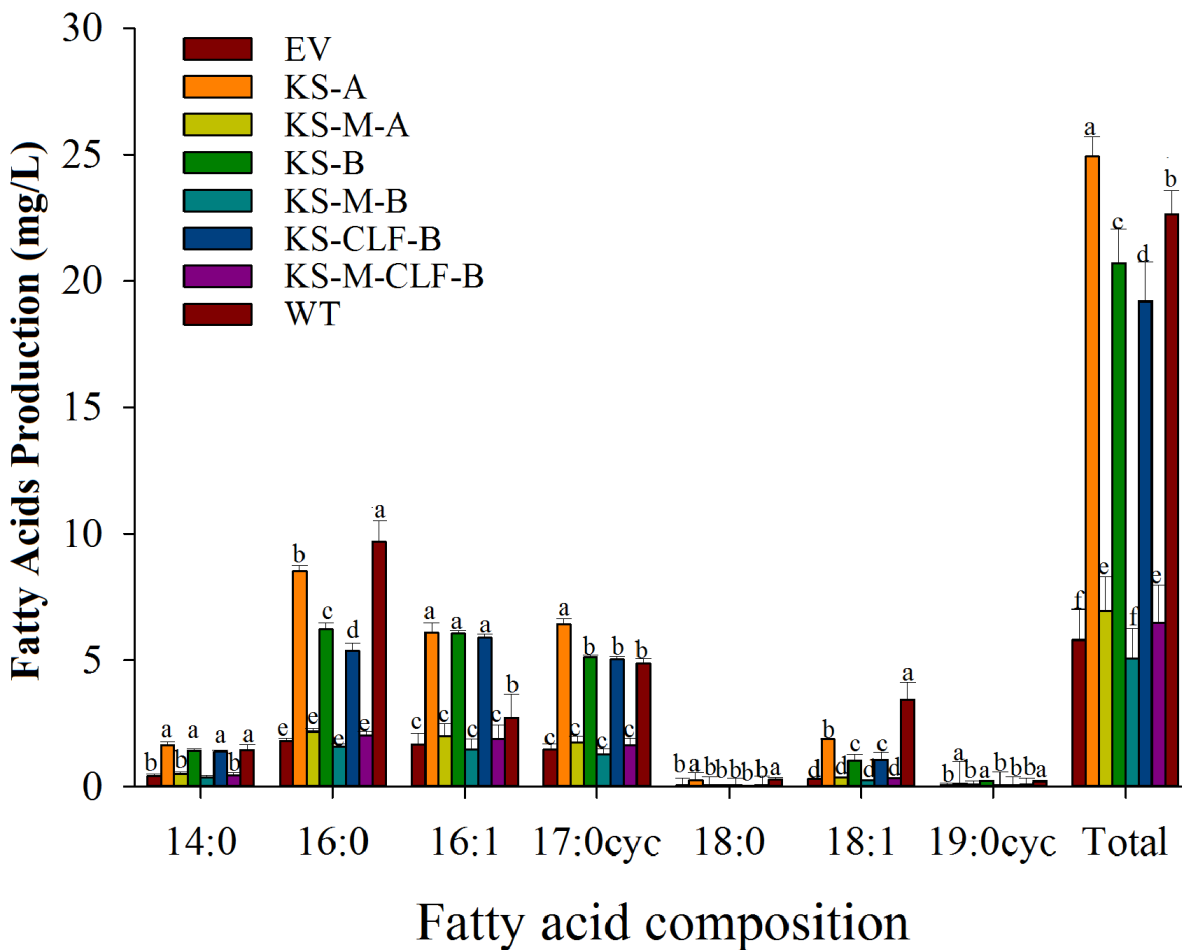
Figure 3.9 The amount of fatty acids produced in an *E. coli FabF* mutant expressing mutagenized KS domains from the PUFA synthase. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12- methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

3.4.5 Overexpression of the KS domains in *E. coli*

To further investigate functional roles of the KS domains, we overexpressed these domains as discrete enzymes in a wild type *E. coli* strain grown at two different temperatures (20°C and 37°C) to see if they could improve production of the fatty acids. At 37°C, the amounts of the total fatty acids produced in the wild type as well as the transformants overexpressing the KS domains were relatively low compared to those at 20°C. In addition, there was no significant difference in the production of either individual fatty acids or the total fatty acids among all the strains. However, at 20°C, all the transformants overexpressing the KS domains produced significantly higher levels of the total fatty acids than that in the wild type. In particular, an overexpression of KS-A improved the production of the total fatty acids significantly over the wild type strain. The amount of the total fatty acids produced in the overexpression reached about 41 mg/L, while that in the wild type was only about 24 mg/L. The approximately 45% increase in the overall fatty acid production was largely realized by substantial increase in 16:0, 16:1-9, 17:0cyc and 18:1-11 (Figure 3.10). Moreover, when the relative levels of UFAs and SFAs produced in these strains were compared, all KS overexpressions had a higher ratio of UFAs/SFAs relative to that in the wild type. The ratio in the KS-B overexpression was the most significant among all the strains, even higher than that in the KS-A overexpression, despite the KS-A overexpression was slightly higher in the production of the total fatty acids than the KS-B overexpression. The higher production of the total fatty acids in the KS-A overexpression was mainly reflected by a higher production in SFAs, as the level of UFAs produced in the KS-A overexpression was slightly lower than that in the KS-B overexpression. This result indicated that overexpression of the two KS domains could significantly improve the overall production of the total fatty acids and the KS-B domain was more effective in production of unsaturated fatty acids in *E. coli* in comparison of the KS-A domain (Table 3.3).
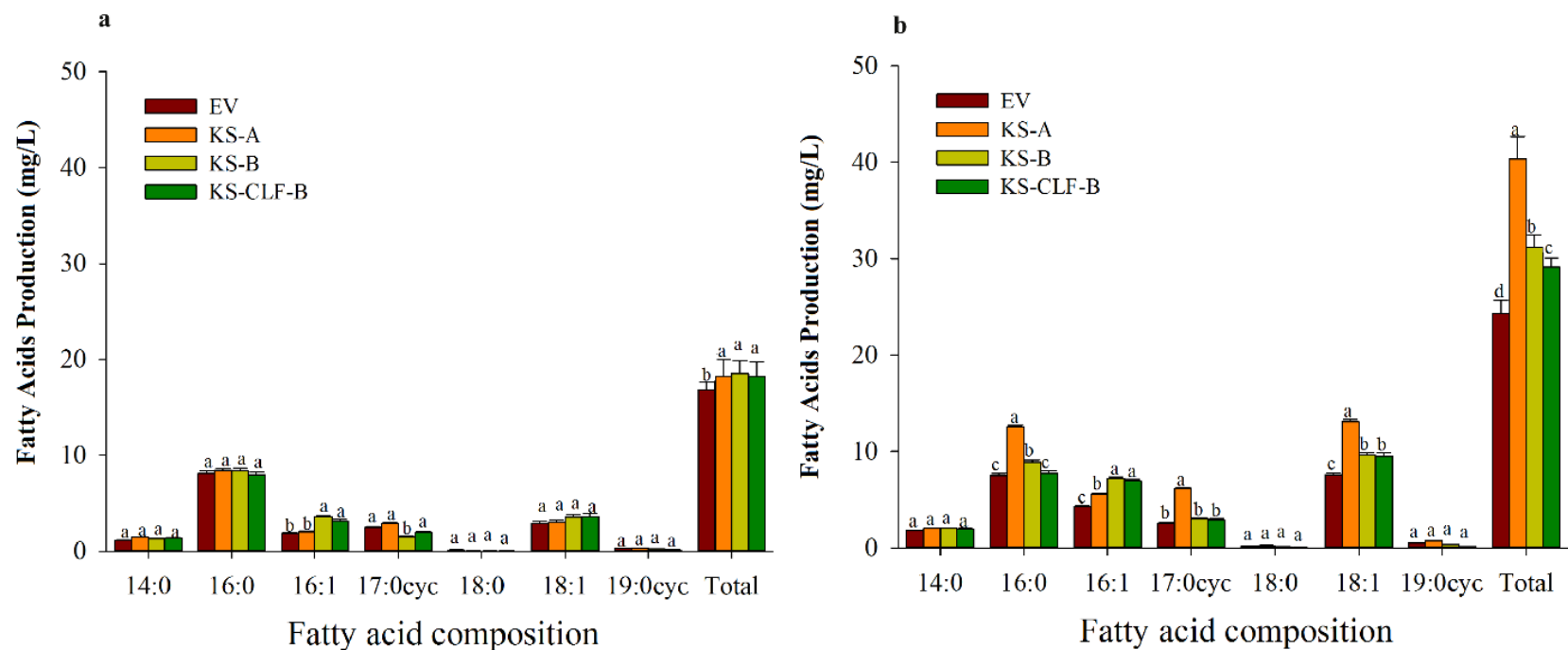
Figure 3.10 Fatty acid production of the KS overexpressions in the wild type *E. coli* at two different temperatures. a. Culture temperature at 37°C, **b.** Culture temperature at 20°C. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12- methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

44

Table 3.3 Production of saturated and unsaturated fatty acids in overexpressions of the KS domains in the wild type *E. coli*.

Recombinants	Total	SFA	UFA	UFA/SFA[a]
EV	24.34±1.33[d]	9.42±1.09[c]	14.92±0.85[d]	1.58
KS-A	40.35±2.40[a]	14.78±1.77[a]	25.56±1.12[a]	1.73
KS-B	31.15±1.69[b]	10.00±1.02[b]	21.14±0.66[b]	2.11
KS-CLF-B	29.13±1.87[c]	9.68±0.95[c]	19.45±0.72[c]	2.01

[a]Signifies the ratio of UFAs to SFAs and the two types of CFAs were considered as UFAs. Statistical analysis of the results from the experiments was conducted using oneway analysis of variance ($P<0.05$).

3.5. Discussion

The PUFA synthase has been identified in many marine microorganisms and is used for the biosynthesis of various VLCPUFAs, such as DHA, EPA, DPA and ARA (Gong *et al*., 2014; Okuyama *et al*., 2007). However, the exact molecular mechanism of this enzyme for intricately positioning multiple double bonds in these VLCPUFAs is unknown. In *Thraustochytrium sp*. 26185, this enzyme comprises three subunits, each with multiple domains predicted by the presence of characteristic active site sequences. However, authentic roles of these putative domains are not biochemically defined. It is assumed that key activities might be conferred by DH and KS domains of the PUFA synthase where DH domains are used for introducing double bonds while KS domains are for retaining these double bonds during the biosynthesis of VLCPUFAs. In this study, we attempted to dissect and express embedded KS domains of the PUFA synthase from the *Thraustochytrium* in *E. coli* to interrogate the function of these domains.

E.coli is a well-studied model using a Type II fatty acid synthase to produce both SFAs and UFAs. FabB is a key enzyme for the biosynthesis of unsaturated fatty acids in catalyzing the condensation of cis-3-decenoyl-ACP (formed by the FabA catalytic reaction) with malonyl-ACP (Lu *et al*., 2004), while FabF, despite a non-essential enzyme, has an important function in elongating long chain acyl-ACP, particularly for the formation of vaccenic acid (C18:1-11c) (Janßen & Steinbüchel, 2014). Both KS-A and KS-B domains from the *Thraustochytrium* PUFA synthase share sequence similarity to FabB and FabF, thus our first step for functional characterization of these putative KS domains is to express them as standalone enzymes in the two *E. coli* mutants. The complementation test of the *FabB* mutant showed that both KS domains could complement the defective phenotype and the mutant transformants expressing both domains could grow well at 42°C, a non-permissive temperature for the *FabB* mutant. Fatty acid analysis of the transformants confirmed that the mutant transformants with both KS domains produced a high level of UFAs (16:1-9 and 18:1-11) to sustain the growth at the high temperature (42°C), as seen in the wild-type strain. Particularly, the KS-B domain could accumulate a higher level of UFAs than the KS-A domain. The complementation test of the *FabF* mutant also showed that both KS domains could complement the defective phenotype in the fatty acid production. Particularly, the mutant transformants with KS-A domain could produce a high level of 18:1-11, as seen in the wild type strain.

Functional roles of the KS domains from the PUFA synthase was further reaffirmed by site-directed mutagenesis and overexpression. The cysteine at the predicted active sites of these domains is highly conserved in discrete ketoacyl-ACP synthase enzymes and KS domains of fatty acid synthase and polyketide synthase in animals, plants and microbes (Cui *et al*., 2015; Keatinge-Clay, 2012; Lomakin *et al*., 2007; Lu *et al*., 2004; Olsen *et al*., 2004; Schweizer & Hofmann, 2004; Smith *et al*., 2003; Xu *et al*., 2013). The thiolate form is believed to serve as a ''nucleophilic elbow'' necessary for the nucleophilic attack in the transacylation reaction (Schweizer & Hofmann, 2004). When this residue in the two putative KS domains was changed to a lysine residue, the substitutions could not complement the defective phenotype of the *FabB* mutant and restore the activity of FabF in the mutant. Overexpression of FabB and FabF have been previously shown to increase fatty acid production in *E. coli* (Jeon *et al*., 2012; Lee *et al*., 2013). As KS-A and KS-B can functionally complement the two enzymes, we then overexpress the two KS domains in a wild type strain to see if they can produce similar effects. Indeed, over-expressions of KS-B and KS-A

result in 1.4-fold, and a 1.45-fold increase in total fatty acid production, respectively, relative to the wild type control at the low temperature. The improvement in the total fatty acid production reconfirms biochemical functions of the putative KS domains from the PUFA synthase in fatty acid biosynthesis in *E. coli*. The strain carrying KS-B accumulates a high level of UFAs by retaining a double bond, as seen in a strain overexpressing the *FabB* gene. The KS-A domain has an authentic ability in elongating 14:0 to 16:0 and C16:1 to C18:1, as seen in a strain overexpressing FabF (Parsons & Rock, 2013). The fact that KS-B contributes a higher ratio of UFAs to SFAs than KS-A, while KS-A can improve overall production of the fatty acids, particularly for production of SFAs in the transformants suggests that KS-A might be comparable to FabF while KS-B might be more similar to FabB in the catalytic function.

VLCPUFAs can be synthesized by either an aerobic or an anaerobic pathway in living organisms. The latter pathway is advantageous in the productivity as it is a *de novo* fatty acid system where the final products are synthesized by a single enzyme complex starting from acetate. On the other hand, the former involves different kinds of enzymes, such as fatty acid synthase for the synthesis of long chain saturated fatty acids (16:0 and 18:0), desaturases and elongases for introducing various double bonds and for elongating fatty acids with various chain lengths. These enzymes can be localized at very different locations in a cell and use different forms of lipids as substrates. For instance, a type II fatty acid synthase can be a plastidic enzyme or an extraplastidic enzyme depending on the microbial species. A desaturase can be associated or not associated with the membrane that can be of endoplasmic reticulum or plastid envelope using acyl-ACP, acyl-CoA or glycerolipids as substrates, while an elongase is usually localized on endoplasmic reticulum using acyl-CoAs as substrates. Due to the complexity in localization and substrate use, VLCPUFA-producers using the aerobic pathway for the biosynthesis usually accumulate a substantial amount of intermediate fatty acids in addition to the main final products, while VLCPUFA-producers using a PUFA synthase for the biosynthesis often produce cleaner final products with fewer and lesser intermediate fatty acids (Ujihara *et al*., 2014), which is considered more desirable for dietary supplementation (Vrinten *et al*., 2013). In addition, since the anaerobic pathway is a *de novo* system using a single enzyme, it might be more efficient for VLCPUFAs production. However, so far little is known about the molecular mechanism of a PUFA synthase for the biosynthesis of VLCPUFAs. The PUFA synthase is a mega-enzyme encoded by large ORFs with usually high G/C content and repetitive ACP coding sequences. Cloning and expression of

the large DNA fragments in heterologous systems for functional characterization is a daunting task. In addition, the VLCPUFA producers using the PUFA system is usually not amenable to genetic manipulation. These make the mechanistic interrogation very challenging. This study attempted a dissection of the entire PUFA synthase sequence from *Thraustochytrium sp.* 26185 into different catalytic domains and then expressed the embedded domains as standalone enzymes in *E. coli* for functional characterization. Successful complementation and functional expression of the KS domains in *E. coli* is the first step towards the future study through *in vitro* assays with purified domain proteins incubated with other components including appropriate radiolabeled substrates for the final elucidation of the molecular mechanism for the biosynthesis of VLCPUFAs.

3.6 Connection to next study

This study showed that the embedded KS domains of the PUFA synthase from *Thraustochytrium* could be dissected and expressed as standalone enzymes in *E.coli* where their functions could be characterized by complementation test, site-directed mutagenesis and overexpression. This opened an opportunity for functional analysis of other key domains of PUFA synthase such as DH domains in the similar way, which was described in the next study (Chapter 4).

CHAPTER 4. FUNCTIONAL ANALYSIS OF THE DH DOMAINS OF A PUFA SYNTHASE FROM *THRAUSTOCHYTRIUM* IN *E.COLI*.

4.1 Abstract

Thraustochytrium sp. 26185, a unicellular marine protist, synthesizes docosahexaenoic acid, an omega-3 VLCPUFAs, by a PUFA synthase comprising three large subunits with multiple catalytic DH domains critical for introducing double bonds at the specific position of fatty acids. To investigate functions of these DH domains, one DH domain from subunit-A and two DH domains from subunit-C of the PUFA synthase were dissected and expressed as stand-alone enzymes in *Escherichia coli*. The results showed that all these DH domains could complement the defective phenotype of an *E. coli FabA* temperature sensitive mutant, despite they have only modest sequence similarity with FabA, indicating they can function as 3-hydroxyacyl-ACP dehydratase for the biosynthesis of unsaturated fatty acids in *E. coli*. Site-directed mutagenesis analysis confirmed the authenticity of active site residues in these domains. In addition, overexpression of the three domains in a wild type *E. coli* strain resulted in the substantial alteration of fatty acid profiles including productions and ratio of unsaturated to saturated fatty acids. A combination of evidence from sequence comparison, functional expression and mutagenesis analysis suggest that the DH domain from subunit-A is similar to DH domains from polyketide synthases, while the DH domains from subunit-C are more comparable to *E. coli* FabA in catalytic functions. Successful complementation and functional expression of the embedded DH domains from the PUFA synthase in *E. coli* is an important step towards for elucidating the molecular mechanism in the biosynthesis of VLCPUFAs in *Thraustochytrium*.

4.2 Introduction

Unsaturated fatty acids in phospholipids play crucial roles in maintaining the functional properties of cell membranes (Parsons & Rock, 2013). The living organisms have developed two distinct approaches to synthesize unsaturated fatty acids. The most popular one is to utilize a desaturase, a special type of oxygenases that can abstract two hydrogens from a saturated hydrocarbon chain, whereby introducing a double bond. This process requires molecular oxygen as the hydrogen acceptor, is thus called the aerobic way of forming a carbon-carbon double bond (Meesapyodsuk

& Qiu, 2012; Zhang *et al*., 2007). The other one is to use a fatty acid synthase (complex) to introduce double bonds during the process of fatty acid synthesis. This approach, so called the anaerobic way, does not require molecular oxygen for double bond formation (Lu *et al*. 2004; Meesapyodsuk & Qiu, 2016). The aerobic way of double bond formation mainly occurs in animals, plants and high eukaryotic microorganisms, while the anaerobic way takes place only in bacteria, protist, microalgae and primitive fungi.

Omega-3 VLCPUFAs, particularly EPA and DHA, have been shown to be essential for appropriate visual and neurological developments (Diau *et al*., 2005; Hadders-Algra, 2010), and play important roles in the prevention and reduction of chronic diseases such as arteriosclerosis, hypertension, arthritis, diabetes and neurological disorders (Horrocks & Yeo, 1999). The biosynthesis of VLCPUFAs in living species goes through either an alternating desaturation and elongation process catalyzed by various desaturases and elongases or a fatty acid synthesis process catalyzed by a polyunsaturated fatty acid synthase, a PKS-like mega-enzyme with multiple catalytic domains such as KS, KR, DH, ER, AT, MAT and ACP. In the former pathway, multiple double bonds are introduced by individual desaturases each with distinct regioselectivity, while in the latter pathway, multiple double bonds are introduced presumably through dehydratase activity conferred by the embedded DH domains in a PUFA synthase.

In a type II fatty acid synthesis system, dehydratase such as FabA in *Escherichia coli*, is a discrete enzyme catalyzing the removal of a water molecule from 3-hydroxyacyl-ACP giving 2-*trans*-enoyl-ACP, and isomerization of the intermediate to 3-*cis*-enoyl-ACP with substrate preference towards a ten-carbon acyl chain (Finzel *et al*., 2015), whereby introducing a *cis* double bond in the middle of the final long chain fatty acid. The DH domains of a PUFA synthase share sequence and structure similarity to the dehydratase, thus presumably they use a similar mechanism to introduce double bonds (Metz *et al*., 2001; Qiu, 2003). However, this assumption has never been confirmed. Moreover, how DH domains in a PUFA synthase for intricately positioning multiple double bonds in VLCPUFAs is completely unknown.

Thraustochytrium sp. 26185, a single cell protist, can produce a high level of DHA and DPA (22:5n-6). Our recent *in vitro* and *in vivo* experiments showed that a PUFA synthase is primarily responsible for the biosynthesis of these VLCPUFAs in this species (Meesapyodsuk and Qiu, 2016). Similar to orthologues from other marine microbes (Metz *et al*. 2001; Okuyama *et al*.

51

2007; Gong *et al*. 2014), the PUFA synthase from *Thraustochytrium* comprises multiple subunits with three DH catalytic domains. In this study, the three DH domains from the PUFA synthase were dissected with the aid of multiple alignments and homology modeling, and then expressed as stand-alone enzymes for functional analysis through complementation test, site-directed mutagenesis, and overexpression in *E. coli*. The results from these studies provide insights into the functions of these DH domains of the PUFA synthase in *Thraustochytrium*.

4.3 Experimental approach

4.3.1 Experimental flow

The experimental approach for this study shown as follows (Figure 4.1).
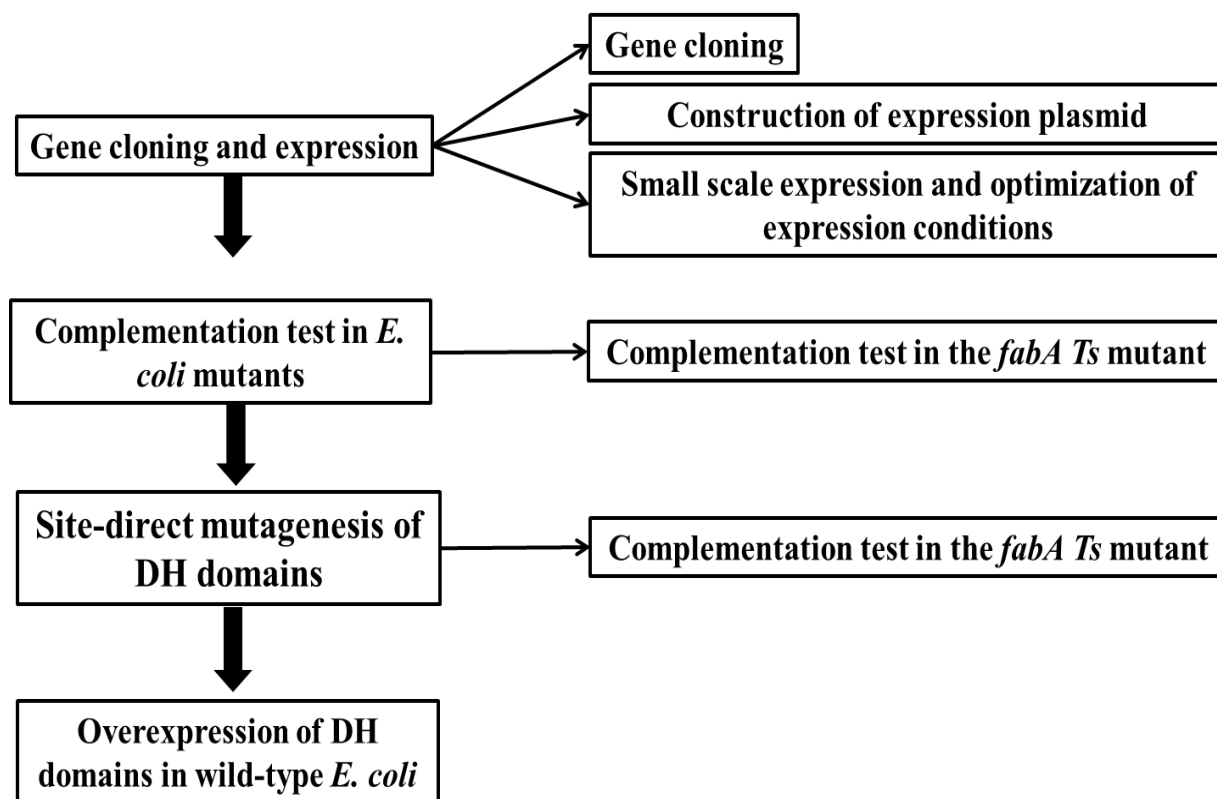
Figure 4.1 Experimental flow of chapter 4.

4.3.2 Strains, Plasmid, media, and reagents

Strains, Plasmid, media

The strain, *Thraustochytrium sp. 26185*, was purchased from the American Type Culture Collection (ATCC 26185). *E.* coli strains *Top*10 and *BL*21 (DE3) were purchased from Invitrogen Biotechnology Co. (Grand Island, USA) grown in Luria-Bertani medium (1% peptone, 0.5% yeast extract, and 1% sodium chloride). The *E.coli FabA* gene was provided by Dr. Charles Rock, Department of Biochemistry, St. Jude Children's Research Hospital, Tennessee, USA (Heath & Rock, 1996). The temperature sensitive (*Ts*) *FabA* mutant (IH131) was ordered from National BioResource Project (NBRP *E. coli*, Microbial Genetics Laboratory, National Institute of Genetics, Japan) (Yamazaki & Sugawara, 2009). Intermediated plasmid pGEM-T was purchased from Promega Co. (Madison, WI, USA). Expression vectors, pBAD and pET28a were purchased from Invitrogen Co. (Carlsbad, CA, USA).

Reagents

The DNA purification and plasmid extraction kits were obtained from Bio Basic Inc. (York, Ontario, Canada). HP Taq DNA polymerase was purchased from Bio Basic Inc. Q5 polymerase, restriction enzymes and dNTP were purchased from New England Biolabs (Ipswich, MA, USA). Primers were synthesized by Sigma-Aldrich (St. Louis, MO, USA).

4.3.3 Sequence analysis of the PUFA synthase in *Thraustochytrium*

Sequence analysis of DH domains from the PUFA synthase in *Thraustochytrium sp. 26185* was performed on the genome sequence (GenBank accession no. PRJNA368972). Three open reading frame (ORF-A, ORF-B and OFR-C) sequences were entered as query sequence to align with online databases. Tools for cataloging FASs/PKSs including BLAST (NCBI), ClusterMine360 (Conway & Boddy, 2013), antiSMASH (Medema *et al*., 2011) and NRPS-PKS (Ansari *et al*., 2004) were used to identify putative domains and active sites. Borderlines of each domain were defined through multi-sequence alignment of homologous sequences from *Schizochytrium*, *Emiliania huxleyi* and *Shewanella fidelis* by DNAMAN and confirmed by on-line homology modeling tools such as the SWISS-MODEL (Biasini *et al*., 2014) and Phyre server (Kelley & Sternberg, 2009).

4.3.4 Cloning and expression of DH domains from the PUFA synthase in *E.coli*.

To clone and express the DH domains in *E. coli*, eight primers were designed for amplifying them from ORF-A (DH-A) and ORF-C (named DH1-C and DH2-C, respectively) with a stop codon at the C-terminus. The DH domain from ORF-A was amplified with primers F DH-A and R DH-A. The DH1-C domain from ORF-C was amplified with primers F DH1-C and R DH1-C while DH2-C domain was targeted by primers F DH2-C and R DH2-C. And the tandem DH domain (DH1-DH2-C) was amplified with the primers F DH1-C and R DH2-C. FabA gene from *E.coli* was amplified with primers F-fabA and R-fabA (Table 4.1). Amplicons with the expected size were first cloned into an intermediate pGEM-T vector for sequencing confirmation.

Table 4.1 Primers used for DH domain amplifications

Name	Sequence (5'-3')	Annotation
F DH-A	<u>GAATTC</u>ATGGTGACCCCGCAGCT	*Eco*R I site (underline)
R DH-A	<u>AAGCTT</u>CTAGAAGGCAAGGCTGTCGG	*Hin*d III site (underline)
F DH1-C	<u>GAATTC</u>ATGGCGCTCCGCGTCA	*Eco*R I site (underline)
R DH1-C	<u>AAGCTT</u>CTACTCCTCGGTGCTGGTGG	*Hin*d III site (underline)
F-DH2-C	<u>GAATTC</u>ATGTCCCGCCTCGGCGA	*Eco*R I site (underline)
R-DH2-C	<u>AAGCTT</u>CTACGAGGCGGCGGGCTC	*Hin*d III site (underline)
F-fabA	<u>GAATTC</u>ATTCCGGGGATCCGTCG	*Eco*R I site (underline)
R-fabA	<u>AAGCTT</u>GAAGCAGCTCCAGCCTAC	*Hin*d III site (underline)
F-pBAD	TTGCTATGCCATAGCATTTTTATCC	For colonies PCR

For expression plasmid construction, the DH and *FabA* sequences from the pGEM-T vector were sub-cloned into a His-tagged vector pBAD for complementation test, meanwhile these DNA sequences were introduced into vector pET28a(+) for overexpression assay (Figure 4.S2). The purified *DH* and *FabA* DNA fragments from pGEM-T and expression vector digested with *Eco*RI and *Hin*dIII were ligated according to the standard procedure (Figure 4.S3). The recombinant expression vectors were transformed into competent *E. coli* Top10 cells for propagation of the recombinant plasmids. The plasmids were confirmed by restriction enzyme digestion and sequencing.

4.3.5 Complementation of *E. coli* mutants defective in dehydratase-isomerase activity with the DH domains

For expression plasmid construction, the DH and *FabA* sequences from the pGEM-T vector were sub-cloned into a His-tagged vector pBAD for complementation test, meanwhile these DNA sequences were introduced into vector pET28a(+) for overexpression assay. The purified *DH* and *FabA* DNA fragments from pGEM-T and expression vector digested with *Eco*RI and *Hin*dIII were ligated according to the standard procedure. The recombinant expression vectors were transformed into competent *E. coli* Top10 cells for propagation of the recombinant plasmids. The plasmids were confirmed by restriction enzyme digestion and sequencing.

4.3.6 Fatty acid analysis

The fatty acid profiles of the cell cultures were analyzed through GC analysis. After inducible expression, the cells were harvested using 6,000 rpm centrifugation for 10 min, 5 volumes of PBS buffer (20 mM PBS pH7.0) were added to wash the cells for three times. Equal biomass of cells (15 mL) was used for fatty acid analysis with an internal standard (heptadecanoic acid, 17:0). Total fatty acids in *E. coli* samples were converted to fatty acid methyl ester (FAMEs) by adding 2 mL of 1% sulfuric methanol and incubated at 80°C for 1 hour. After transmethylation, the sample was added 1 mL of 0.9% NaCl and 2 mL of hexane and then mixed by vortex and centrifuged at 2,200 rpm for 5 minutes for phase separation. The upper hexane phase was transferred to a new glass tube and dried with N_2 gas. After drying, the sample was resolubilized in hexane and used for GC analysis by using an Agilent 7890A system installed with a DB-23 column.

4.3.7 Site-directed mutagenesis of the DH domains

   Site-directed mutagenesis was conducted to generate DH domains with mutative active site through a PCR-based strategy (Ho *et al*., 1989). The wild-type DH domain sequences cloned into the expression vector pET28a were used as the DNA template for site-directed mutagenesis with the mutagenic and flanking primers listed in Table 4.2. First, two overlapping sequences with mutagenized site and complementary region were amplified separately by PCR using mutagenic and flanking primers respectively. Second, two overlapping fragments were fused through overlapping PCR with flanking primers. The resulting mutagenized fragment was digested with restriction enzyme and sub-cloned into pBAD and pET28a vectors. The recombinant plasmids were confirmed by restriction digestion and DNA sequencing. The confirmed plasmids were transformed into *E. coli* strains for functional analysis.

Table 4.2 Primer used for site-directed mutagenesis of DH domains

Name	Sequence (5'-3')	Annotation
F DH-M-A	CGCCCAC<u>GCA</u>CCCTTCGTGAACGAC	mutation site (underline, Asp to Ala)
R DH-M-A	CGAAGGG<u>TGC</u>GTGGGCGTCCGTGT	mutation site (underline, Asp to Ala)
F DH1-M-C	GTCTCG<u>GCA</u>GGCTGCAGCCAGAT	mutation site (underline, Asp to Ala)
R DH1-M-C	GCAGCC<u>TGC</u>CGAGACGAGCGAGC	mutation site (underline, Asp to Ala)
F DH2-M-C	GGCGTC<u>GCA</u>TCCATGTTCCAGCTC	mutation site (underline, Glu to Ala)
R DH2-M-C	CATGGA<u>TGC</u>GACGCCGAGGCTG	mutation site (underline, Glu to Ala)

4.3.8 Overexpression of the DH domains in wild-type *E. coli*

Overexpression was performed in wild-type strain *E.coli BL*21(DE3). The expression plasmids and control vector pET28a were transformed into the competent cells. The transformants were then spread on the plates containing 50 µg/mL kanamycin. A single colony of the positive transformant harboring the expression plasmid was grown overnight in 10 mL of LB medium with 50 µg/mL kanamycin at 37 °C. A 500 µL aliquot of the overnight culture was inoculated into 100 mL of fresh LB medium containing kanamycin and incubated aerobically at 37 °C until an A_{600} of 0.4 reached. After that, the cell culture was aliquoted into three equal volumes and incubated at 20°C for inducible expression. Fatty acids profiles of *E. coli BL*21 (DE3) transformants expressing different DH domains were analyzed by GC as described above.

4.4 Results

4.4.1 Sequence analysis of DH domains of a PUFA synthase in *Thraustochytrium*

To investigate the biosynthetic mechanism of VLCPUFAs in *Thraustochytrium* sp. 26185, the whole genome of this protist was recently sequenced (Zhao *et al*., 2016). From the genome sequence, A PUFA synthase solely responsible for the biosynthesis of DHA in the species was identified, and it was comprised of three subunits with three DH catalytic domains (Meesapyodsuk & Qiu, 2016) (Figure 4.2A). One DH domain was located at the C-terminus of subunit-A (DH-A) sharing approximately 24% sequence identity with the DH domain from a polyketide synthase (Rouhiainen *et al*., 2004), but low sequence similarity to *E. coli* FabA and FabZ (<10%). A predicted active site residue histidine was found in the N-terminal region of DH-A, while another active site residue aspartic acid was identified in the middle of the domain. Two DH domains resided adjacently at the N-terminal region of subunit-C (DH1-C and DH2-C) and shared only 15% sequence identity with each other, but 32% and 27% identity with *E. coli* FabA, and 11% and 20% with *E. coli* FabZ, respectively. By contrast, both DH1-C and DH2-C shared very low sequence identity (<10%) to the DH domain of polyketide synthase (Figure 4.3). Two predicted active residues, histidine and aspartic acid/glutamic acid, located in the middle of the domains. Noticeably, DH1-C and DH2-C domains contained a long linker sequence upstream and a FabA-like core domain. This long undefined amino acid sequence predicted a pseudo-domain sharing the similar secondary structure as the FabA-like DH domain, but without a conserved active site

histidine residue. A combination of pseudo-domain and FabA-like domain contributed a structure that might be critical for catalytic activity (Pasta *et al*., 2007). The boundaries of these DH domains of the PUFA synthase were determined by sequence alignment and homology modeling. The DH-A domain ranged from residue 3161 to 3472 in subunit-A, while two DH domains (DH1-C and DH2-C) in subunit-C were from residue 1 to 413 and 414 to 1030, respectively.
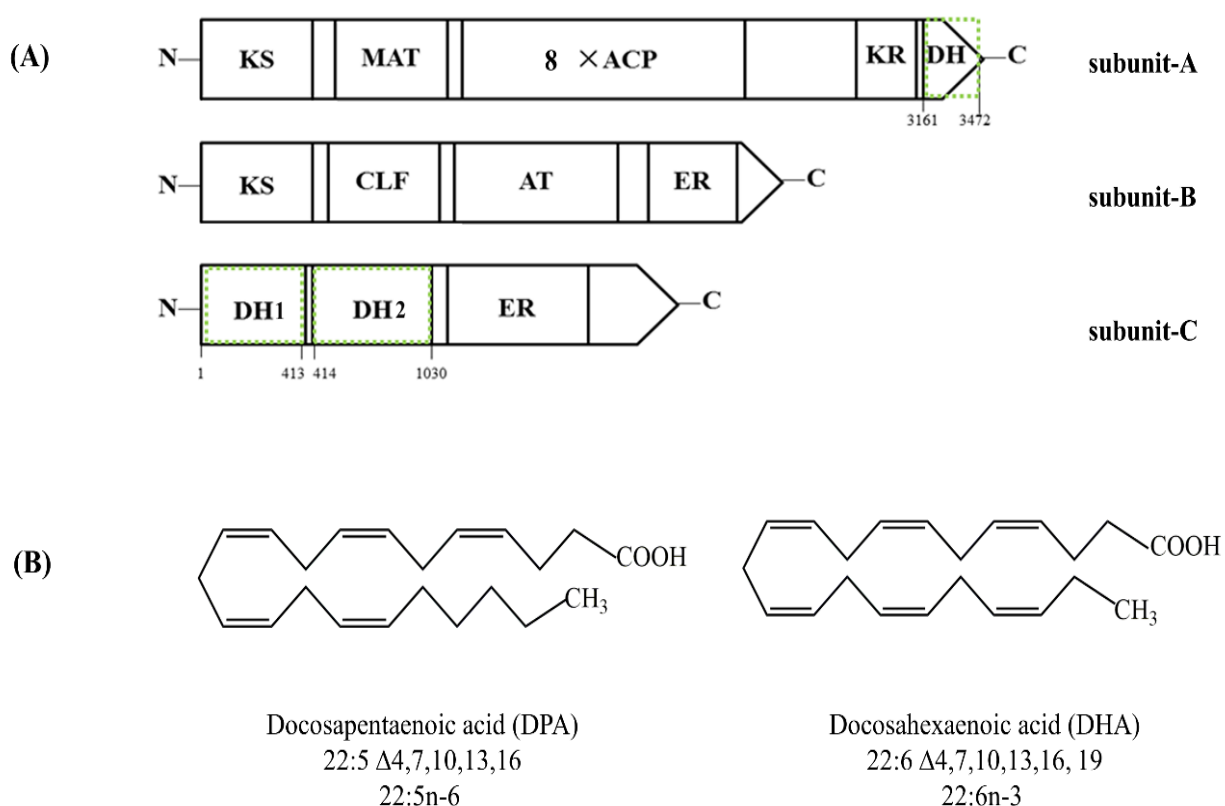
Figure 4.2 The domain organization of the PUFA synthase and primary VLCPUFAs in *Thraustochytrium sp.* 26185. (A) The domain organization of the PUFA synthase. Three DH domains were highlighted by green dotted rectangular line, the number indicating the amino acids positions. ACP, acyl carrier protein; KS, 3-ketoacyl-ACP synthase; CLF, chain length factor; KR, 3-ketoacyl-ACP reductase; DH, 3-hydroxyacyl-ACP dehydratase; ER, enoyl-ACP reductase; MAT, malonyl acyl transferase. (B) Structures of two main VLCPUFAs produced by *Thraustochytrium.*

(A)

DH-A   RKISAKSNAELDDHVIQGRRVLPMTLAIGSLAETCLGLFPGYSLWAIDDAQLFKGVTVDG 60
PKS DH RRLTLRGNPEVYDHVIGGNAVLPAMCALAWITNSCEQLYKGYHFLSCHNYQVLKGVVFDK 60

DH-A   DVN--CEIALTPSEGQAGRVNVQATLKTFSSGKLVPAYRAVVVLSSQGAPPANATMQPPS 118
PKS DH SLANLYCLDLTEVEKTEDEIKFEALIWSETAKGIPLYHRAIINITKQVIAERKLEESIQ 120

DH-A   LSADPAAQSAVYDGKTLFHGPAFRGLDEVLSCSKSQLVAKCRAVPGSDAARAEFATDTDA 178
PKS DH PVKEPSLNFQPYQEGVLFHQPRFQGLKKILEINNNELLFNCYLPPISTQDQGQFS--VQS 178

DH-A   HDPFVNDLAFQAMLVWWRRTLGQAALPNSIQRIVQHRPVPQDKPFYLTLRSNQAGGHSQH 238
PKS DH FNSYTADILFQCLLVWWRKHYNSGSLPLKLNELKQFSSLPFNQEFWLKLSINEHSDTKVI 238

DH-A   KHALQFHNEQGDLFIDVQASVIATDSLAF   267
PKS DH ANALSYHS-QGKIYLEATNMEVT   260

(B)

DH1-C   LDGGISMFFFEYDCYVNGRLLIEMRDGCAGFFTNEELAAGKGVVFTRADLAARAKVAKQD 228
DH2-C   MLGEMGVHRFSFELSVDDVVFYKGTTSFGWFVP---------EVFAAQAGLDSGRKSEPW 231
FabA   ------------------MVDKRE----------------S-YTKEDLLASGR----   18
FabZ   ------------------MTTNTHT--------------L--QIEEILELLP-----   18

DH1-C   ISPYAVAPCLHKTTLSEKEMQTLVDKDWASVFGSKNGMPEINYKLCARKMLMIDRVPKID 288
DH2-C   FVEQKVPAGQVVSYDVRPG-----ASGRTSLFANAPSG--AQLNRRTNQGQYLDKIDLVS 284
FabA   --------------GE-----------LFGAKG----P--QLPAPNMLMMDRVVKMT   44
FabZ   ---------------------------H-----RFPFLLVDRVLDFE   33

DH1-C   HTGGIYGLGLIVGEKILERDHWYFPCIFVKDEVMAGSLVSDGCSQMLKMYMIWLGLHLTT 348
DH2-C   GSGKQ-GLGYGHGVKAVNPNDWFFACHFWFDSVMPGSLGVESMFQLVEAIAVQDDLAGKH 343
FabA   ETGGNFDKGYVEAELDINPDLWFFGCHFIGDPVMPGCLGLDAMWQLVGFYLGWMGGEGKG 104
FabZ   EGR------FLRAVKNVSVNEPFFQGHFPGKPIFPGVLILEMAQATGILAFKSVGKLEP   87

DH1-C   GPFDFRPVSGH-PNKVRCRGQISPHKGKLVYVMEIKEIGFDEKNDPYALADVNIIDIDYE 407
DH2-C   GIANPTITHAPGKISWKYRGQLTPKSKQMDSEVHVTSI-----EAHDGVVDV------VA 392
FabA   RALGVG--------EVKFTGQVLPTAKKVTYRIHFKRIVN---RR--LIMGL-------A 144
FabZ   GELYYFAG----IDEARFKRPVVPGD-QMIMEVTFEKT-----RR--GLTRF------- 127

DH1-C   KGQTFEMSRLGDYGKGDLNKKLVVDFKGIALKMQKRSTSTEE   449
DH2-C   DGYLWADGLR-VYQVNNIRVRLASGEPAAS   421
FabA   DGEVLVDGRL-IYTASDLKVGLFQDTSAF   172
FabZ   KGVALVDGKV----VCEATMMCARSREA   151

Figure 4.3 Alignment of the DH domains with FabA and FabZ from *E. coli* and homologous DH domain from polyketide synthase. (A) Sequence alignment of the DH domain from subunit-A of the PUFA synthase with the DH domain from *Microcystis aeruginosa* polyketide synthase. (B) Sequence alignment of two DH domains from subunit-C of the PUFA synthase with FabA (accession no. P0A6Q3.2) and FabZ (accession no. P0A6Q6.1) from *E.coli*. DH1-C, DH1 domain from Subunit-C; DH2-C, DH2 domain from subunit-C; FabA and FabZ, dehydratases from *E.coli*. The putative active sites were highlighted by red rectangular line, and the consensus amino acids were marked by shade.

4.4.2 Heterologous complementation of *E. coli FabA* mutant with DH domains

In *E.coli*, FabA can catalyze the dehydration of 3-hydroxyacyl-ACP to into a *cis*-3-enoyl-ACP with a ten-carbon chain (Rock, 2008), which can then be used for subsequent condensation reaction, whereby a *cis* double bond is retained in an acyl chain. FabA is essential for synthesizing UFAs in *E. coli* to maintain the normal function of cell membrane (Rock *et al.*, 1996). Therefore, the complementation test of the DH domains was carried out in a *FabA* (*Ts*) mutant with a temperature sensitive phenotype, i.e. it could only grow at 30°C, but not at a non-permissive temperature (40°C). As shown in Figure 4.4, the mutant expressing three DH domains from the PUFA synthase as well as wild-type *FabA* were all capable of growing at both temperatures (30°C and 40°C), in contrast to the mutant with the empty vector that ceased growth at 40°C. This result clearly demonstrated that all DH domains from the PUFA synthase could functionally replace FabA as 3-hydroxyacyl-ACP dehydratase for the biosynthesis of unsaturated fatty acids in *E. coli*.
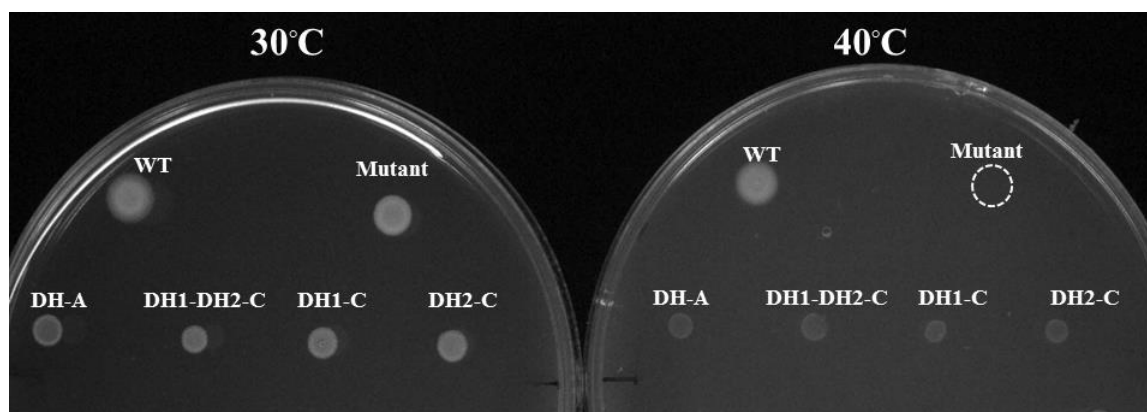
Figure 4.4 The complementation plate assay of DH domains in *E. coli FabA Ts* mutant. DH domains were inserted behind the inducing promoter P$_{BAD}$ individually. The recombinant plasmids were transformed into the *FabA Ts* mutant. The transformants grown on selection plates with 0.02% L-arabinose were incubated overnight at 30°C and 40°C, respectively. WT, *E. coli* wild-type W3110; Mutant, *E. coli FabA Ts* mutant with empty vector; DH-A, the mutant expressing the DH domain from subunit-A; DH1-C, the mutant expressing the DH1 domain from subunit-C; DH2-C, the mutant expressing the DH2 domain from subunit-C; DH1-DH2-C, the mutant expressing both DH1 and DH2 from subunit-C.

To obtain further insight into fatty acid profiles in the complementary transformants, six major fatty acids were measured in these strains grown at the permissive temperature, including myristic acid (14:0), palmitoleic acid (16:1-9), palmitic acid (16:0), 9,10-methylene-hexadecanoic acid (17:0-cyclopropane fatty acid, 17:0-cyc), *cis*-vaccenic acid (18:1-11) and 11,12-methylene-octadecanoic acid (19:0-cyclopropane fatty acid, 19:0-cyc). As shown in Figure 4.5, compared to the wild-type strain, the production of these fatty acids was severely compromised in the *FabA* mutant, particularly for unsaturated fatty acids 16:1-9 and 17:0-cyc, which led to the significant reduction of both unsaturated fatty acids and total fatty acids in the mutant strain. On the other hand, expression of all DH domains from the PUFA synthase in the mutant reversed the defective phenotype. Specifically, when DH-A was expressed in the mutant, the amounts of fatty acids such as 16:0, 16:1-9, 17:0-cyc, 18:1-11 were significantly higher than the mutant with empty vector. When DH1-C and DH2-C expressions were compared, the amounts of 16:0 and 18:1, as well as the total fatty acids produced in DH1-C complementary strain, were slightly higher than those in DH2-C complementary strain. The mutant expressing both domains (DH1-DH2-C) produced the highest amounts of 16:0, 17:0-cyc and 18:1-11 as well as total fatty acids than those expressing three individual DH domains and the wild-type control. However, the ratio of UFAs to SFAs was the highest in the mutant expressing DH-A, and lowest in the mutant expressing two tandem DH domains (DH1-DH2-C) among these complementary strains (Table 4.3).
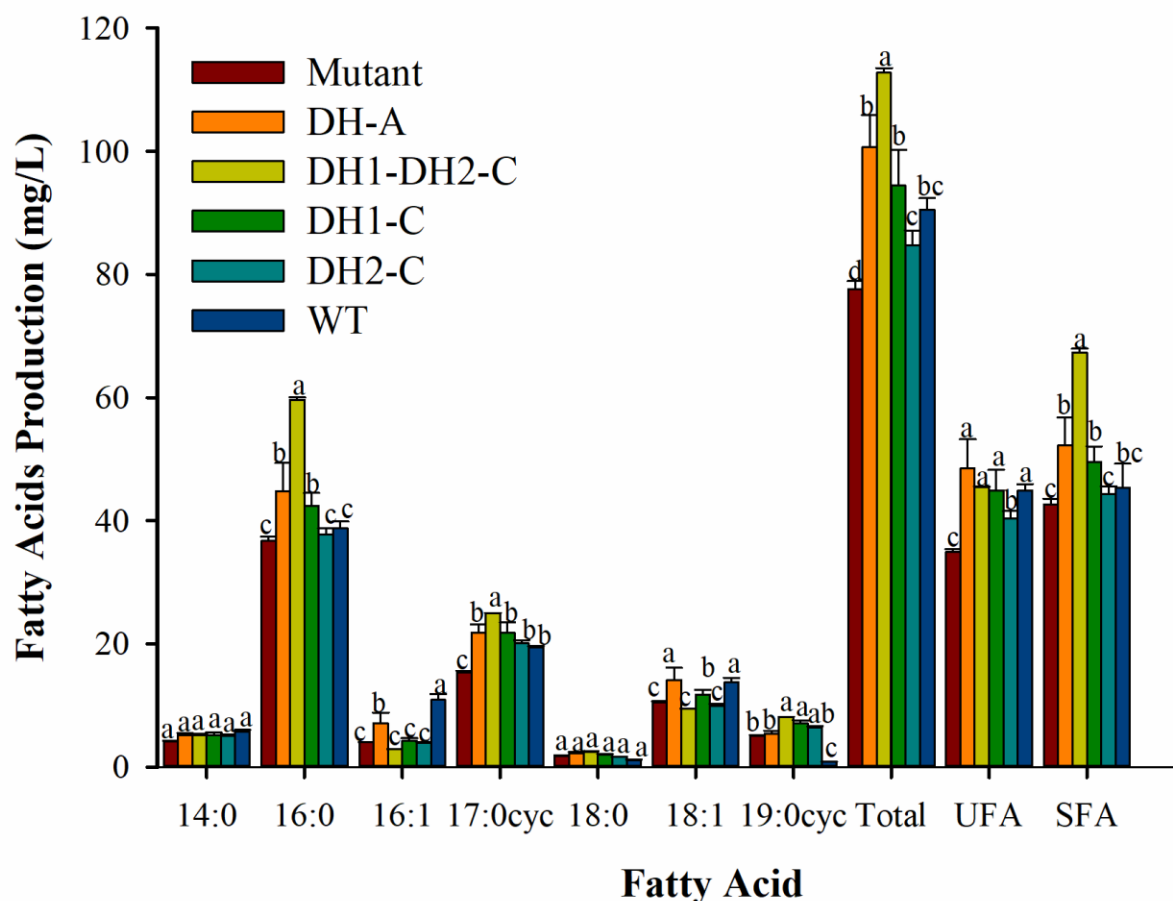
Figure 4.5 The amount of fatty acids produced in the *E. coli FabA* mutant expressing the DH domains from the PUFA synthase at 30°C. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12-methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

Table 4.3 Production of saturated and unsaturated fatty acids in *E. coli FabA* mutant expressing the DH domains[a]. EV, mutant with the empty vector control; FabA, mutant expressing *FabA*; DH-A, mutant expressing DH domain from subunit-A; DH1-DH2-C, mutant expressing DH1-DH2 domain from subunit-C; DH1-C, mutant expressing DH1 domain from subunit-C; DH2-C, mutant expressing DH2 domain from subunit-C.

Strain	The amount of fatty acids produced (mg/L)[b]			UFA/SFA ratio
	Total	UFA	SFA	
EV	77.60 ± 1.33^D	34.94 ± 1.07^C	42.66 ± 1.16^C	0.82
DH-A	100.65 ± 4.98^B	48.49 ± 4.71^A	52.17 ± 4.35^B	0.93
DH1-DH2-C	112.74 ± 1.27^A	45.42 ± 0.59^A	67.32 ± 0.65^A	0.67
DH1-C	94.45 ± 4.72^B	44.92 ± 3.31^A	49.53 ± 2.51^B	0.91
DH2-C	84.71 ± 2.44^C	40.36 ± 1.18^B	44.35 ± 1.25^C	0.91
WT	90.54 ± 1.88^{BC}	44.88 ± 1.03^A	45.36 ± 3.38^{BC}	0.98

[a]Production of saturated (SFA) and unsaturated fatty acids (UFA) in the expression of DH domains in *E. coli FabA* mutant; [b]Values were reported as the means ± standard deviations for three biological replicates. The means with the same letters are not statistically significantly different, while the means with different letters are statistically significantly different according to the one-way analysis of variance ($P < 0.05$).

4.4.3 Site-directed mutagenesis analysis of the DH domains

The DH domains in the PUFA synthase was predicted according to the conserved characteristic residues at their putative active sites. To confirm authenticity of these residues for the function of the DH domains, the aspartic acid or glutamic acid residue at the predicted active sites of the DH domains was replaced by an alanine residue using site-directed mutagenesis. Individual mutated domains were then expressed as stand-alone enzymes for the complementation analysis. The result showed that, like the mutant with empty vector, all the transformants

expressing single mutated DH domain could grow only at 30°C, but not at the non-permissive temperature 40°C. In contrast, the mutant with wild type DH domain that could grow at both temperatures (Figure 4.6). This result reaffirmed that the putative active residue is essential for the catalytic function of these DH domains. In addition, these mutant domains were also over-expressed in a wild type *E. coli* strain. Intriguingly, expression of individual mutated DH domains from subunit-C (DH1-M-C or DH2-M-C), but not mutated DH domain from subunit A (DH-M-A), resulted in significantly reduced fatty acid production relative to the wild type control. Expression of two tandem DH domains from subunit-C with either being mutated revealed slightly stronger effect of the mutated DH1-C (DH1-M-DH2-C) over the mutated DH2-C (DH1-DH2-M-C) (Figure 4.S1).
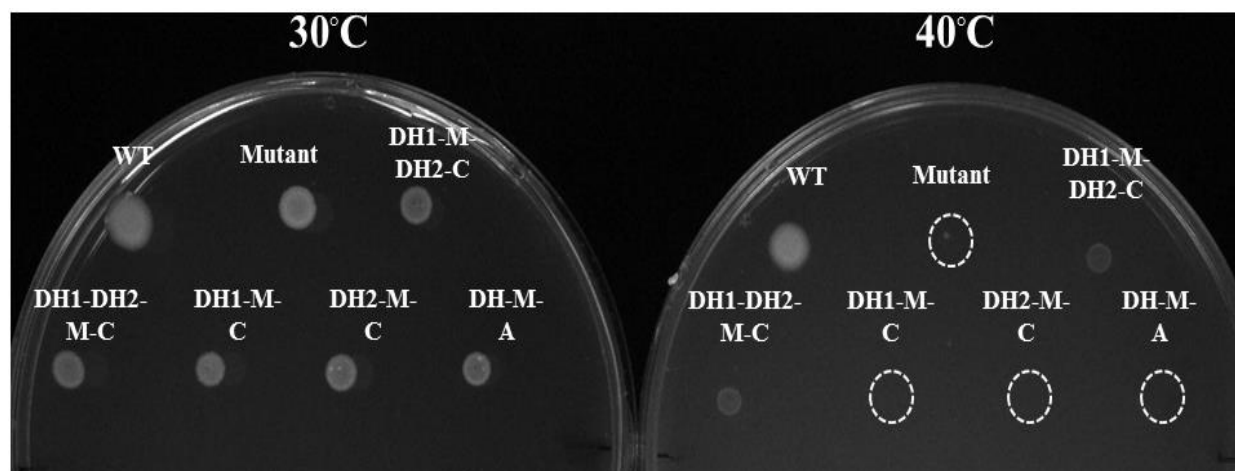
Figure 4.6 The complementation plate assay of DH domains in *E. coli FabA Ts* mutant as conducted in Figure 4.4. DH-M-A, the mutant expressing mutated DH domain from subunit-A; DH1-M-C, the mutant expressing mutated DH1 domain from subunit-C; DH2-M-C, the mutant expressing mutated DH2 domain from subunit-C. DH1-M-DH2-C, the mutant expressing mutated DH1 along with DH2 from subunit-C; DH1-DH2-M -C, the mutant expressing mutated DH2 along with DH1 from subunit-C.

4.4.4 Overexpression of the DH domains in *E. coli*

To further study functional roles of the DH domains, these domains were over-expressed under a strong promoter in a wild type *E. coli* strain. As shown in Figure 4.7, all the transformants expressing the DH domains produced significantly higher levels of the total fatty acids, compared with the control and FabA expression. In particular, overexpression of DH-A enhanced the UFA level most significantly with the amount of UFAs including 16:1-9, 17:0-cyc and 18:1-11 being increased by 38.4%. Overexpression of DH1-C produced a slightly higher amount of the total fatty acids than overexpression of DH2-C in the wild type strain. Overexpression of two tandem DH domains improved the fatty acid production most significantly among the overexpressions. The overall production of fatty acids was at 119 mg/L which was increased by approximately 50% over the control. When the relative levels of UFAs and SFAs produced in these strains were compared, DH-A overexpression produced a ratio of UFAs/SFAs higher than that in the control, while the ratios of UFAs/SFAs in the overexpression of DH1-C, DH2-C and DH1-DH2-C, as in the overexpression of FabA, were all lower than the control (Table 4.4). The lower ratios of UFAs/SFAs in these overexpressions were mainly attributed to a higher production of overall fatty acids, particularly SFAs. This result indicated that overexpression of all the DH domains could significantly improve the overall production of the total fatty acids, however, the DH domain from subunit-A was able to produce UFAs more efficiently and the DH domain from subunit-C could improve the production of the total fatty acids especially SFAs more efficiently in *E. coli*.
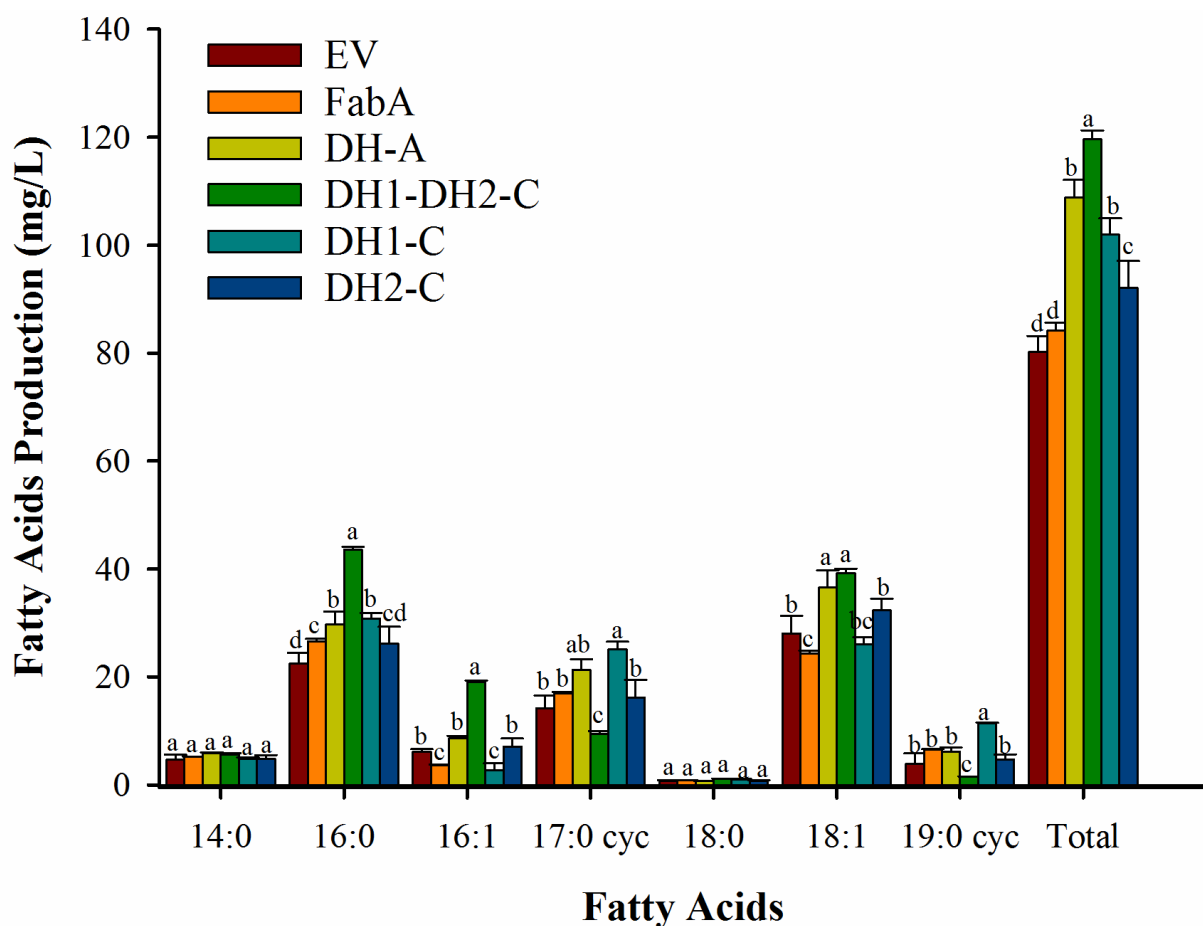
Figure 4.7 The fatty acids production of DH over-expressions in wild-type *E. coli*. EV, *E. coli BL*21 with empty vector; FabA, *E. coli BL*21 overexpressing FabA; DH-A, *E. coli BL*21 overexpressing the DH domain from subunit-A; DH1-DH2-C, *E. coli BL*21 overexpressing DH1 and DH2 domains from subunit-C; DH1-C, *E. coli BL*21 overexpressing the DH1 domain from subunit-C; DH2-C, *E. coli BL*21 overexpressing the DH2 domain from subunit-C. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Fatty acid abbreviations are the following: myristic acid (14:0), palmitic acid (16:0), palmitoleic acid (16:1-9), 9,10-methylene-hexadecanoic acid (17:0-cyc), octadecanoic acid (18:0), cis-vaccenic acid (18:1-11) and 11,12-methylene-octadecanoic acid (19:0-cyc). Statistical analysis of the results was conducted using the oneway analysis of variance (P<0.05).

Table 4.4 Production of saturated and unsaturated fatty acids in wild-type *E. coli BL*21 overexpressing the DH domains[a]. EV, the empty vector control; FabA, *E.coli BL*21overexpressing *FabA*; DH-A, *E.coli* BL21 overexpressing DH domain from subunit-A; DH1-DH2-C, *E.coli BL*21 overexpressing DH1-DH2 domain from subunit-C; DH1-C, *E.coli* BL21 overexpressing DH1 domain from subunit-C; DH2-C, *E.coli BL*21 overexpressing DH2 domain from subunit-C.

| Strain | The amount of fatty acids produced (mg/L)[b] | | | UFA/SFA ratio |
	Total	UFA	SFA	
EV	80.19 ± 3.02^D	52.23 ± 1.81^C	27.96 ± 2.08^D	1.87
FabA	84.18 ± 1.95^D	51.48 ± 1.04^C	32.69 ± 0.46^C	1.57
DH-A	108.85 ± 3.22^B	72.60 ± 2.47^A	36.26 ± 2.68^B	2.00
DH1-DH2-C	119.58 ± 1.71^A	69.35 ± 1.59^A	50.23 ± 1.02^A	1.38
DH1-C	101.97 ± 3.43^B	65.30 ± 1.31^A	36.67 ± 1.09^B	1.78
DH2-C	93.12 ± 5.64^C	60.42 ± 2.33^B	32.71 ± 1.36^C	1.83

[a]Production of saturated (SFA) and unsaturated fatty acids (UFA) in the overexpression of DH domains in wild-type *E. coli*.

[b]Values were reported as the means $\pm$ standard deviations for three biological replicates. The means with the same letters are not statistically significantly different, while the means with different letters are statistically significantly different according to the one-way analysis of variance ($P < 0.05$).

4.5 Discussion

PUFA synthases have been identified in many marine microorganisms such as bacteria, protists, fungi and algae where that are primarily responsible for the biosynthesis of VLCPUFAs such as EPA, DPA and DHA (Okuyama *et al*., 2007; Gong *et al*., 2014; Meesapyodsuk & Qiu, 2016). However, the exact molecular mechanism of a PUFA synthase for intricately positioning multiple double bonds in VLCPUFAs remains unknown. The PUFA synthase in *Thraustochytrium sp*. 26185 comprises three large subunits, each with multiple domains predicted by the presence of characteristic active site motifs. It has been assumed that DH and KS domains of a PUFA synthase play key roles in the biosynthesis of VLCPUFAs where DH domains catalyze introduction of double bonds while KS domains are responsible for retaining the double bonds (Metz *et al*., 2001; Qiu, 2003; Xie *et al*., 2017). However, this assumption has never been proved.

E. coli is able to synthesize both SFAs and UFAs through a Type II fatty acid synthesis system using two different types of dehydratases. FabZ catalyzes dehydration of 3-hydroxyacyl-ACPs with a wide range of chain length to corresponding *trans*-2-enoyl-ACPs, which are then reduced by enoyl-ACP reductase to saturated acyl-ACPs. FabA is a bifunctional enzyme that catalyzes dehydration of 3-hydroxyacyl-ACP of ten-carbon chain length to *trans*-2-enoyl-ACP and subsequently isomerization of *trans*-2-enoyl-ACP to *cis*-3-enoyl-ACP (Feng & Cronan, 2009), which is then used for subsequent condensation reaction whereby the double bond is retained. FabA works as a dimer where each monomer comprises a central α-helix wrapped by two six-stranded antiparallel β-sheets (Leesong *et al*., 1996) and two active sites are formed at the interface with His and Asp/Glu active residues contributed by two monomers. This unique structure is important for the dehydratase activity in producing UFAs (Cronan & Thomas, 2009). The FabA *Ts* mutant has low dehydratase activity and is UFA-auxotrophy at the non-permissive temperature where the mutated FabA is unable to form the appropriate dimer structure for synthesizing UFAs (Rock *et al*., 1996). DH-A of the PUFA synthase from *Thraustochytrium* has higher sequence similarity to DH domains of polyketide synthases and very low sequence similarity to *E. coli* FabA and FabZ, while DH1-C and DH2-C have relatively higher sequence similarity to *E. coli* dehydratases, particularly to FabA, but low similarity to DH domains of polyketide synthases. However, expressions of all three DH domains of the PUFA synthase can restore the defective phenotype, indicating all these domains can functionally replace FabA for the biosynthesis of unsaturated fatty acids in *E. coli*.

As DH domains from the PUFA synthase can functionally complement *FabA* mutant, we then overexpressed the DH domains in a wild type *E.coli* strain to interrogate if they have any implication in fatty acid production. It is noticed that the most drastic implication occurs when the two tandem DH domain from subunit-C was overexpressed where the amount of total fatty acids was increased by 50% over the wild type control. Similar to overexpression of FabA, overexpressions of either tandem or individual DH domains from subunit-C produce SFAs more efficiently than UFAs (Table 4.4). By contrast, overexpression of DH-A domain produced UFAs more efficiently. These results suggest that DH1-C and DH2-C, other than DH-A, are functionally (not just structurally) more similar to FabA. Site-directed mutagenesis were then performed to examine the functions of DH domains from the PUFA synthase. The aspartic acid or glutamic acid residue at the predicted active site of these domains is highly conserved in dehydratases, and DH domains of PUFA synthases and polyketide synthases (Keatinge-Clay, 2012; Lu *et al*., 2004; Rawsthorne, 2002; Smith *et al*., 2003). When this residue in the DH domains of the PUFA synthase was changed to an alanine residue, all the substitutions could not complement the defective phenotype of the *FabA* mutant, confirming the critical role of this amino acid in defining the catalytic activity in those DH domains. Intriguingly, overexpression of mutated DH domains from subunit-C (DH1-M-C and DH1-M-C), but not mutated DH domain from subunit A (DH-M-A), resulted in reduced fatty acid production relative to the wild type control (Figure 4.S1). The lower production of total fatty acids in *E.coli* transformants with mutated DH1-C or DH2-C might result from the phenomenon called "dominant negative mutation effect" (Ira 1987; Sheppard, 1994) where the competition of a non-functional enzyme (mutated DH1-M-C or DH2-M-C) with a functional enzyme (endogenous FabA) for the same substrate or dimerization of one mutated monomer (mutated DH1-M-C or DH2-M-C) and one intact monomer (endogenous FabA) occurs. Regardless, this result restates the functional and structural similarity between DH domains from subunit-C and FabA. On the basis of these results, we believe that DH-A might function as a general dehydratase mainly responsible for introducing a 2-*trans* double bond, while DH1-C and DH2-C might introduce a 3-*cis* double bond during the biosynthesis of VLCPUFAs in *Thraustochytrium*. The *trans* double bond produced by DH-A could be reduced by either of two ER domains in the PUFA synthase while the *cis* double bond produced by DH-C could be retained in VLCPUFAs.

So far, functional analysis of PUFA synthases has been limited to the expression of the mega-enzyme as a whole in a heterologous system such as *E. coli* (Amiri-Jami *et al*., 2014; Meesapyodsuk & Qiu, 2016) with very few reports tapping into the function of individual catalytic domains. The main reason for this might be the absence of suitable substrates, purified domain proteins and other cofactors for the *in vitro* assays of fatty acid synthesis. This study attempted to dissecting the embedded catalytic DH domains from the PUFA synthase in *Thraustochytrium* and expressing these domains as stand-alone enzymes in *E. coli* dehydratase mutant and wild type strains. Successful complementation and functional expression of individual DH domains is the first step towards the final elucidation of the molecular mechanism for the PUFA synthase underlying the biosynthesis of VLCPUFAs.

4.6 Supplemental Data
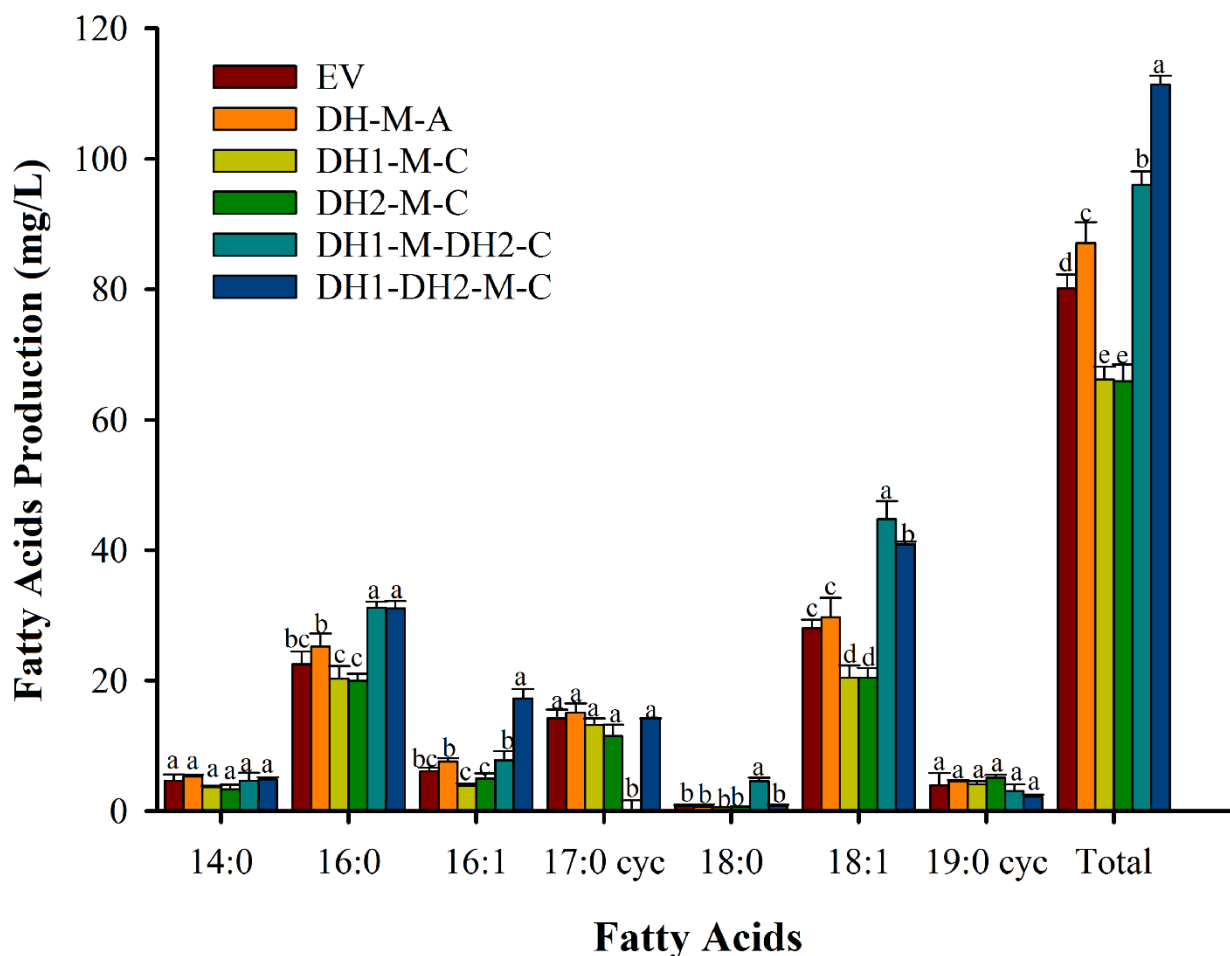
Figure 4.S1 Fatty acid production in a wild type *E. coli* strain over-expressing the mutated DH domains. EV, *E. coli* BL21 with empty vector; DH-M-A, *E. coli* BL21 overexpressing the mutated DH domain from subunit-A; DH1-M-C, *E. coli* BL21 overexpressing mutated DH1 domain from subunit-C; DH2-M-C, *E. coli* BL21 overexpressing mutated DH2 domain from subunit-C; DH1-M-DH2-C, *E. coli* BL21 overexpressing mutated DH1 along with DH2 from subunit-C; DH1-DH2-M-C, *E. coli* BL21 overexpressing mutated DH2 along with DH1 from subunit-C. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis was of the results from experiments conducted using the one-way analysis of variance ($P<0.05$).
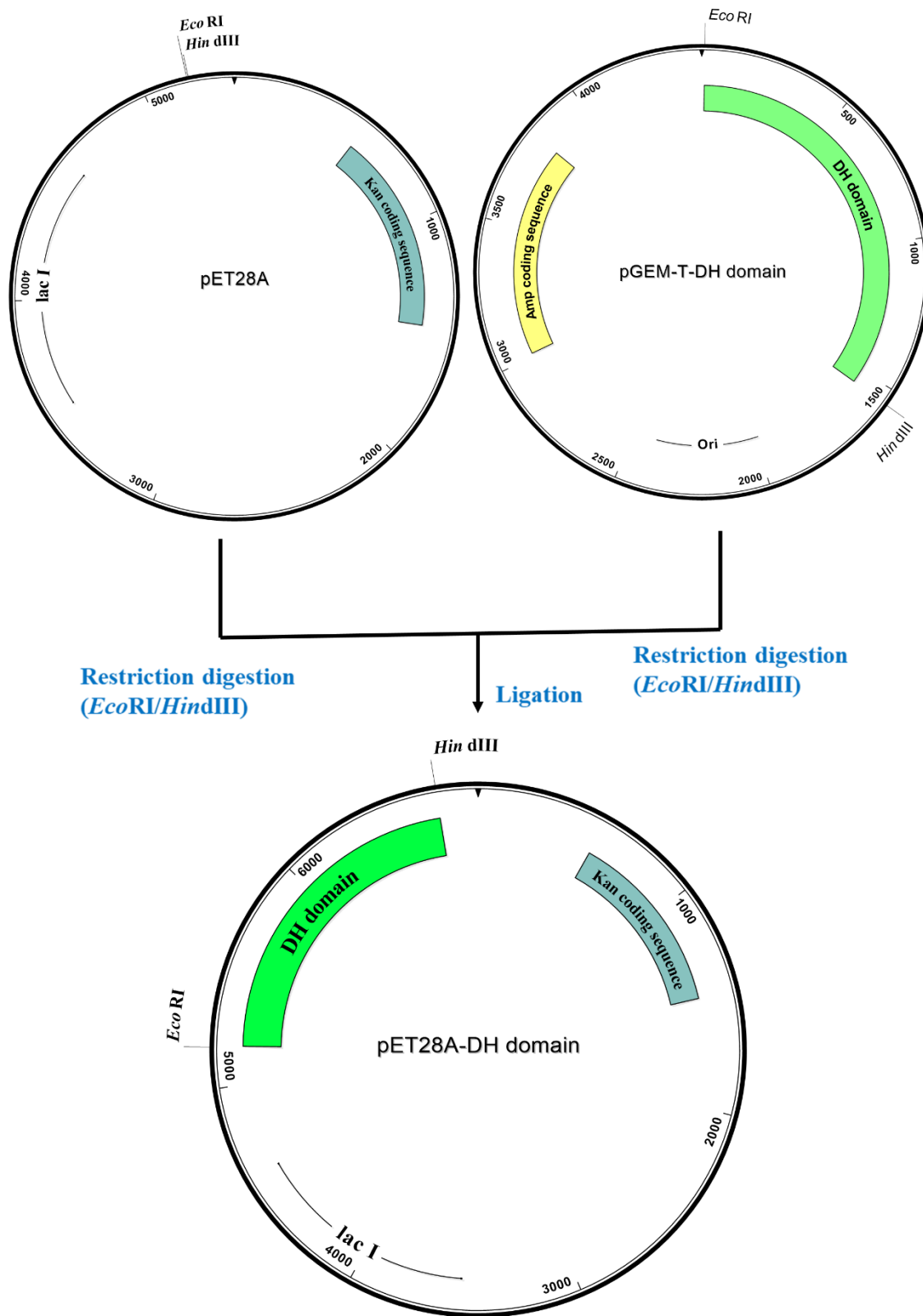
Figure 4.S2 Construction of the expression plasmids for overexpression.
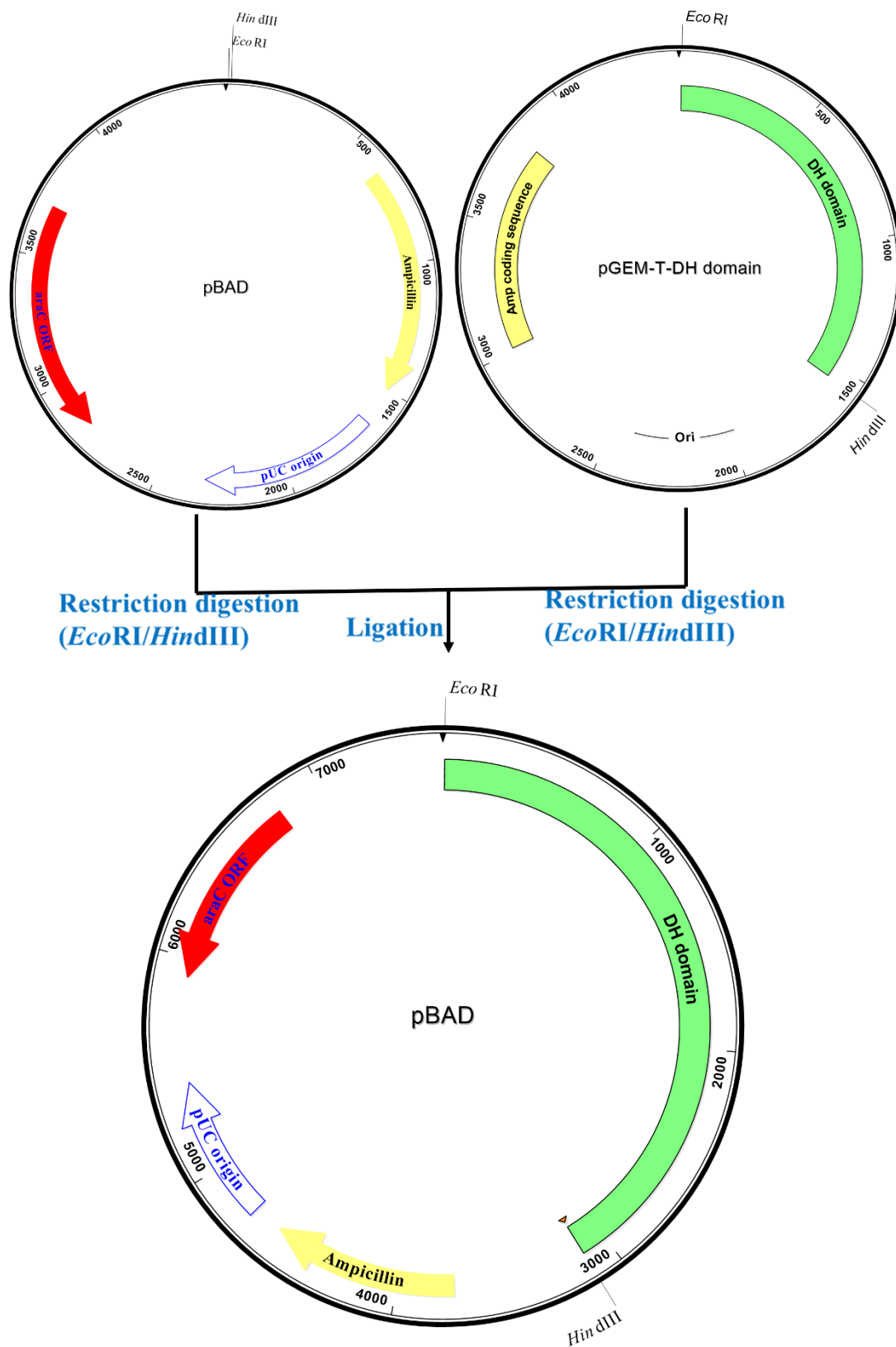
Figure 4.S3 Construction of the expression plasmids for complementation test.

4.7 Connection to next study

The study in Chapter 3 showed that a KS domain from subunit-B of PUFA synthase could complement an *E. coli* mutant defective in ketoacyl-ACP synthase I (*FabB*). To further functionally analyze this domain, it was expressed in wild type and mutant (*KASI* knockout) *Arabidopsis thaliana.* The result was described in the next study (Chapter 5).

CHAPTER 5. ENHANCING OIL PRODUCTION IN *ARABIDOPSIS* THROUGH EXPRESSION OF A KETOACYL-ACP SYNTHASE DOMAIN OF THE PUFA SYNTHASE FROM *THRAUSTOCHYTRIUM*.

5.1 Abstract

Thraustochytrium is a marine protist that can produce a high level of nutritionally important VLCPUFAs using a PUFA synthase, a PKS-like fatty acid synthase comprising three large subunits each with multiple catalytic domains. However, the exact functions of these domains were not well understood. A previous study showed that a KS domain from subunit-B could complement an *Escherichia coli* mutant defective in ketoacyl-ACP synthase I (*FabB*). In this study, this KS domain from the PUFA synthase was further functionally analyzed in *Arabidopsis thaliana*. The result showed that the plastidial expression of the domain could complement the defective phenotypes of a *kasI* knock-out mutant. Seed-specific expression of the domain significantly increased seed weight and seed oil, and altered the unsaturation level of fatty acids in seeds. In addition, overexpression of the domain also promoted seed germination and early seedling growth. These results indicate that the condensation process of fatty acid biosynthesis in plants is a limiting step and overexpression of the KS domain from a PUFA synthase of microbial origin offers a new strategy to increase oil production in commercial oilseed crops.

5.2 Introduction

Plant seed oil is an important bioresource not only for nutrition providing high energy saturated fatty acids and nutraceutical unsaturated fatty acids, but also for industrial bioproducts such as biofuel and biopolymers replacing non-renewable and environmentally unfriendly petro-polymers and petro-fuels (Kumar *et al*., 2016; Robinson & Mazurak, 2013; Wang *et al*., 2014). Therefore, increasing oil content in seeds has long been one of the primary targets in the breeding programs of oilseed crops. In the past few years, the success has been made through various molecular breeding approaches, such as manipulation of transcriptional factors in carbon partitioning and regulation of the biosynthetic process of lipids, and overexpression of key enzymes in fatty acid biosynthesis and assembly. For example, overproduction of the regulator WRINKLED1 resulted in the enlarged seeds and an 10-40% increase of oil in *Arabidopsis thaliana* (Liu *et al*., 2010).

Over-expression of a yeast glycerol-3-phosphate dehydrogenase for the formation of glycerol-3-phosphate at the starting point of the Kennedy pathway boosted the lipid content by 40% in the seeds of oilseed rape (Vigeolas *et al*., 2007). Seed-specific expression of acyl-CoA: diacylglycerol transferase 1 (DGAT1) for the biosynthesis of TAG by acylating diacylglycerol (DAG) with acyl-CoA contributed to about 20% enhancement of seed oil content in *Arabidopsis thaliana* (Jako, 2001). In addition, increases of the activities of plastidial enzymes such as biotin carboxyl carrier protein isoform 2 (BCCP2) in acetyl-coenzyme carboxylase (ACCase), pyruvate transporter (BASS2) and purple acid phosphatase 2 could also enable the augmentation of oil levels in plants (Gu *et al*., 2017; Roesler *et al*., 1997; Zhang *et al*., 2012).

In plant, fatty acid biosynthesis occurs in plastids where ACCase catalyzes the formation of malonyl-CoA from acetyl-CoA initially, and then a type II FAS complex uses malonyl thioester as an extender for the synthesis of long chain saturated fatty acids such as 16:0 and 18:0 through a repetitive cycle of four catalytic reactions: condensation, keto-reduction, dehydration and enoyl-reduction (Magnuson *et al*., 1993). After synthesized, these two long chain fatty acids are often desaturated by a soluble acyl-ACP Δ9-desaturase in the stroma of plastids giving 16:1-9 and 18:1-9. Subsequently, these saturated and monounsaturated long chain fatty acids are exported to cytosol where they can be further modified by desaturations to introduce more double bonds or elongations to extend the chain length. With few exceptions (Meesapyodsuk *et al*., 2018), higher plants do not possess capacity to synthesize very long chain VLCPUFA with the chain length of 20C or more and the double bonds of two or more.

De novo biosynthesis of VLCPUFAs occurs mainly in microorganisms where their biosynthesis goes through either an aerobic pathway catalyzed by desaturases and elongases or an anaerobic pathway catalyzed by a PUFA synthase. *Thraustochytrium* is a protist producing a high level of a nutritionally important VLCPUFA DHA. Biosynthesis of this fatty acid is catalyzed by a PUFA synthase, a polyketide synthase-like mega-enzyme (Meesapyodsuk & Qiu, 2016; Zhao *et al*. 2016). Structurally, PUFA synthase shares similarity to both type I and type II fatty acid synthases in catalytic domains and motifs. Functionally, both PUFA synthase and Type I and II fatty acid synthases catalyze the synthesis of fatty acids through similar reiterative reactions such as condensation, keto-reduction, dehydration and enoyl-reduction using acyl-ACPs as substrates (Janßen & Steinbüchel, 2014; Schweizer & Hofmann, 2004; Yoshida *et al*., 2016). The PUFA

synthase in *Thraustochytrium* comprises three large subunits; each with multiple catalytic domains. However, the exact functions of these domains remain poorly understood. Our previous study showed KS domains from the PUFA synthase could functionally complement the defective phenotype of two *E. coli* mutants of acyl-ACP synthases. Particularly, overexpression of one KS domain from subunit-B (KS-B) resulted in accumulation of a high level of total fatty acids as well as unsaturated fatty acids (Xie *et al*. 2017).

The aim of this study was to extend our previous efforts on this domain to model plant *Arabidopsis* for further functional analysis. The result showed that the KS domain was functional in plant and could complement the defective phenotype of *Arabidopsis KASI* knockout mutant and the seed specific expression of this domain in wild type *Arabidopsis* significantly enhanced seed weight and seed oil, and promoted seed germination and seedling growth. These results emphasize that the condensation process of the fatty acid biosynthesis in plants is a limiting step and the enhanced activity by overexpression of the KS domain from a PUFA synthase of microbial origin offers an attractive biotechnological approach to increase oil content in commercial oilseed crops.

5.3 Experimental approach

5.3.1 Experimental flow
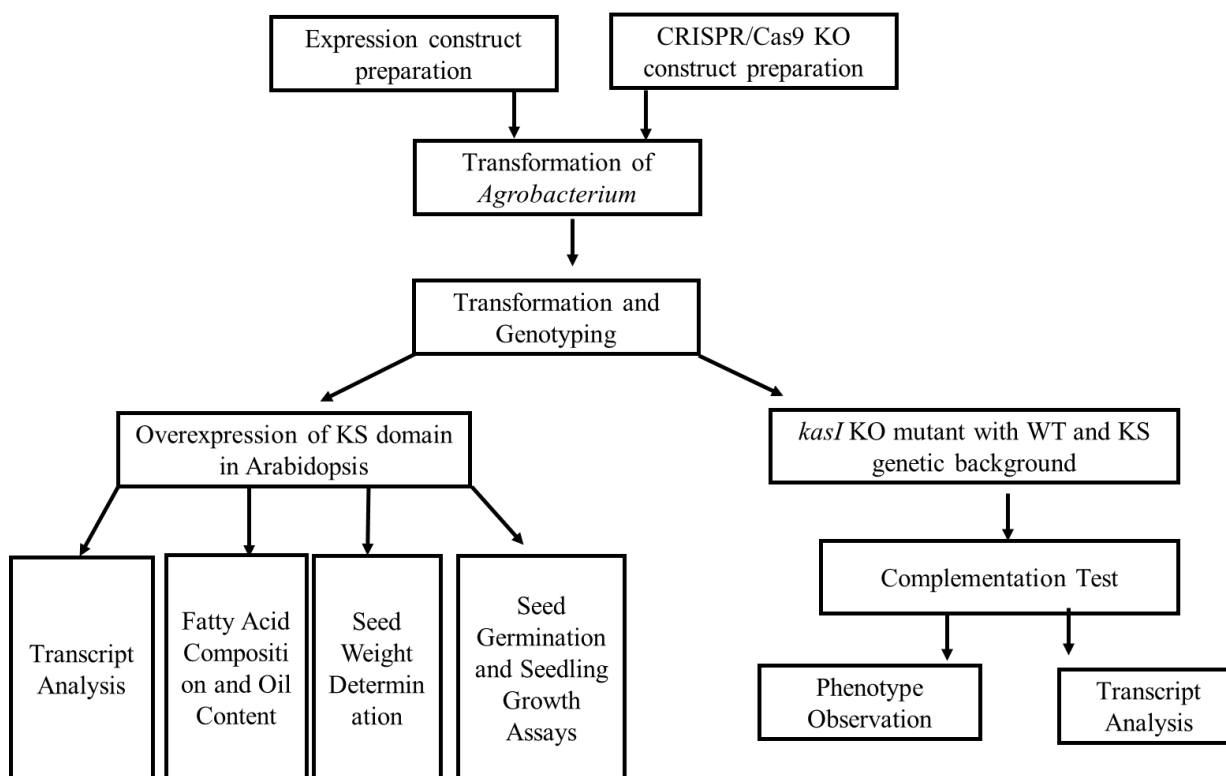
Figure 5.1 Experimental flow of chapter 5

5.3.2 Material and reagent

Agrobacterium tumefaciens and *E. coli* media were purchased from Bio Basic Inc. (York, Ontario, Canada). Fatty acids and the multiple standards were obtained from Nu-Chek-Prep, Inc. (Elysian, MN, USA). Q5 DNA polymerase, restriction enzymes, T7 Endonuclease and deoxynucleotide triphosphate (dNTP) were purchased from New England Biolabs (Ipswich, MA, USA). HP Taq DNA polymerase was obtained from Bio Basic Inc. (York, Ontario, Canada). Primers were synthesized from Sigma-Aldrich (St. Louis, MO, USA). GC grade solvents Fisher Scientific (Ottawa, ON, Canada). All other chemicals were purchased from Sigma-Aldrich Ltd (Oakville, ON, Canada). The intermediate vector and expression vectors for clustered regularly

interspaced short palindromic repeats/CRISPR-associated protein-9 nuclease (CRISPR/Cas9) system were purchased from Addgene (Watertown, MA, USA).

5.3.3 Construction of plant expression vectors

To express the KS domain in *Arabidopsis thaliana*, two different constructs with different selection and screening marker genes were built. The plastidial overexpression construct contains a KS domain from PUFA synthase (accession no. PRJNA368972) fused with a functional chloroplasts transit peptide (CTP) (for anchoring the KS domain protein to chloroplast) at the N-terminus (Lee *et al*., 2006) and a *Ds*Red2 from *Discosoma sp*. encoding a red fluorescent protein, each under the control of a seed-specific napin promoter (Josefsson *et al*., 1987), as well as an antibiotic kanamycin-resistance gene under the control of a constitutive nos promoter.

For disrupting *Arabidopsis KASI* (*AtKASI*), a CRISPR/Cas9 construct targeting *At*KASI was built. Two single guide RNAs (sgRNA) targeting *KASI* gene (AT5G46290) were designed using the previous protocol (website: (http://skl.scau.edu.cn) (Xie *et al*. 2017)). The sgRNA sequences were then inserted into an intermediate plasmid behind an AtU3d promoter. The recombinant AtU3d promoter-driving sgRNA cassettes were then inserted into pYLCRISPR/Cas9P$_{ubi}$-B vector (Ma and Liu, 2016). For transgenic seed screening, a napin promoter was used to guide a modified green fluorescent protein (eGFP) gene at the SpeI site of pYLCRISPR/Cas9P$_{ubi}$-B.

5.3.4 Transformation of *Agrobacterium* with recombinant vector

Electrocompetent cells of *Agrobacterium tumefaciens* GV3101 were prepared as follows. *Agrobacterium* cells were grown for 24 h in LB medium with 50 ug/mL gentamycin. When the A$_{600}$ reached 0.7, the cells were chilled on ice and pelleted by centrifugation (5,000 rpm for 10 min at 4°C). The pellet was washed in 1, 0.5, and 0.02 volumes of cold 10% (v/v) sterile glycerol and resuspended in 0.01 volume of cold 10% (v/v) glycerol. The electrocompetent cells were then frozen in liquid N$_2$ and stored at ≤-70°C. The *Agrobacterium* cells were transformed by electroporation with 20 to 50 ng of DNA according to the manufacturer's instructions. Transformed cells were plated on a selective medium (LB broth with corresponding antibiotic), and incubated for 48 h at 28°C. Single transformed cells were grown for 16 h (28°C, 225 rpm) in 5 mL LB broth with corresponding antibiotic and 50ug/mL gentamicin. The fidelity of the construct was rechecked by PCR before floral dipping transformation.

5.3.5 Transformation of *Arabidopsis* and genotyping of transgenics

The flowering plants in the growth chamber were used for transformation following the floral dip method (Clough & Bent, 1998). The transgenic seeds were selected by fluorescence, and grown on soil for the next generation. The transgenic plants were genotyped every generation and the seeds from positive lines were harvested and planted for the next generation. Genomic DNA of wild type and transgenic plants was extracted with Edward method (Edwards *et al*. 1991). The PCR condition was 30 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 30 s.

5.3.6 Transcriptional expression analysis

The samples of 7th-8th rosette leaf and developing siliques at 12 to 14 days after flowering (DAF) from T3 overexpressing lines and wild type were collected and frozen in liquid N_2. RNA was extracted by Qiagen RNeasy Mini Kit (Germantown, MD, USA) and treated with DNase I for 30 min to digest contaminating DNA in samples. The biosynthesis of complementary DNA was carried out using SuperScript™ III Reverse Transcriptase from Invitrogen (Carlsbad, CA, USA). Quantitative RT-PCR was performed with PowerUp™ SYBR™ Green Master Mix from Fisher Scientific (Carlsbad, CA, USA). The PCR conditions were as follows: 50◦C 2 min, 95◦C for 2 min, 40 cycles of 95◦C for 5 s, 60◦C for 1min, and one cycle of 95◦C for 15 s, 60◦C for 1 min, and 95◦C for 15 s for melt curve stage. The expression of the housekeeping gene *At*Actin-2 (AT3G18780) was used as references (Wang *et al*., 2014). The expression level was normalized by that of ACTIN-2 and the primer pairs used for the PCR were listed in Table.S1.

5.3.7 Fatty acid composition and oil content analysis

For fatty acid analysis, single seed was placed in a glass tube with a screw cap and 5 µl of 10 µg/mL heptanoic acid (C17:0) was added as an internal standard for quantification. 2 mL of 1% (v/v) H_2SO_4 with methanol was added to the glass tube. Samples were incubated for 10 min at 80◦C, and then seeds were pressed by glass rod and incubated at 80◦C for 1h to transmethylate fatty acids into their FAMEs. FAMEs were extracted with hexane and 1.0 mL of 0.9% NaCl was added to enhance phase separation. FAMEs were quantified using an Agilent 7890A system installed with a DB-23 column (30 m, 0.25 mm, 0.25 µm).

5.3.8 Seed weight determination

To examine seed weight, T2 seeds of three individual plants from different pots with the same genotype were mixed together and dried in open tubes for 3 days at room temperature. The seeds weight form wild type and four transgenic plants were collected and weight four times. For each replicate of four, 300 seeds of each sample collections were collected and weighed.

5.3.9 Seed germination and seedling growth

The surface-sterilized seeds were plated on agar plates of MS (Murashige and Skoog) medium containing 1% sucrose, and incubated in darkness for 3 days at 4 °C. The plates were then placed vertically in a growth chamber set to 22 °C under a 16-h-light (120 $\mu Em^{-2} s^{-1}$)/8-h-dark photoperiod for 4 days. Then the germination rate and seedling length were measured.

5.3.10 CRISPR/Cas9 knockout mutant screening and complementation test

The knockout mutant generated by CRISPR/Cas9 was genotyped in T2 plants by sequencing of target genes. The authentic mutant lines were used for the complementation test. The phenotype for the mutant and mutant complemented lines were observed and recorded every 7 days.

5.4 Results

5.4.1 Sequence comparison of the KS domain with 3-ketoacyl-ACP synthase I from plants

The PUFA synthase from *Thraustochytrium* was comprised of three subunits encoded by three large open reading frames (ORFs), each with multiple catalytic domains predicted by the presence of characteristic active site sequence motifs or residues (Meesapyodsuk and Qiu, 2016). One KS domain (KS-B) from residue 1 to 469 was identified in subunit-B of the PUFA synthase. Comparison of this KS domain with representative plant 3-ketoacyl-ACP synthase I (KASI) (Figure 5.2) showed that the KS domain shared only about 20% amino acid sequence identity with KASI enzymes from *Arabidopsis thaliana, Jatropha curca*, *Nicotiana tabacum*, *Oryza sativa* throughout the entire sequence. However, key residues in the active site such as a cysteine (C), two histidine (H) and a lysine (K) residues essential for the decarboxylation during the condensation process were completely conserved among these sequences. In addition, a glycine (G) residue presumably located at the accepting entrance of the substrate-binding tunnel, and two threonine (T) residues likely involved in hydrogen bonding with the ACP phosphopantetheine

moiety were also highly conserved. Furthermore, a glycine-rich motif located in the C-terminus contributing to the formation of an oxyanion hole for decarboxylation reaction was also observed among these sequences (Davies *et al*., 2000; White *et al*., 2005; Zhang *et al*., 2005).
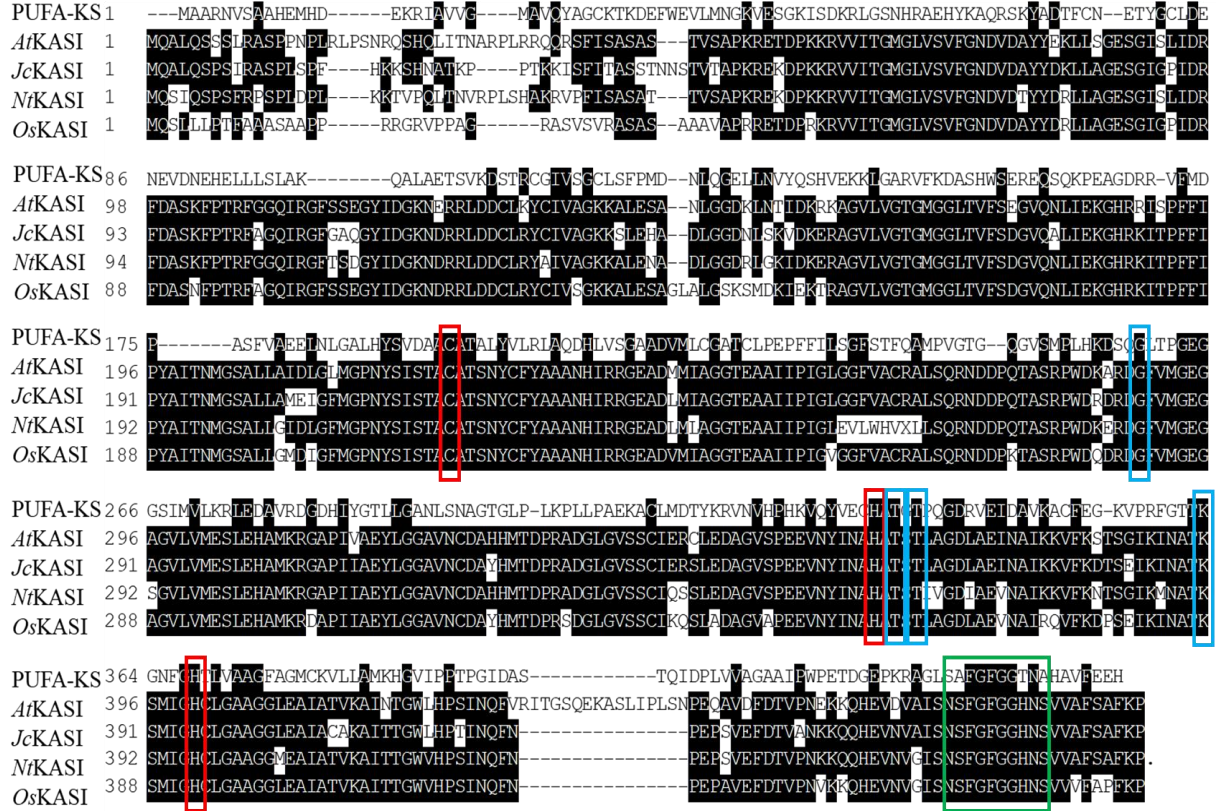
Figure 5.2 Comparison of the KS domain with KASI from plants. Red boxes represent conserved residues directly involved in the catalysis. Blue boxes represent conserved residues interacting with substrate, the green box represents a glycine motif in the C-terminus. PUFA-KS, PUFA synthase (PRJNA368972); *At, Arabidopsis thaliana* (NM_001036941); *Jc, Jatropha curcas L.* (KDP37953) *Nt, Nicotiana tabacum* (KX033513); *Os, Oryza sativa* (LOC_Os06g09630).

5.4.2 Generation of the constructs disrupting endogenous *KASI* and expressing the KS domain in *Arabidopsis*

To functionally analyze the KS domain in *Arabidopsis*, two binary plasmids were constructed. The first one was used to disrupt endogenous *KASI* genes, and the other was used to express the KS domain as a standalone enzyme from the PUFA synthase. To generate a construct to disrupt the *KASI* genes in *Arabidopsis*, two sets of overlapping sgRNA primer sequences targeting the conserved coding region of the gene were designed. The adaptor formed from each set of primers with a sgRNA sequence was inserted behind an *Arabidopsis* U3 promoter in an intermediate plasmid pLYsgRNA-AtU3d, giving a sgRNA expression cassette. Two sgRNA expression cassettes were amplified with universal primer sets (Ma and Liu, 2016) and assembled into a CRISPR/Cas9 binary vector pYLCRISPR/Cas9-P$_{ubi}$-B by a single golden gate ligation, giving the *KASI*-disrupting plasmid. Besides the two sgRNA expression cassettes for guiding the disruption of targeted genes, this construct was also comprised of a cassette expressing *Cas9* under a constitutive ubiquitin promoter for cutting the targeted sites, a cassette expressing a herbicide resistant *Bar* gene under a constitutive 35S promoter for selecting transgenic plants, and a cassette expressing a modified green fluorescent protein (eGFP) from *Aequorea victoria* under a seed-specific napin promoter for screening transgenic seeds (Figure 5.3a).

The second binary plasmid was constructed to express the KS domain in *Arabidopsis*. It was comprised of three expression cassettes. The first cassette was to express the domain as a standalone protein with a chloroplast transit peptide from *Arabidopsis* fused to the N-terminus under the control of the seed-specific promoter. The second cassette was to express an antibiotic kanamycin-resistance gene under the control of a constitutive promoter for selecting transgenic plants. The third cassette was to express a red fluorescent protein gene (*DsRed2*) from *Discosoma* under the seed specific promoter for screening transgenic seeds. (Figure 5.3 b).

After the confirmation of the structures of the two constructs by sequencing, they were introduced into *Arabidopsis* by floral dipping through an *Agrobacterium tumefaciens* transformation approach. Transgenic seeds with fluorescence were selected under a fluorescent microscope and grown to the next generation for further analysis.
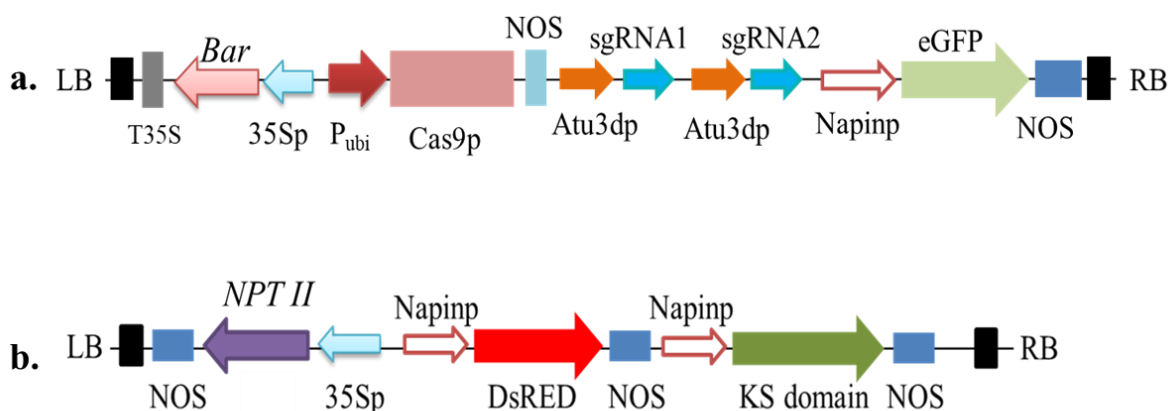
Figure 5.3 Schematic representation of the T-DNA region of two binary vectors. a. The T-DNA for of *AtKASI*-disrupting plasmid; **b**. The T-DNA region of the KS-domain expression plasmid.

5.4.3 Complementation of *Arabidopsis kasI* mutant by the KS domain

Arabidopsis KASI is essential for the biosynthesis of fatty acids in *Arabidopsis*, knockout mutants for this enzyme might not be easily obtained for the functional complementation test. Therefore, for the complementation assays of the KS domain in *Arabidopsis*, two approaches were attempted. The first approach was to generate a *kasI* mutant in wild type *Arabidopsis* using a CRISPR/Cas9 technique and to express the domain in the mutant subsequently, if it was amenable. The second approach was to express the domain first in wild type *Arabidopsis* and then to disrupt the *KASI* by the CRISPR/Cas9 technique under the KS-expressed background.

To knock out *KASI* in *Arabidopsis*, the disrupting construct was first used to transform wild type *Arabidopsis thaliana*. The transgenic seeds were selected under a fluorescent microscope and grown to the next generation for genotyping and phenotyping. Genotyping transgenic T2 plants by sequencing the amplicons of targeted genes identified three homozygous lines with deletion mutation between two targeted sites of the *KASI* gene (*kasI*-1,4,11). However, all these lines exhibited similar abnormal growth, smaller and shorter seedlings at the early growth stage, and semi-dwarf at the late growth stage (Figure 5.4), compared to the wild type. The similar

phenotype was previously observed in a *kasI* mutant generated by a T-DNA insertion (Wu & Xue, 2010). Due to the severe growth defect, re-transformation of the mutant with the KS-expression construct for the complementation test was not successful.

Figure 5.4 Complementation of the *kasI* mutant in *Arabidopsis*. a. Growth phenotypes of *kasI* mutant line, *kasI* complementation line (*KS/kasI*) at 23 days after growing (DAG). Bars = 2 cm. b. Growth phenotypes of *kasI* mutant line, *kasI* complementation line (*KS/kasI*) at 35 DAG. Bars = 2 cm.

As the first approach for the complementation assay was unsuccessful, we then resorted the second approach, expressing the KS domain in *Arabidopsis* first and then disrupting *KASI* under the KS-expressed background. A transgenic plant expressing the KS domain obtained using the second construct for the first transformation was re-transformed with the disrupting construct of CRISPR/Cas9, and transgenic seeds were selected on both green (by eGFP) and red (by DsRed) fluorescence signals and grown the next generation for genotyping and phenotyping. Genotyping analysis identified two mutative T2 lines (*KS/kasI*-9, *KS/kasI*-15) with insertion mutation in the target genes. One had a single nucleotide insertion, while the other had a two-nucleotide insertion close to the targeted sites; both resulted in shifting the open reading frames. Phenotypic observations showed that the two transgenic plants had comparable growth and development to the wild type, and no abnormal phenotypes were detected under the growth condition (Figure 5.4).

To confirm the expression of the KS domain, two types of reverse transcription PCR (RT-PCR) were performed in the complementation line. The total RNAs were isolated from the 7th-8th rosette leaves and the developing siliques at 14 days after flowering, and then reverse-transcribed to cDNAs (Baud *et al*., 2008). A semi-quantitative PCR analysis of the cDNAs showed that no expression of the KS domain was observed in the untransformed *Arabidopsis* control, but the expression of the KS domain could be obviously detected in both leaves and developing siliques of the complementation line (Figure 5.5a). A quantitative PCR analysis showed that the transcription level of the KS domain in developing siliques was 10-fold higher than that in leaves (Figure 5.5b). This result indicated that the KS domain guided by a seed-specific napin promoter indeed possessed significantly higher expression in developing seeds than leaves, but it did express in leaves, providing the function for complementing the *KASI* defective phenotype in *Arabidopsis*.

Figure 5.5 Transcriptional analysis of the *KS* domain in leaves and developing siliques of a complementation line. a. Expression of the *KS* domain and *AtActin-2* in leaves and developing siliques of *KS/kasI*-9 analyzed by semi-quantitative RT-PCR. **b**. Expression of the *KS* domain in leaves and developing siliques of *KS/kasI*-9 analyzed by quantitative RT-PCR. *KS/kasI*, *kasI* mutant complementation line; WT-L, leaves of wild type; WT-S, siliques of wild type; CE-L, leaves of complementation line; CE-S, siliques of complementation line. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis of the results was conducted using the one-way analysis of variance (P<0.05)

88

5.4.4 Seed-specific expression of the KS domain in wild type *Arabidopsis*

To examine the impact of the KS domain in fatty acid biosynthesis and oil accumulation in *Arabidopsis*, the second construct expressing the domain under the seed-specific promoter was introduced into wild type *Arabidopsis*. Transgenic seeds were selected on a red fluorescence signal and propagated to the next generation. To confirm the expression of the KS domain, the total RNAs were isolated from the developing siliques of four transgenic lines (OE-6, OE-8, OE-13, and OE-16) and reverse transcribed to cDNAs. A semi-quantitative PCR analysis of the cDNAs showed that the expression of the KS domain was much higher than that of the housekeeping gene *AtActin-2* (Figure 5.6a). A quantitative PCR analysis showed that the transcription levels of the KS domain were highly varied among the four lines. The expression level of the KS domain in OE-6 was the highest, followed by OE-16 and OE-13, and OE-8 had the least expression (Figure 5.6b). Subsequently, the oil and mass of the seeds in these four transgenic lines were measured. As shown in Figure 5.7, all the transgenic lines accumulated more oil relative to the wild type control. The amount of oil in OE-6, OE-16, OE-13, and OE-8 was increased by 80%, 77%, 69% and 48%, respectively, over the control. In particular, the level of monounsaturated fatty acids such as 18:1and 20:1 was significantly higher in these transgenic lines over the wild type, the indicator of the seed triacylglycerol production previously seen (Lemieux *et al*., 1990). In addition, the decrease of medium-chain fatty acid 14:0 and the increase of long-chain fatty acids of C16 and C18 and very long chain saturated fatty acid 20:0 were also observed in the transgenic lines (Table 5.1). Intriguingly, the transgenic seeds were also significantly heavier, representing 1.4, 1.2, 1.4 and 1.5-fold referring to the wild type, respectively. The percentages of lipid content in these transgenic lines exhibited the increases by 14.7% to 28.1% over the control. Nevertheless, no significant difference in the seed protein content was detected between these transgenic lines and the control (Figure 5.S2). These results indicate that expressing the KS domain can significantly increase both seed oil and seed mass in *Arabidopsis*.
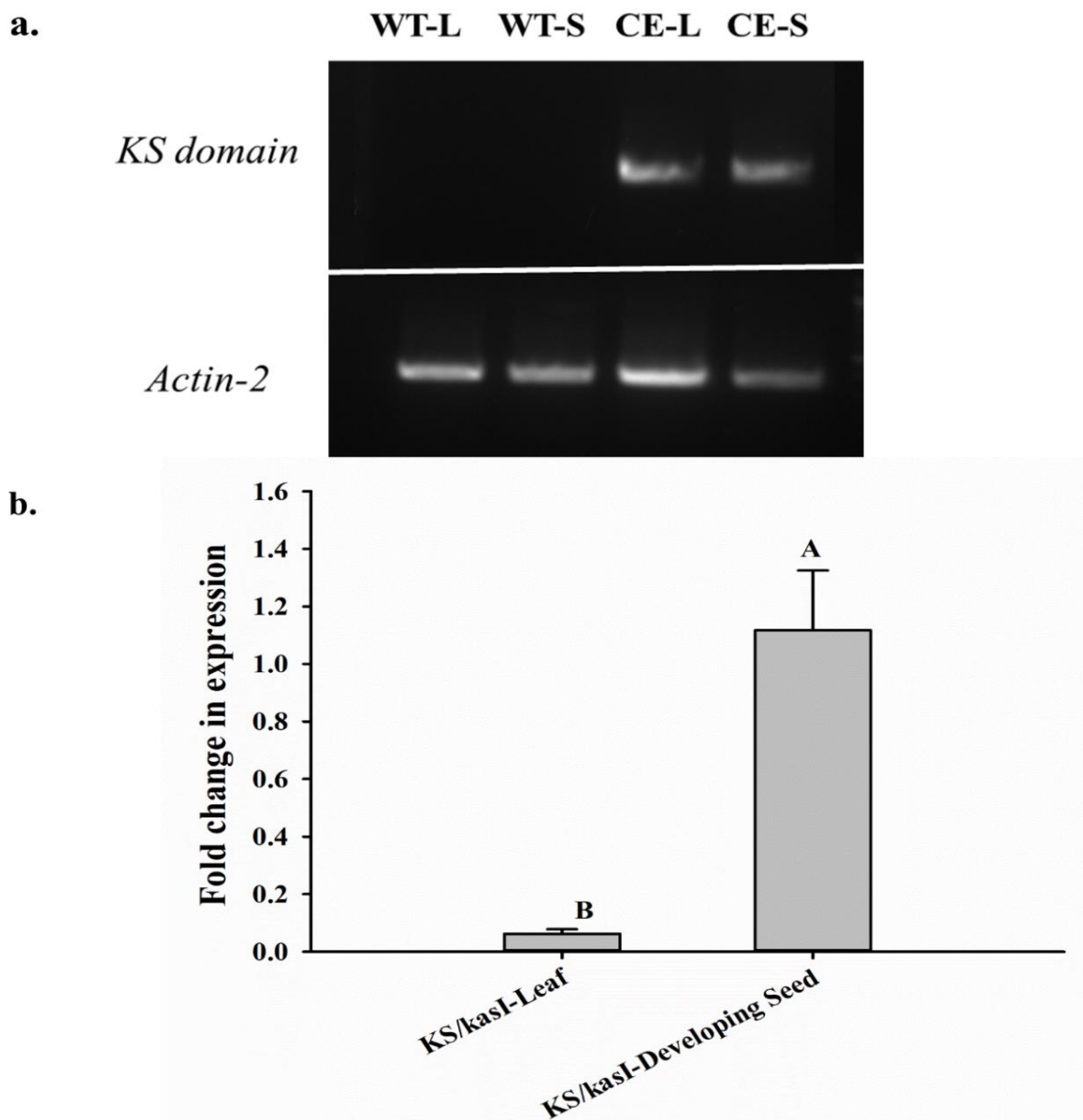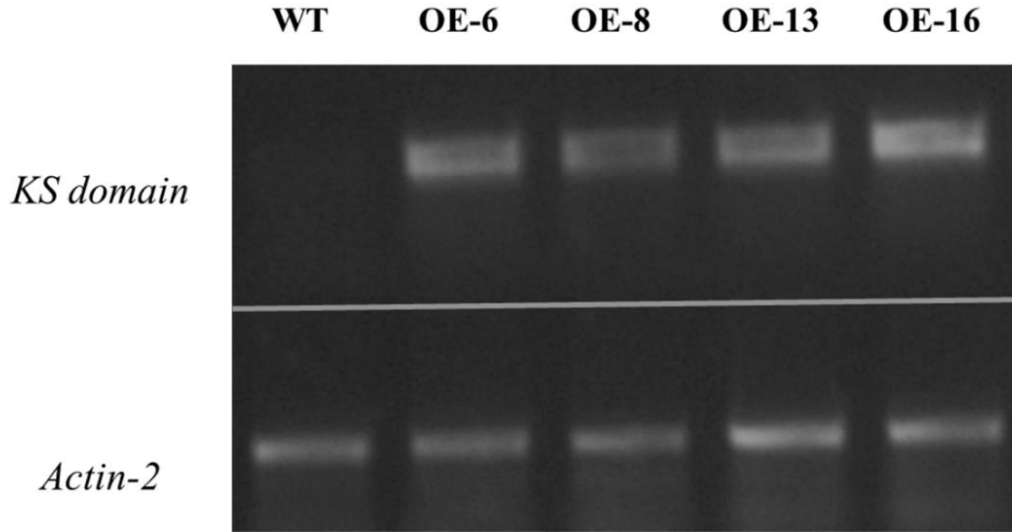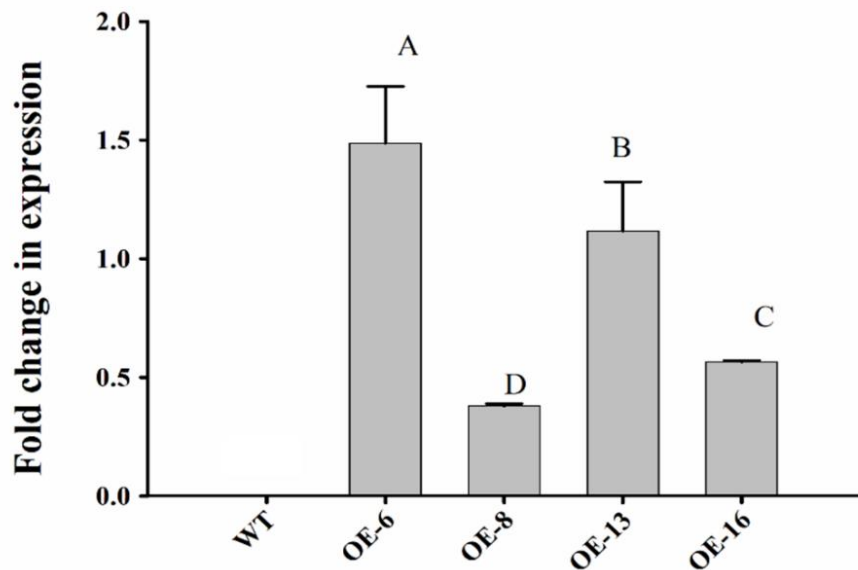
Figure 5.6 Transcriptional analysis of the *KS* domain in developing siliques of overexpression lines. a. Expression of the *KS* domain and *AtActin-2* in the developing siliques of *KS/WT* analyzed by semi-quantitative RT-PCR. **b**. Expression of the *KS* domain in leaves and developing siliques of *KS/WT* analyzed by quantitative RT-PCR. WT, wild type; OE, overexpression. Values are reported as the means of three biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis of the results was conducted using the one-way analysis of variance (P<0.05)
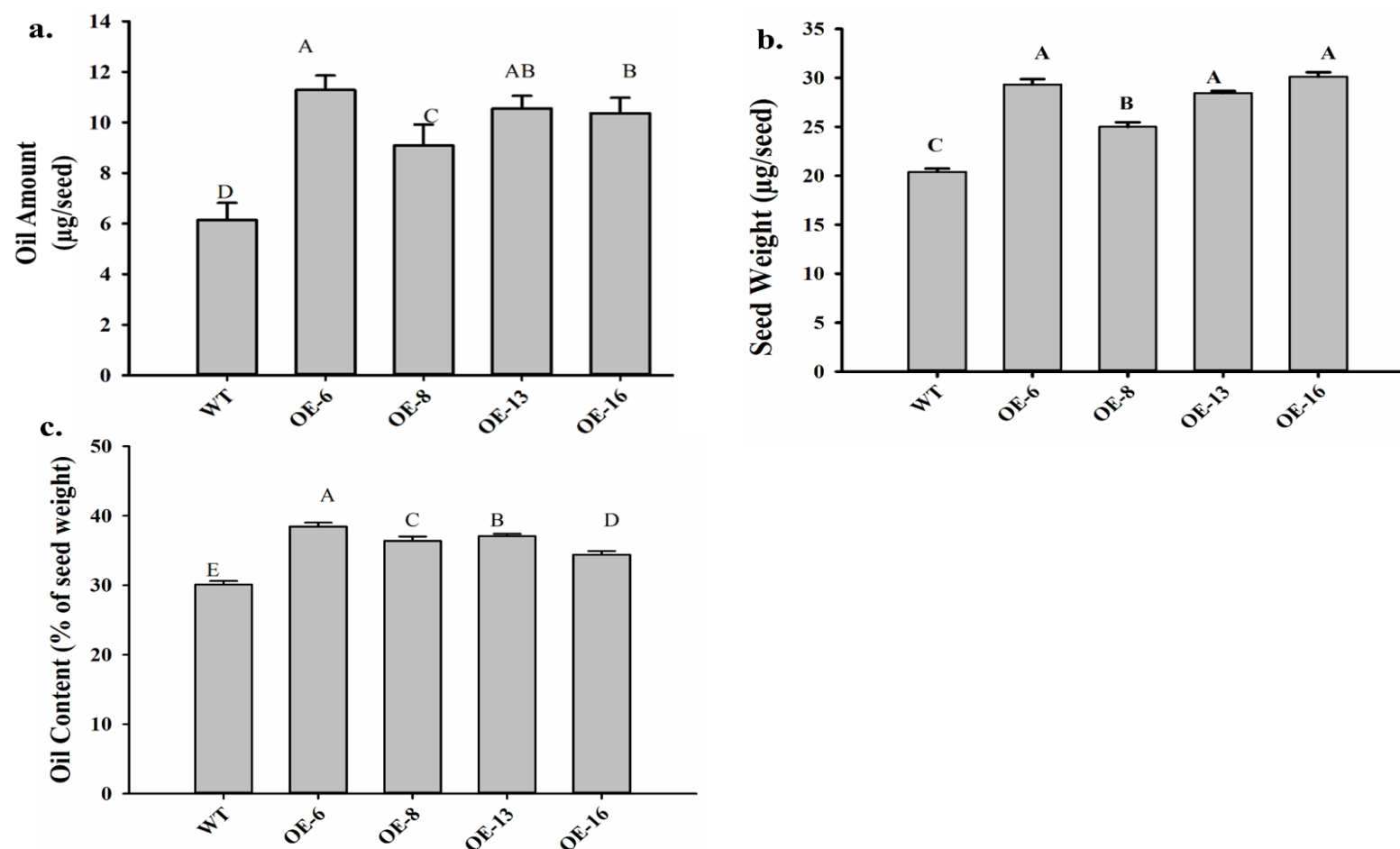
Figure 5.7 The amount of oil, seed weight and oil content in transgenic seeds overexpressing the KS domain. a. the amount of oil per seed in the T2 seeds of overexpression lines; **b.** the seed weight of the transgenic lines; **c.** seed oil content of the transgenic lines. WT, wild type; *KS/kasI*, *kasI* mutant complementation line; OE: overexpression lines. Values are reported as the means of 10 biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis of the results was conducted using the one-way analysis of variance (P<0.05)

Table 5.1 Analysis of fatty acid methyl esters (FAME) of the FA composition (mol%) from WT and OE lines. Numbers in each column refer to the relative molar ratios of the different FA with the total being 100%. Means and standard deviation of four independent samples are presented; US/S refers to the ratio of unsaturated/saturated FA. The asterisk indicates significant difference in fatty acids composition between transgenic plants and WT (* $p < 0.05$, ** $p < 0.01$).

FA species	WT	OE-6	OE-8	OE-13	OE-16
14:0	4.26±0.52**	1.40±0.18*	1.43±0.18*	0.90±0.19	0.64±0.20
16:0	11.64±1.17*	10.07±0.20	11.72±0.26*	10.27±0.39	10.13±0.23
16:1-9	0.21±0.10	0.74±0.03	0.80±0.06	0.66±0.05	2.11±0.98**
18:0	8.17±1.44	7.29±0.38	9.68±0.37**	8.38±0.33*	7.71±0.52
18:1-9	10.28±1.10	17.38±0.70**	11.98±0.77*	17.93±0.61**	12.89±1.78*
18:1-11	0.90±0.45	1.90±0.12	2.03±0.16*	2.81±0.23*	4.57±1.97**
18:2	21.96±1.87*	20.49±0.56	19.94±0.77	20.50±0.41	21.10±0.81*
18:3	14.94±1.69	14.35±0.53	15.59±0.80*	12.30±0.26	14.31±1.47
C16-18	68.10±1.72	72.23±1.52**	71.74±3.19**	72.85±1.29**	72.81±1.77**
20:1-11	14.58±0.85	18.23±0.27**	16.85±0.43*	16.20±0.38*	18.64±0.35*
SFA	31.94±1.66**	23.03±0.73	29.06±0.73*	24.72±1.11	22.17±0.89
UFA	67.39±2.66	76.97±0.82**	70.94±1.13*	75.28±0.71**	77.83±0.56**
UFA/SFA	2.11	3.34	2.44	3.05	3.51

5.4.5 Effect of the KS domain expression on early seedling growth in *Arabidopsis*

To examine the growth effects of the KS domain, the phenotype of the transgenic *Arabidopsis* plants was carefully observed during the life cycle. No obvious phenotypical differences could be detected between transgenic and untransformed wild type control except for seed germination and early seedling growth. The seed germination rates of four transgenic lines were considerably higher (217%, 205%, 211 % and 203%) than that of the control (Figure 5.8a). The lengths of four-day seedlings of the transgenic lines were all longer compared to the wild type (Figure 5.8b). This result indicated the expression of the KS domain from the PUFA synthase in *Arabidopsis* could promote seed germination and seedling growth.
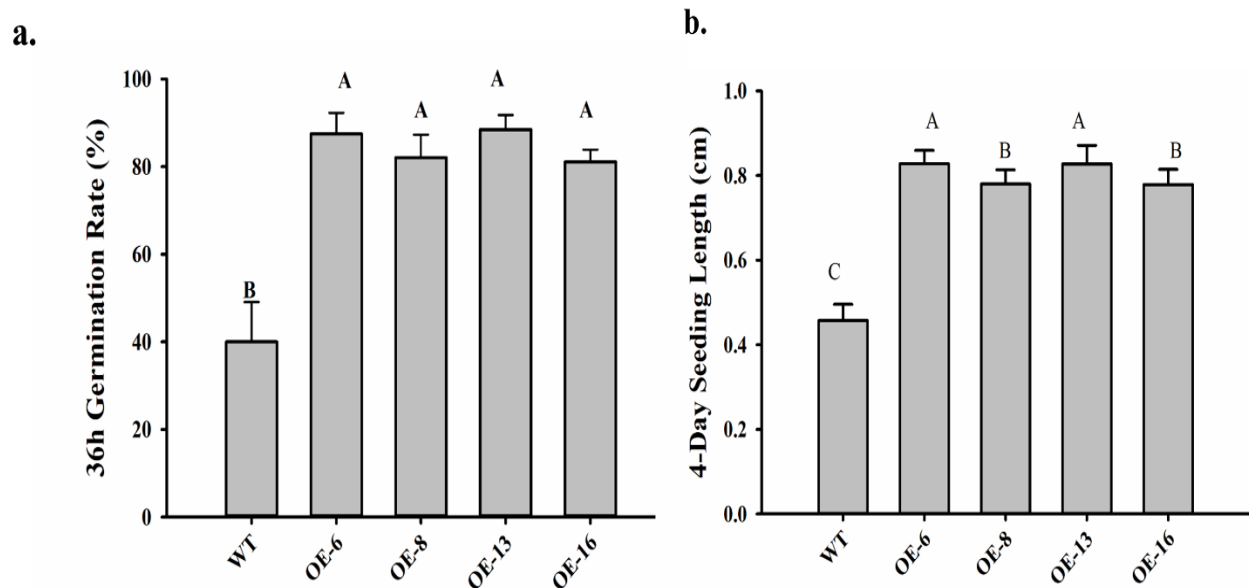
Figure 5.8 Effect of KS domain expression on seed germination and early seedling growth in *Arabidopsis*. a. Germination rates of wild type (WT) and transgenic lines (OE) at 36 h of imbibition. **b.** The 4-day seedling length of WT and OE lines. WT, wild type; OE, overexpression lines. Values are reported as the means of 10 biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis of the results was conducted using the one-way analysis of variance (P<0.05)

5.5 Discussion

In *Thraustochytrium*, the biosynthesis of VLCPUFAs is catalyzed by a PUFA synthase, a PKS-like mega-enzyme comprising three subunits each with multiple catalytic domains predicted on the characteristic active site residues (Meesapyodsuk and Qiu, 2016). However, the exact functions of these domains in the biosynthesis of these fatty acids are not characterized. Our previous study indicated that a dissected KS domain from subunit-B of the PUFA synthase from *Thraustochytrium* could complement the temperature sensitive phenotype of *E. coli* β-ketoacyl-ACP synthase I mutant (*FabB*) and enhance unsaturated fatty acid and total fatty acid production in a wild type *E. coli* strain (Xie *et al*., 2017). In the present study, this domain was further functionally analyzed in *Arabidopsis thaliana*. As the *kasI* mutant of *Arabidopsis* was defective in growth and development, and not amenable for re-transformation to express the KS domain for the complementation test, we resorted expressing the KS domain in *Arabidopsis* first and disrupting the endogenous *KASI* genes subsequently using a CRISPR/Cas9 system. Two homozygous insertion mutants of the target genes on the KS-expressed background were identified. Both resulted in the open reading frame shifting, which leads to the inactivation of the *KASI* genes (Schiml & Puchta, 2016). In the two complementation lines, the defective phenotypes of *kasI* knockout mutant such as small seedlings and dwarf plants were not observed. This result indicates that the KS domain from the PUFA synthase can functionally replace KASI for the biosynthesis of fatty acids in *Arabidopsis* (Figure 5.4). The RT-PCR analysis showed that transcription of the KS domain was detected in both leaves and developing seeds under the napin promoter, although the transcription level in developing siliques was 10 times higher than that in leaves. The similar expression pattern of the napin promoter has been seen in transgenic *Brassica napus* and tobacco (Ellerström *et al*., 1996; Kjell *et al*., 2016). Nevertheless, the level of the KS domain expression under the promoter was enough for the functional complementation in *Arabidopsis*.

After the success of functional complementation, we then overexpress the KS domain in wild type *Arabidopsis* under the seed-specific promoter. The result showed that seed weight was increased by 20%–40%, and seed oil was increased by 14.70% to 28.16% in these transgenic lines. The degree of the increases was coincided with the expression level of the KS domain in the siliques of these transgenic lines (Figure 5.7). In addition, expression of the KS domain also had positive effects in seed germination and early seedling growth. This result suggests that the

expression of the KS domain from the PUFA synthase could boost the *de novo* biosynthesis of fatty acids, thereby increasing the oil content in transgenic seeds.

In plant, the biosynthesis of long chain fatty acids occurs in plastids by a type II fatty acids synthase complex comprising multiple discrete enzymes (Magnuson *et al*., 1993). Among these enzymes, three β-ketoacyl-ACP synthases (KASI, KASII, KASIII) are key enzymes responsible for the condensing reaction, the first of four chemical reactions in a repetitive cycle of fatty acid synthesis by the addition of a two-carbon unit a time donated by malonyl-ACP to a primer acyl chain. KASIII catalyzes the first condensation of fatty acid synthesis between acetyl-CoA/ACP and malonyl-ACP, giving a C4 acyl chain. KASI catalyzes the subsequent condensation reactions from C4 to C16 acyl chains. KASII catalyzes the final condensation of fatty acid synthesis giving a C18 acyl chain. After the biosynthesis, long chain fatty acids are modified in cytosol and assembled into storage glycerolipids TAGs through the Kennedy pathway (Baud & Lepiniec, 2010).

In the past few years, several molecular strategies have been attempted to increase oil production in oilseed crops. One strategy is to increase fatty acid assembly into triacylglycerols by overexpressing acyl-CoA acyltransferase in the Kennedy pathway. For instance, an overexpression of lysophosphatidic acid acyltransferase (LPAAT) or DGAT1 significantly increase oil contents in seeds. Fatty acid profile in the oil would be determined by substrate specificity of acyl-CoA acyltransferases used for fatty acid acylations (Jako, 2001; Savadi *et al*., 2015). Another strategy is to increase the efficiency of fatty acid synthesis prior to the assembly by overexpression of enzymes directly involved in the biosynthetic process. For instance, over-expression of *ACCase* resulted in the increase of oil content (Roesler *et al*., 1997; Cui *et al*., 2017; Gu *et al*., 2017). In this strategy, fatty acid profile of the oil would be similar to the wild type. In the present study, the expression of the KS domain from the PUFA synthase significantly enhanced oil content in transgenic lines and at the same time the fatty acid profile of seed oil was also altered to the higher proportion of unsaturated fatty acids. This result indicates the condensation process, particularly the condensation steps from C4 to C16, during the fatty acid biosynthesis might be the limiting step, over-expression of the KS domain, although it only shares about 20% amino acid identity to *Arabidopsis* KASI, can increase seed oil and seed mass simultaneously in *Arabidopsis*. The reason why the amount of unsaturated fatty acids was also increased in the transgenics remains to be determined. It is probably due to the unique activity of the KS domain of microbial original.

Overexpression of the domain may not only enhance the condensation reaction, but also stimulate the activity of down-stream fatty acid modifying enzymes such as acyl-ACP and phospholipid desaturases and FAE type elongases for altering fatty acid profile.

In summary, seed oil is an important renewable bioresource for food and bioproducts. This research provides a new strategy to enhance oil and unsaturated fatty acid production by expressing a KS domain from type I fatty acid synthase-like enzymes such as PUFA synthases and polyketide synthases of microbial origin in oilseed crops for nutritional and industrials uses.

5.6 Supplemental Data

Table 5.S1 Primers used for this study

Category	Primer	Sequence	Annotation
Expression construct preparation	F-CTP	ATGGCTTCCTCTATGCTCTCTTCCG	Primers for cloning CTPs
	R-CTP	GAATTCCTTGTCGTCGTCGTCCTTGTAGTC	
	F-KS-B	GAATTCATGGCCGCGCGCAAC	Primers for cloning KS-CLF domain
	R-KS-L-B	AAGCTTCTACATGTTGTTCTCGCCGCC	
Genotyping	KS-B TEST-F	GCACGAGATGCACGATGAGAA	Primers for sequencing and qRT-PCR
	KS-B TEST-R	ATGGCTTTGGTACACGTTGAGG	
	F-KASI-1-seq	GTGCTTCTCCTCCAAACC	Primers for sequencing
	R- KASI-seq	CAACAAACCCAATAACCAAA	
Crisper/Cas9 KO construct preparation	SpeI-Forward	CCGACTAGTTAATGAGATATGCGAGACG	cloning GFP with napin promoter and terminator
	SpeI-Reverse	CCGACTAGTTGTGATTGTATGTGTGTGCT	
	AtU3d-KASI-F1	gtcaTCTTCGCAAAGGTCTCGCAT	
	AtU3d-KASI-F2	gtcaCGTTTAGGAGCGGAGACAG	sgRNA primers
	AtU3d-KASI-R1	aaacATGCGAGACCTTTGCGAAGA	
	AtU3d-KASI-R2	aaacCTGTCTCCGCTCCTAAACG	
qRT-PCR	F-ACTIN2	GATTCCTGGACCTGCCTCAT	Actin-2 primers for qPCR
	R-ACTIN2	TACCCGATGGGCAAGTCA	
	KASI-qPCR-F	AGCCATCGCCACCGTGAA	AtKASI primers for qPCR
	KASI-qPCR-R	CAAGTGGTGGTGATGGATGATG	

Figure 5.S1 Comparison of sequences surrounding the sgRNA sequences in the *KASI* genes. *KASI*, original *KASI*; *KasI*, *kasI* knockout mutant; *KS/kasI*, *KasI* knockout mutant with KS domain; Blue boxes represent PAM sites and Green boxes represent the target sites.
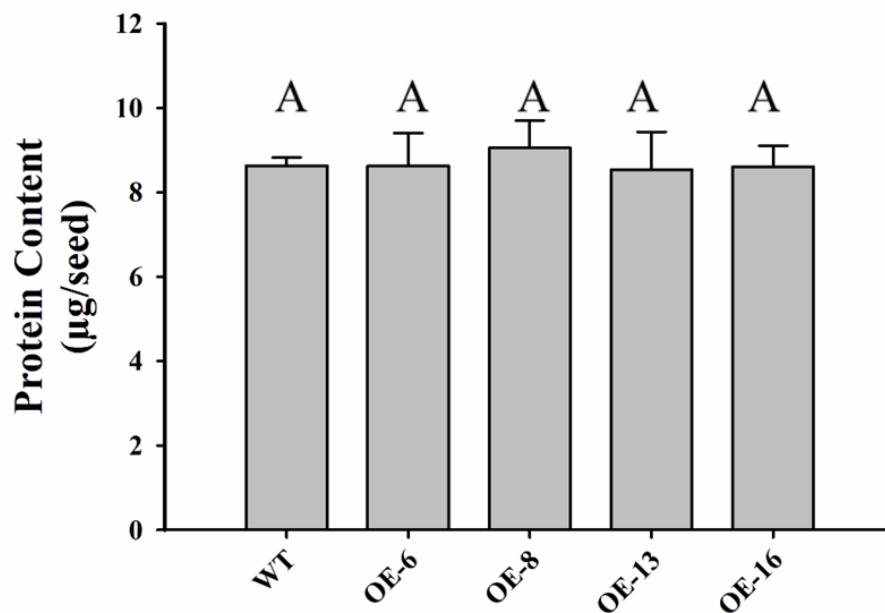

Figure 5.S2 Protein content in transgenic seeds overexpressing the KS domain. Values are reported as the means of 4 biological replicates along with standard deviation. The means with the same letters are not statistically significantly different. Statistical analysis of the results was conducted using the one-way analysis of variance (P<0.05)

CHAPTER 6. GENERAL CONCLUSIONS AND DISCUSSIONS

The present doctoral thesis was guided by the following hypotheses: DHA biosynthesis by a PUFA synthase in *Thraustochytrium* involves coordinated activity of multiple catalytic domains such as KS, KR, MAT, ACP, DH and ER domains. Among these domains, KS and DH domains are for positioning double bonds during the biosynthetic process. Specifically, **(1)** the KS domains are capable to catalyze the condensation of unsaturated acyl-ACP with malonyl-ACP, resulting in retaining a double bond in the acyl chain. **(2)** The DH domain can catalyze the specific dehydration and isomerization of hydroxyl acyl-ACP to introduce a cis double bond in the acyl chain. Coordination of these two types of domains can position multiple double bonds in VLCPUFAs intricately. **(3)** KS-B domain from a PUFA synthase could cooperate with other components of type II FAS for fatty acid biosynthesis in bacteria and plants. The overall objective of this research was to understand VLCPUFA biosynthesis catalyzed by the PUFA synthase in *Thraustochytrium* with focus on the functions of two catalytic KS and DH domains.

(1) Functional analysis of KS domains from PUFA synthase of *Thraustochytrium*

Biosynthesis of VLCPUFAs in *Thraustochytrium* is solely catalyzed by a PUFA synthase with multiple catalytic domains (Meesapyodsuk & Qiu, 2016). However, limited information is available on how multiple double bonds are positioned in VLCPUFAs and what the functions of these domains are, particularly DH and KS domains for introducing and retaining double bonds during the biosynthesis of VLCPUFAs.

In the first study, we aimed at functionally analyzing two KS domains from a PUFA synthase with three subunits in *Thraustochytrium*. On the basis of the sequence analysis including homology search, multiple sequence alignment and homology modeling, one KS domain (KS-A) was identified in subunit-A while the other KS domain (KS-B) was found in subunit-B of the synthase. In addition, one CLF domain, which shares sequence similarity to KS domain, but lack the active site residues, was also found to be adjacent to KS-B domain. The two KS domains have amino acid sequence identity with each other at about 20% and both share similar sequence similarity with FabB and FabF. To characterize functions of these two KS domains of PUFA synthase, both KS domains were expressed in *E.coli* β-ketoacyl-ACP synthase mutants (*FabB* and *FabF*) for

complementation assay, site-directed mutagenesis analysis. The complementation test of a *FabB Ts* mutant showed that both KS domains could complement the defective growth phenotype. Specifically, the *FabB* mutant expressing KS domains could grow normally at a non-permissive temperature. Fatty acid analysis confirmed that the mutant transformants with KS domains from either subunit-A and subunit-B produced a high level of UFAs to sustain the growth at the non-permissive temperature (42°C), as seen in the wild-type strain. In addition, expression of KS domain from both subunit-A and subunit-B in a *FabF* knockout mutant could complement the defective phenotype by increasing the fatty acids production. However, KS-B domain is more similar to FabB as it could accumulate a higher level of UFAs as FabB did, while KS-A, like FabF, could improve the overall production of fatty acids, particularly for the production of SFAs, suggesting that KS-A is comparable to FabF in catalytic function.

Studying the molecular mechanism of the PUFA synthase for the biosynthesis of VLCPUFAs is challenging since this mega-enzyme is encoded by three large ORFs along with high G/C content and repetitive domain sequences; thus, cloning and expression of this large DNA fragments in heterologous systems for functional characterization are difficult. The PUFA synthase is a mega-enzyme, the borderlines of all the catalytic domains remain unclear, therefore, efforts on defining the boundary of domains are required. In addition, *Thraustochytrium* is not amenable to genetic manipulation and effective heterologous expression systems are not available. In this study, we utilized bioinformatics tools to dissect the PUFA synthase sequence into small fragments containing intact catalytic domains, and then expressed the embedded domains as standalone enzymes in *E. coli* for functional analysis. Successful complementation and functional expression of the KS domains in *E. coli* is the first step towards the future *in vitro* enzymatic assays with purified domain proteins and other components for elucidating the molecular mechanism for the condensation reactions in the biosynthesis of VLCPUFAs.

(2) Functional analysis of the DH domains of PUFA synthase from *Thraustochytrium*

Intrigued by the finding from the first study of this thesis that the embedded KS domains of PUFA synthase could function as standalone enzymes in *E. coli*, we then functionally analyzed DH domains from the same PUFA synthase in the similar way. According to sequence analysis, there are three DH domains in the PUFA synthase of *Thraustochytrium*, one in the C-terminal region of subunit-A (DH-A) and the other two in the N-terminal region of subunit-C (DH1-C and

DH2-C). The sequence comparison showed that DH-A shares very low sequence similarity to DH1/DH2-C and *E.coli* FabA/Z, and modest sequence similarity including conserved active site motifs to the DH domains from PKS. Two DH domains from subunit-C (DH1-C and DH2-C) share only 15% sequence identity with each other, but 32% and 27% identity with *E. coli* FabA, including conserved active site motifs.

To characterize the functions of DH domains, these domains were expressed in an *E.coli* temperature-sensitive mutant defective in β-hydroxyacyl-ACP dehydratase (FabA) activity. Results showed that the mutant expressing DH domains of PUFA synthase, like the control, could grow on a non-permissive temperature. Fatty acids analysis showed that the amounts of fatty acids such as 16:0 and 18:1-11 were significantly higher in the mutant expressing the DH domains than those in the mutants with empty vector. This result indicates that the DH domains possess similar functions of β-hydroxyacyl-ACP dehydratase. Site-directed mutagenesis was conducted on DH domains by replacing the putative active site residues (Asp/Glu) to alanine. The mutated DHs were expressed in both *FabA* mutant and wild type *E. coli* strains. Complementation assay in the *FabA* mutant with mutated DHs showed when the conserved residue (Asp/Glu) was changed to an alanine residue, the substitutions could not correct the defective phenotype of the mutant, confirming the critical role of this amino acid in defining the catalytic activity in those DH domains. Interestingly, overexpression of mutated DH domains from subunit-C (DH1-M-C and DH1-M-C), but not mutated DH domain from subunit A (DH-M-A), showed the "dominant negative mutation effect", resulting in a significant reduction of fatty acid production relative to the wild type control. A combination of evidence from sequence comparison, functional expression and mutagenesis analysis indicate that DH domain from subunit-A might function as a common dehydratase mainly responsible for introducing a 2-*trans* double bond, while tandem DH domains from subunit-C might introduce a 3-*cis* double bond during the biosynthesis of VLCPUFAs in *Thraustochytrium*.

(3) Functional analysis of the KS-B domain of PUFA synthase in *Arabidopsis*

The result of study 1 indicates KS-B domain plays a role comparable to FabB or β-ketoacyl-ACP synthase, catalyzing the main condensation reactions in bacterial type II fatty acid synthesis. To investigate how KS-B functions in plant type II fatty acid synthesis, this domain was expressed in model plant *Arabidopsis thaliana*.

Comparison of the KS-B domain with KASI enzymes from plants showed that it shares only about 20% amino acid sequence identity with KAS I from *Arabidopsis thaliana, Jatropha curca, Nicotiana tabacum* and *Oryza sativa* throughout the entire sequence; however, key residues including important ones for catalytic function were highly conserved in the domain. The function of KS-B was first analyzed in *Arabidopsis kasI* mutants which were generated by CRISPR/Cas9. Defective growth phenotypes such as small seedlings and dwarf plants were observed in *kasI* knockout mutants, but not in a *kasI* mutant expressing KS-B domain. This result indicates that the KS-B domain from the PUFA synthase can functionally replace KASI for the biosynthesis of fatty acids in *Arabidopsis*. Subsequently, KS-B domain was overexpressed in wild type *Arabidopsis* under a seed expression promoter. The result showed that overexpression of the domain in *Arabidopsis* could enhance seed oil content and seed weight, and promote seed germination and early seedling growth. These data indicate that expressing a KS domain from mega-enzymes such as PUFA synthases and polyketide synthases of microbial origin can be used to improve oil production in oilseed crops for nutritional and industrials uses.

(4) Significance

PUFA synthase responsible for the biosynthesis of VLCPUFAs in *Thraustochytrium* comprises multiple catalytic domains including KS, KR, MAT, ACP, DH and ER domains. However, the catalytic functions of these domains are largely unknown. The research work presented in this thesis attempted to analyze the functions of two important domains, KS and DH, from a PUFA synthase in *Thraustochytrium* by complementation assay, site-directed mutagenesis and overexpression in *E. coli* and *Arabidopsis*. Results obtained from this thesis research provides an insight into molecular mechanism underlying the biosynthesis of VLCPUFAs by PUFA synthase and offers novel biotechnological strategies for improving fatty acid and oil production in oilseed crops using catalytic domains of PUFA synthase of microbial origin.

CHAPTER 7. FUTURE DIRECTIONS

Demand for VLCPUFAs such as ARA, EPA and DHA in the global market is increasing due to the public awareness of potential health benefits of these fatty acids such as vision development, brain performance, and protection against cardiovascular and immunological diseases (Reich *et al*., 2001; Diau *et al*., 2005; Deckelbaum *et al*., 2006; Mukherjee *et al*., 2007; Hadders-Algra, 2010). Traditional source for VLCPUFAs is marine fish such as salmon, trout, mackerel, and others. However, this source of VLCPUFAs is unsustainable due to overexploitation, environmental pollution and climatic change leading to reduced fish population. Engineering VLCPUFAs in oilseed plants and oleaginous microbes is an attractive alternative and has recently drawn much attention from lipid research community and nutraceutical industries. *Thraustochytrium* is a marine protist which can produce more than 50% DHA, and 63% VLCPUFAs in storage lipids (Meesapyodsuk & Qiu, 2016); thus, understanding the mechanism for VLCPUFA biosynthesis in this species could provide an opportunity for us to develop a new source of VLCPUFAs through reconstructing the VLCPUFA biosynthetic pathway in plants and microbes.

The aerobic pathway for DHA biosynthesis in microorganisms was first discovered by identifying a Δ4 desaturase in *Thraustochytrium* (Qiu *et al*., 2001). Different from that in mammals where DHA is synthesized by two elongations of EPA (20:5n-3), one Δ6 desaturation and one β-oxidation, biosynthesis of DHA in this species can go through one elongation of EPA and one Δ4 desaturation. This discovery opens opportunities for constructing an artificial DHA biosynthetic pathway in plants (Wu *et al*., 2005) where a number of desaturases including Δ4 desaturase and elongases are employed. As a result, DHA can be produced in transgenic plants by heterologous enzymes from microbes. Since then, tremendous progress has been made in producing DHA in oilseed crops by using this pathway (Petrie *et al*., 2012; Usher *et al*., 2015). However, engineered aerobic pathway in plants can produce undesirable intermediate fatty acids such as SDA (18:4n-3), DGLA (20:4n-3) and ETA (20:4n-3) (Qiu, 2003).

Recently, the anaerobic pathway in the biosynthesis of VLCPUFAs catalyzed by PUFA synthase has drawn attention since it produces a low level of intermediate side-fatty acids. This pathway was first discovered in *Schizochytrium* (Metz *et al*., 2001) and later reconstituted in

bacteria such as *E. coli* and *Lactococcus lactis* (Meesapyodsuk & Qiu, 2016; Amiri-Jami *et al*., 2014). Engineered microorganisms successfully produced EPA and/or DHA. Very recently, the pathway has also been engineered in *Arabidopsis* and Canola successfully using four ORFs (ORFA, B, and C) of a PUFA synthase along with a phosphopantetheinyl transferase (PPT) (Metz *et al*., 2006; Walsh *et al*., 2016). Transgenic seeds are capable of synthesizing DHA and DPA. However, both *Arabidopsis* and Canola transgenic seeds produced only 0.8% DHA over total fatty acids. The possible reason for low DHA production might be the low activity of heterologous PUFA synthase in transgenic plants. Hence, further optimization of the pathway by improving activity of PUFA synthase in oilseeds is required. PUFA synthase from *Thraustochytrium* is a large enzyme comprising more than 7000 amino acids, and some of the sequences or domains might be redundant in function or not essential for DHA biosynthesis. Thus, functional analysis of domains and sequence motifs would help us to understand the catalytic mechanism of the PUFA synthase and optimally implement this pathway in heterologous systems.

Several functional analyses of domains in PUFA synthases have recently been made by dissection and expression in microorganisms. Overexpression of key domains of a PUFA synthase could improve the yield of DHA produced. For example, introducing more ACP domains into an engineered PUFA synthase of *Shewanella oneidensis* enhanced DHA production in *E. coli* (Shohei *et al*., 2016). In addition, overexpression of pks3 gene (encoding ER and DH domain) of a PUFA synthase in *Aurantiochytrium limacinum* could enhance both DHA and total fatty acid content in this species (Liu *et al*., 2018). Therefore, our identification and characterization of key domains such as KS and DH domains would not only provide insights into the functions of these domains in PUFA synthase but also offer strategies for improving the fatty acid biosynthetic efficiency through engineering these domains. For instance, our study showed that overexpression of KS-A domain could significantly improve total fatty acid production and expression of KS-B domain could enhanced UFAs. Consequently, improved expression of KS-B by doubling the domain in the PUFA synthase might boost the DHA production. On the other hand, two tandem DH domains in subunit-C might have overlapping activity; thus, deleting one DH domain might be possible without compromising the dehydratase activity required for DHA synthesis. Our study also showed that KS-B domain could work with other components of plant type II fatty acid synthase leading to improved oil production in model plant *Arabidopsis*. This research could be extended

to oilseed plants to see if the same phenomenon would be observed. If so, this will open an opportunity for improving oil production in oilseed crops.

PUFA synthase in *Thraustochytrium* comprises three subunits each with multiple domains. Function of a protein depends on its three-dimensional structure. Therefore, crystal structure analysis of the PUFA synthase is necessary for the complete elucidation of its catalytic mechanism. However, the crystal structure of PUFA synthase is currently not available due to several reasons. (1) The PUFA synthase contains more than 7000 amino acids which is too large to express at a high level in heterologous systems. (2) Isolation of this mega-enzyme in the high degree of purity for crystallization is difficult; (3) It is challenging for the mega-enzyme to maintain the active structure during the crystallization process. However, analyzing the crystal structure of a functional domain from the PUFA synthase would be amenable. Our study showed that dissected KS and DH domains could be highly expressed as the soluble enzymes in *E.coli.* Thus, these domain proteins would be easily purified through a nickel column if a His-Tag is introduced in combination with other chromatography techniques for crystal structure analysis. Three dimensional structure information of the crystallized domains would provide insight into catalytic mechanisms of individual domains of the PUFA synthase for the biosynthesis of DHA in *Thraustochytrium*.

REFERENCES

Abe, T., Sakuradani, E., Asano, T., Kanamaru, H., & Shimizu, S. (2006). Functional characterization of Δ9 and ω9 desaturase genes in *Mortierella alpina* 1S-4 and its derivative mutants. *Applied Microbiology and Biotechnology*, *70*(6), 711–719.

Ajith, T. A, & Jayakumar, T. G. (2019). Omega-3 fatty acids in coronary heart disease: recent updates and future perspectives. *Clinical and Experimental Pharmacology and Physiology*, *46*(1), 11-18.

Ames, B. D., Nguyen, C., Bruegger, J., Smith, P., Xu, W., Ma, S., Tsai, S. C. (2012). Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proceedings of the National Academy of Sciences*. *109*(28):11144–11149.

Amiri-Jami, M., Lapointe, G., & Griffiths, M. W. (2014). Engineering of EPA/DHA omega-3 fatty acid production by *Lactococcus lactis* subsp. cremoris MG1363. *Applied Microbiology and Biotechnology*, *98*(7), 3071–3080.

Ansari, M. Z., Yadav, G., Gokhale, R. S., & Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Research*, *32*, 405–413.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., & Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*(1), 2006-2008.

Baud, S, & Lepiniec, L. (2010). Physiological and developmental regulation of seed oil production. *Progress in Lipid Research*, *49*(3), 235–249.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C., & Lepiniec, L. (2008). Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. *The Arabidopsis Book*. *6*: e0113.

Bazinet, R. P., & Layé, S. (2014). Polyunsaturated fatty acids and their metabolites in brain function and disease. *Nature Reviews Neuroscience*, *15*(12), 771–785.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., & Schwede, T. (2014). SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, *42*, 252-258.

Broekman, J. H. F. F., & Steenbakkers, J. F. (1973). Growth in high osmotic medium of an unsaturated fatty acid auxotroph of *Escherichia coli* K 12. *Journal of Bacteriology*, *116*(1), 285–289.

Broekman, J. H. F. F., & Steenbakkers, J. F. (1974). Effect of the osmotic pressure of the growth medium on fabB mutants of *Escherichia coli. Journal of Bacteriology*, *117*(3), 971–977.

Brown, A., Affleck, V., Kroon, J., & Slabas, A. (2009). Proof of function of a putative 3-hydroxyacyl-acyl carrier protein dehydratase from higher plants by mass spectrometry of product formation. *FEBS Letters*, *583*(2), 363–368.

Campbell, A. A., & Nikolau, B. J. (2011). Fatty acid elongase (FAE) systems: An investigation of genetic redundancy. (Doctoral dissertation) Retrieved from https://lib.dr.iastate.edu/etd/10279/.

Campbell, J. W., & Cronan, J. E. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annual Review of Microbiology*, *55*, 305-332.

Cao, Y. J., Yang, J. M., Xian, M., Xu, X., & Liu, W. (2010). Increasing unsaturated fatty acid contents in *Escherichia coli* by coexpression of three different genes. *Applied Microbiology and Biotechnology*, *87*(1), 271–280.

Chirala, S. S., & Wakil, S. J. (2004). Structure and function of animal fatty acid synthase. *Lipids*, *39*(5), 1045–1053.

Chung, C. T., Niemela, S. L., & Miller, R. H. (1989). One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution (recombinant DNA). *Biochemistry*, *86*, 2172–2175.

Clough, S. J., & Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, *16*(6), 735-43.

Conway, K. R., & Boddy, C. N. (2013). ClusterMine360: a database of microbial PKS/NRPS

biosynthesis. N*ucleic Acids Research*, *41*, 402-407.

Cox, R. J., & Simpson, T. J. (2009). Fungal type I polyketide synthases. *Methods in Enzymology*, *459*, 49-78.

Crawford, M. A., Costeloe, K., Ghebremeskel, K., Phylactos, A., Skirvin, L., & Stacey, F. (1997). Are deficits of arachidonic and docosahexaenoic acids responsible for the neural and vascular complications of preterm babies? *American Journal of Clinical Nutrition*. *66*(4):1032S–1041S.

Cui, W., Liang, Y., Tian, W., Ji, M. J., & Ma, X. F. (2015). Regulating effect of β-ketoacyl synthase domain of fatty acid synthase on fatty acyl chain length in *de novo* fatty acid synthesis. *Biochimica et Biophysica Acta-Molecular and Cell Biology of Lipids*, *1861*(3), 149–155.

Daiello, L. A., Wellenius, G., Ott, B. R., & Buka, S. L. (2015). Role of supplemental docosahexaenoic acid (DHA) for cognition in *Alzheimer's* disease and mild cognitive impairment: A systematic review and meta-analysis of randomized controlled trials. *Alzheimer's & Dementia*, *11*(7), 611.

Davies, C., Heath, R. J., White, S. W., & Rock, C. O. (2000). The 1.8 Å crystal structure and active-site architecture of β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. *Structure*, *8*(2), 185-195.

Deckelbaum, R. J., Worgall, T. S., & Seo, T. (2006). n-3 fatty acids and gene expression. *The American Journal of Clinical Nutrition*, *83*(3), 1520–1525.

Diau, G. Y., Hsieh, A. T., Sarkadi-Nagy, E., Wijendran, V., Nathanielsz, P. W., & Brenna, J. T. (2005). The influence of long chain polyunsaturate supplementation on docosahexaenoic acid and arachidonic acid in baboon neonate central nervous system. *BMC Medicine*, *3*, 11.

Diep, Q. N., Touyz, R. M., & Schiffrin, E. L. (2000). Docosahexaenoic acid, a peroxisome proliferator-activated receptor-α ligand, induces apoptosis in vascular smooth muscle cells by stimulation of p38: mitogen-activated protein kinase. *Hypertension*. *36*(5):851–855.

Dutta, S., Whicher, J. R., Hansen, D., Hale, W., Chemler, J., Congdon, G. R., & Skiniotis, G. (2014). Structure of a modular polyketide synthase. *Nature*, *510*(7506), 512–517.

Edwards, K., Johnstone, C., & Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, *19*(6): 1349.

Ellerström, M., Stålberg, K., Ezcurra, I., & Rask, L. (1996). Functional dissection of a napin gene promoter: identification of promoter elements required for embryo and endosperm-specific transcription. *Plant Molecular Biology*, *32*, 1019–1027.

Feng, Y. J., & Cronan, J. E. (2009). *Escherichia coli* unsaturated fatty acid synthesis: complex transcription of the fabA gene and *in vivo* identification of the essential reaction catalyzed by FabB. *Journal of Biological Chemistry*, *284*(43), 29526–29535.

Finzel, K., Nguyen, C., Jackson, D. R., Gupta, A., Tsai, S. C., & Burkart, M. D. (2015). Probing the substrate specificity and protein-protein interactions of the *E. coli* fatty acid dehydratase, FabA. *Chemistry and Biology*, *22*(11), 1453–1460.

Freedman, S. D., Blanco, P. G., Zaman, M. M., Shea, J. C., Ollero, M., Hopper, I. K., Weed, D. A., Gelrud, A., Regan, M. M., Laposata, M., Alvarez, J. G., O'Sullivan, B. P. (2004). Association of Cystic Fibrosis with Abnormalities in fatty acid metabolism. *New England Journal of Medicine*. *350*(6):560–569.

Giusto, N. M., Pasquaré, S. J., Salvador, G. A., Castagnet, P. I., Roque, M. E., & Ilincheta De Boschero, M. G. (2000). Lipid metabolism in vertebrate retinal rod outer segments. *Progress in Lipid Research*. *39*(4):315–391.

Gong, Y. M., Wan, X., Jiang, M. L., Hu, C. J., Hu, H. H., & Huang, F. H. (2014). Metabolic engineering of microorganisms to produce omega-3 very long-chain polyunsaturated fatty acids. *Progress in Lipid Research*, *56*(1), 19–35.

Grogan, D. W., & Cronan, J. E. (1997). Cyclopropane ring formation in membrane lipids of bacteria. *Microbiology and Molecular Biology Reviews*, *61*(4), 429–441.

Gu, Y. N., He, L., Zhao, C. J., Wang, F., Yan, B. W., Gao, Y. Q., Li, J. T., Yang, K. J. & Xu, J. Y. (2017). Biochemical and transcriptional regulation of membrane lipid metabolism in maize leaves under low temperature. *Frontiers in Plant Science*, *8*, 1–13.

Hadders-Algra, M. (2010). Effect of long-chain polyunsaturated fatty acid supplementation on neurodevelopmental outcome in full-term infants. *Nutrients*, *2*, 790–804.

Hayashi, S., Satoh, Y., Ujihara, T., Takata, Y., & Dairi, T. Enhanced production of polyunsaturated fatty acids by enzyme engineering of tandem acyl carrier proteins. *Scientific Reports*, *18*;6:35441.

Heath, R. J., & Rock, C. O. (1996a). Inhibition of β-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in *Escherichia coli*. *Journal of Biological Chemistry*, *271*(18), 996-1000.

Heath, R. J., & Rock, C. O. (1996b). Roles of the FabA and FabZ β-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *Journal of Biological Chemistry*, *271*(44), 27795–27801.

Herskowitz, I. (1987). Functional inactivation of genes by dominant negative mutations. *Nature*, *329*(6136), 219–222.

Hitz, W. D., Carlson, T. J., Booth, J. R., Kinney, A. J., Stecca, K. L., & Yadav, N. S. (1994). Cloning of a higher-plant plastid omega-6 fatty acid desaturase cDNA and its expression in a *cyanobacterium*. *Plant Physiology*, *105*(2), 635–641.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, *77*(1), 51–59.

Horrocks, L., & Yeo, Y. K. (1999). Health benefits of docosahexaenoic acid (DHA). *Pharmacological Research*, *40*(3), 211–225.

Jako, C., Kumar, A., Wei, Y., Zou, J., Barton, D. L., Giblin, E. M., Covello, P. S., & Taylor, D. C. (2001). Seed-specific over-expression of an *Arabidopsis* cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight. *Plant Physiology*, *126*(2), 861–874.

Janine, G. B., Andrew, C., Carl, A. M., Zhang, X. J. & Peter, J. T. (2014). Substrate recognition by β-ketoacyl-ACP synthases. *ACS Biochemistry*, *71*(11), 3831–3840.

Janßen, H. J., & Steinbüchel, A. (2014). Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels*, *7*(7), 1–26.

Jeon, E., Lee. S., Lee, S., Han, S. O., Yoon, Y. J., & Lee, J. (2012). Improved production of long-chain fatty acid in *Escherichia coli* by an engineering elongation cycle during fatty acid synthesis (FAS) through genetic manipulation. *Journal of Microbiology and Biotechnology*, *22*(7), 990–999.

Ji, X. J., Mo, K. Q., Ren, L. J., Li, G. L., Huang, J. Z., & Huang, H. (2015). Genome sequence of *Schizochytrium* sp. CCTCC M209059, an effective producer of docosahexaenoic acid-rich lipids. *Genome Announcements*. *3*(4). pii: e00819–15.

John, A., Umashankar, V., Samdani, A., Sangeetha, M., Krishnakumar, S., & Deepa, P. R. (2016). In silico structure prediction of human fatty acid synthase-dehydratase: a plausible model for understanding active site interactions. *Bioinformatics and Biology Insights*, *10*, 143–154.

John, E. C. & Jacob, T. (2009). Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods in Enzymology*, *459*, 395–433.

Josefsson, L. G., Lenman, M., Ericson, M. L., & Rask, L. (1987). Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*. *Journal of Biological Chemistry*, *262*(25), 12196-12201.

Jump, D.B. (2009). Mammalian fatty acid elongases. *Methods in Molecular Biology*, *579*, 375–389.

Kajikawa, M., Yamato, K. T., Kohzu, Y., Nojiri, M., Sakuradani, E., Shimizu S., Sakai, Y., Fukuzawa, H. & Ohyama, K.(2004). Isolation and characterization of delta(6)-desaturase, an ELO-like enzyme and delta(5)-desaturase from the liverwort Marchantia polymorpha and production of arachidonic and eicosapentaenoic acids in the methylotrophic yeast *Pichia pastoris*. *Plant Molecular Biology*, *54*(3), 335–352.

Kautharapu, K. B., & Jarboe, L. R. (2012). Genome sequence of the psychrophilic deep-sea bacterium *moritella marina* MP-1 (ATCC 15381). *Journal of Bacteriology*. *194*(22):6296–6297.

Kavanagh, K. L., Jörnvall, H., Persson, B., & Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: functional and

structural diversity within a family of metabolic and regulatory enzymes. *Cellular and Molecular Life Sciences*, *65*(24), 3895–3906.

Keatinge-clay, A. (2008). Crystal structure of the erythromycin polyketide synthase Dehydratase. *Journal of Molecular Biology*, *384*(4), 941–953.

Keatinge-Clay, A. T. (2012). The structures of type I polyketide synthases. *Natural Product Reports*, *29*(10), 1050.

Kelley, L. A., & Sternberg, M. J .E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, *4*(3), 363–371.

Khosla, C., Gokhale, R. S., Jacobsen, J. R., & Cane, D. E. (2002). Tolerance and specificity of polyketide synthases. *Annual Review of Biochemistry*, *68*, 219-53.

Kimber, M. S., Martin, F., Lu, Y., Houston, S., Vedadi, M., Dharamsi, A., & Rock, C. O. (2004). The structure of (3R)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, *279*(50), 52593–52602.

Kjell, S., Mats, E., Staffan, S., Ines, E., & Paul, W., & Lars, R. (2016). Heterologous and homologous transgenic expression directed by a 2S seed storage promoter of *Brassica napus*. *CEUR Workshop Proceedings*, *1621*(1998), 36–43.

Kulmacz, R. J. (1998). Cellular regulation of prostaglandin H synthase catalysis. *FEBS Letters*. *430*(3), 154-157.

Kumar, A., Sharma, A., & Upadhyaya, K. (2016). Vegetable oil: nutritional and industrial perspective. *Current Genomics*, *17*(3), 230–240.

Lauro, F. M., Stratton, T. K., Chastain, R. A., Ferriera, S., Johnson, J., Goldberg, S. M., Yayanos, A. A., Bartlett, D. H. (2013). Complete genome sequence of the deep-sea bacterium *Psychromonas* Strain CNPT3. *Genome Announcements*. *1*(3). pii: e00304–13.

Lee, D. W., Lee, S., Lee, G. J., Lee, K. H., Kim, S., Cheong, G. W., & Hwang, I. (2006). Functional characterization of sequence motifs in the transit peptide of *Arabidopsis* small subunit of rubisco. *Plant Physiology*, *140*(2), 466–483.

Lee, S., Lee, S., Yoon, Y. J., & Lee, J. (2013). Enhancement of long-chain fatty acid production

in *Escherichia coli* by coexpressing genes, including fabF, involved in the elongation cycle of fatty acid biosynthesis. *Applied Biochemistry and Biotechnology*, *169*(2), 462–476.

Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., & Smith, J. L. (1996). Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure*, *4*(3), 253–264.

Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M., & Smith, J. L. (1996). Structure of a dehydratase–isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure*, *4*(3), 253–264.

Lemieux, B., Miquel, M., Somerville, C., & Browse, J. (1990). Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. *Theoretical and Applied Genetics*. *80*(2), 234-240.

Li, D., Jin, C. Y., Duan, S. W., Zhu, Y. N., Qi, S. H., Liu, K. G., & Chen, M. X. (2017). MYB89 transcription factor represses seed oil accumulation. *Plant Physiology*, *173*(2), 1211–1225.

Li, Y., Dodge, G. J., Fiers, W. D., Fecik, R. A., Smith, J. L., & Courtney, C. (2015). Functional characterization of a dehydratase domain from the pikromycin polyketide synthase. *Journal of the American Chemical Society*, *37*(22):7003–6.

Liu, J., Hua, W., Zhan, G., Wei, F., Wang, X., Liu, G., & Wang, H. (2010). Increasing seed mass and oil content in transgenic *Arabidopsis* by the overexpression of wri1-like gene from *Brassica napus*. *Plant Physiology and Biochemistry*, *48*(1), 9–15.

Liu, Z., Zang, X. N., Cao, X. X., Wang, Z.D., Liu, C., Sun, D. G, Guo, Y. L., Zhang, F., Yang Q., Hou, G., & Peng, C. H. (2018) Cloning of the pks3 gene of *Aurantiochytrium limacinum* and functional study of the 3-ketoacyl-ACP reductase and dehydratase enzyme domains. *PLoS One*, *13*(12), e0208853.

Lomakin, I. B., Xiong, Y., & Steitz, T. A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell*, *129*(2), 319–332.

Lu, Y. J, Zhang, Y. M., & Rock, C. O. (2004). Product diversity and regulation of type II fatty acid synthases. *Biochemistry and Cell Biology*, *82*(1), 145–155.

Ma, X. L., & Liu, Y. G. (2016). CRISPR/Cas9-based multiplex genome editing in monocot and dicot plants. *Current Protocols in Molecular Biology*, *115*(1), 1–21.

Magnuson, K., Jackowski, S., Rock, C. O., & Cronan, J. E. (1993). Regulation of fatty acid synthesis. *Biochemical Society Transactions*, *57*(3), 522.

Maier, T., Jenni, S., & Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science*. *311*(5765):1258–62.

Manan, S., Chen, B. B., She, G. B., Wan, X. C., & Zhao, J. (2017). Transport and transcriptional regulation of oil production in plants. *Critical Reviews in Biotechnology*, *37*(5), 641–655.

Masliah, E., Mallory, M., Hansen, L., Alford, M., DeTeresa, R. & Terry, R. (1993). An antibody against phosphorylated neurofilaments identifies a subset of damaged association axons in *Alzheimer's* disease. *The American Journal of Pathology*, *142*(3):871-882.

McNamara R. K. (2013). Deciphering the role of docosahexaenoic acid in brain maturation and pathology with magnetic resonance imaging. *Prostaglandins, leukotrienes, and essential fatty acids*, *88*(1), 33–42.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, *39*, 339-346.

Meesapyodsuk, D., & Qiu, X. (2014). Structure determinants for the substrate specificity of acyl-CoA Δ9 desaturases from a marine copepod. *ACS Chemical Biology*, *9*(4), 922–934.

Meesapyodsuk, D., & Qiu, X. (2016). Biosynthetic mechanism of very long chain polyunsaturated fatty acids in *Thraustochytrium* sp. 26185. *Journal of Lipid Research*, *57*(10), 1854-1864 .

Meesapyodsuk, D., & Qiu. X. (2012). The front-end desaturase: Structure, function, evolution and biotechnological use. *Lipids*, *47*, 227–237.

Meesapyodsuk, D., Ye, S. J., Chen, Y., Chen, Y., Chapman, R. G., & Qiu, X. (2018). An engineered oilseed crop produces oil enriched in two very long chain polyunsaturated fatty

acids with potential health-promoting properties. *Metabolic Engineering*, *49*, 192–200.

Metz, J. G., Roessler, P., Facciotti, D., Levering, C., Dittrich, F., Lassner, M., Browse, J. (2001). Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science*, *293*, 290–292.

Moretti, L., Cristofori, I., Weaver, S. M., Chau, A., Portelli, J. N., & Grafman, J. (2012). Cognitive decline in older adults with a history of traumatic brain injury. *The Lancet Neurology*. *11*(12), 1103-1112.

Mukherjee, P. K., Chawla, A., Loayza, M. S. & Bazan, N. G. (2007). Docosanoids are multifunctional regulators of neural cell integrity and fate: significance in aging and disease. *Prostaglandins Leukot Essent Fatty Acids*, *77*(5–6), 233–238.

Neschen, S., Moore, I., Regittnig, W., Yu, C. L., Wang, Y., Pypaert, M., & Shulman, G. I. (2002). Contrasting effects of fish oil and safflower oil on hepatic peroxisomal and tissue lipid content. *American Journal of Physiology-Endocrinology and Metabolism*. *282*(2), 395-401.

Newell, M., Baker, K., Postovit, L. M., & Field, C. J. (2017). A critical review on the effect of docosahexaenoic acid (DHA) on cancer cell cycle progression. International Journal of *Molecular Sciences*, *18*(8), 1784.

Niemoller, T. D., & Bazan, N. G. (2010). Docosahexaenoic acid neurolipidomics. *Prostaglandins and Other Lipid Mediators*, *91*(3-4), 85-89.

Okuyama, H., Orikasa, Y., Nishida, T., Watanabe, K., & Morita, N. (2007). Bacterial genes responsible for the biosynthesis of eicosapentaenoic and docosahexaenoic acids and their heterologous expression. *Applied and Environmental Microbiology*, *73*(3), 665–670.

Olsen, J. G., Rasmussen, A. V., von Wettstein-Knowles, P. & Henriksen, A. (2004). Structure of the mitochondrial beta-ketoacyl-[acyl carrier protein] synthase from *Arabidopsis* and its role in fatty acid synthesis. *FEBS Letters*, *577*(1–2), 170–174.

Organisciak, D. T., Darrow, R. M., Jiang, Y. L., & Blanks, J. C. (1996). Retinal light damage in rats with altered levels of rod outer segment docosahexaenoate. *Investigative Ophthalmology and Visual Science*, *37*(11), 2243–2257.

Oyola-Robles, D., Gay, D. C., Trujillo, U., Sánchez-Parés, J. M., Bermúdez, M. L., Rivera-Díaz, M., & Baerga-Ortiz, A. (2013). Identification of novel protein domains required for the expression of an active dehydratase fragment from a polyunsaturated fatty acid synthase. *Protein Science*, *22*(7), 954–963.

Parsons, J. B., & Rock, C. O. (2013). Progress in lipid research bacterial lipids : metabolism and membrane homeostasis. *Progress in Lipid Research*, *52*(3), 249–276.

Pasta, S., Witkowski, A., Joshi, A. K., & Smith, S. (2007). Catalytic residues are shared between two pseudosubunits of the dehydratase domain of the animal fatty acid synthase. *Chemistry and Biology*, *14*(12), 1377–1385.

Petrie, J. R., Shrestha, P., Zhou, X. R., Mansour, M. P., Liu, Q., Belide, S., & Singh, S. P. (2012). Metabolic engineering plant seeds with fish oil-like levels of DHA. *PLoS ONE*, *7*(11):e49165.

Qi, B. X., Fraser, T., Mugford, S., Dobson, G., Sayanova, O., Butler, J., Lazarus, C. M. (2004). Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. *Nature Biotechnology*, *22*(6), 739–745.

Qiu X. (2003). Biosynthesis of docosahexaenoic acid (DHA, 22:6-4, 7, 10, 13, 16, 19): two distinct pathways. *Prostaglandins Leukotrienes and Essential Fatty Acids*, *68*(2), 181–186.

Qiu, X., Hong, H., & MacKenzie, S. L. (2001). Identification of a Δ4 fatty acid desaturase from *Thraustochytrium sp*. involved in the biosynthesis of docosahexaenoic acid by heterologous expression in *Saccharomyces cerevisiae* and *Brassica juncea*. *Journal of Biological Chemistry*, *276*(34), 31561–31566.

Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., & Walsh, C. T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carder protein domains in peptide synthetases. *Biochemistry*. *37*(6):1585–95.

Rawsthorne, S. (2002). Carbon flux and fatty acid synthesis in plants. *Progress in Lipid Research*, *41*, 182–196.

Reich, E. E., Markesbery, W. R., Roberts, L. J., Swift, L. L., Morrow, J. D., & Montine, T. J. (2001). Brain regional quantification of F-ring and D-/E-ring isoprostanes and

neuroprostanes in *Alzheimer's* disease. *The American Journal of Pathology*, *158*(1), 293–297.

Ridi, R. E., Aboueldahab, M., Tallima, H., Salah, M., Mahana, N., Fawzi, S., & Fahmy, O. M. (2010). *In vitro* and *in vivo* activities of arachidonic acid against *Schistosoma mansoni* and *Schistosoma haematobium*. *Antimicrob Agents Chemother*, *54*(8), 3383–3389.

Rink, C., & Khanna, S. (2010). Significance of brain tissue oxygenation and the arachidonic acid cascade in stroke. *Antioxidants & Redox Signaling*, *14*(10), 1889–1903.

Robinson, L. E., & Mazurak, V. C. (2013). n-3 polyunsaturated fatty acids: Relationship to inflammation in healthy adults and adults exhibiting features of metabolic syndrome. *Lipids*, *48*(4), 319–332.

Rock, C. O. (2008). Fatty acid and phospholipid metabolism in prokaryotes. *Biochemistry of Lipids, Lipoproteins and Membranes (Fifth Edition)*, 59–96.

Rock, C. O., Tsay, J. T., Heath, R., & Jackowski, S. (1996). Increased unsaturated fatty acid production associated with a suppressor of the *fabA*6(Ts) mutation in *Escherichia coli*. *Journal of Bacteriology*, *178*(18), 5382–5387.

Roesler, K., Shintani, D., Savage, L., Boddupalli, S., & Ohlrogge, J. (1997). Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. *Plant Physiology*, *113*(1), 75–81.

Rosenfeld, I. S., D'Agnolo, G., & Vagelos, P. R. (1973). Synthesis of unsaturated fatty acids and the lesion in *fab B* mutants. *Journal of Biological Chemistry*, *248*(7), 2452–2460.

Rouhiainen, L., Vakkilainen, T., Siemer, B. L., Buikema, W., Haselkorn, R., & Sivonen, K. (2004). Genes coding for hepatotoxic heptapeptides (Microcystins) in the *Cyanobacterium Anabaena* Strain 90. *Applied and Environmental Microbiology*, *70*(2), 686–692.

Ruiz-Gutierrez, V., Perez-Espinosa, A., Vazquez, C. M., & Santa-Maria, C. (1999). Effects of dietary fats (fish, olive and high-oleic-acid sunflower oils) on lipid composition and antioxidant enzymes in rat liver. *British Journal of Nutrition*, *82*(3), 233-41.

Ruiz-López, N., Sayanova, O., Napier, J. A., & Haslam, R. P. (2012). Metabolic engineering of

the omega-3 long chain polyunsaturated fatty acid biosynthetic pathway into transgenic plants. *Journal of Experimental Botany*, *63*(7), 2397–2410.

Rustan, A. C., & Drevon, C. A. (2006). Fatty Acids: Structures and properties. *In Encyclopedia of Life Sciences*, 1–7.

Rustan, A. C., Nossen, J. O., Christiansen, E. N., & Drevon, C. A. (1988). Eicosapentaenoic acid reduces hepatic synthesis and secretion of triacylglycerol by decreasing the activity of acyl-coenzyme A:1,2-diacylglycerol acyltransferase. *Journal of Lipid Research*, *29*(11), 1417-1426.

Savadi, S., Naresh, V., Kumar, V., & Bhat, S. R. (2015). Seed specific overexpression of *Arabidopsis* DGAT1 in india mustard (*Brassica juncea*) increases seed oil and seed weight. *Botany*, *184*, 1–27.

Schiml, S., & Puchta, H. (2016). Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods*, *12*(1), 1–9.

Schweizer, E., & Hofmann, J. (2004). Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiology and Molecular Biology Reviews* , *68*(3), 501–517.

Scorletti, E., & Byrne, C. D. (2013). Omega-3 fatty acids, hepatic lipid metabolism, and nonalcoholic fatty liver disease. *Annual Review of Nutrition*, *33*, 231–248.

Semple, K. S., & Silbert, D. F. (1975). Mapping of the fabD locus for fatty acid biosynthesis in *Escherichia coli*. *Journal of Bacteriology*, *121*(3), 1036–1046.

Serhan, C. N. (2005). Novel eicosanoid and docosanoid mediators: resolvins, docosatrienes, and neuroprotectins. *Current Opinion in Clinical Nutrition and Metabolic Care*. *8*(2):115-21.

Shanab, S. M. M., Hafez, R. M., & Fouad, A. S. (2018). A review on algae and plants as potential source of arachidonic acid. *Journal of Advanced Research*, *11*, 3–13.

Shanklin, J., & Cahoon, E. B. (1998). Desaturation and related modifications of fatty acids. *Annual Review of Plant Physiology and Plant Molecular Biology*, *49*, 611–641.

Sheppard, D. Dominant negative mutants: Tools for the study of protein function. *American*

Journal of Respiratory Cell and Molecular Biology, *11*(1), 1-6

Silbert, D. F., & Vagelos, P. R. (1967). Fatty acid mutant of *E. coli* lacking a beta-hydroxydecanoyl thioester dehydrase. *Proceedings of the National Academy of Sciences of the United States of America*, *58*(4), 1579–1586.

Smith, S., Witkowski, A., & Joshi, A. K. (2003). Structural and functional organization of the animal fatty acid synthase. *Progress in Lipid Research*, *42*, 289–317.

Soh, J. Y. K., Russell, C. W., Fenlon, S. N., & Chen, S. L. (2018). Complete genome sequence of *Photobacterium leiognathi* strain JS01. *Genome Announcements*. *6*(1): e01396–17.

Spener, F., Dennis, E., Subramaniam, S., Van Nieuwenhze, M., Glass, C., Brown, H., & Murphy, R. (2013). A comprehensive classification system for lipids. *In Handbook of Biochemistry and Molecular Biology (Fourth Edition),* 165–167.

Sprague, M., Betancor, M. B., & Tocher, D. R. (2017). Microbial and genetically engineered oils as replacements for fish oil in aquaculture feeds. *Biotechnology letters*, *39*(11), 1599–1609.

Sprecher, H. (2002). The roles of anabolic and catabolic reactions in the synthesis and recycling of polyunsaturated fatty acids. *Prostaglandins Leukot Essent Fatty Acids*, *67*(2–3), 79–83.

Tallima, H., & El, R. (2018). Arachidonic acid : physiological roles and potential health benefits – a review. *Journal of Advanced Research*, *11*, 33–41.

Tang, Y., Tsai, S. C., & Khosla, C. (2003). Polyketide chain length control by chain length factor. *Journal of the American Chemical Society*, *125*(42), 12708–12709.

Tavares, S., Grotkjær, T., Obsen, T., Haslam, R. P., Napier, J. A., & Gunnarsson, N. (2011). Metabolic engineering of *Saccharomyces cerevisiae* for production of eicosapentaenoic acid, using a novel Δ5-Desaturase from *Paramecium tetraurelia*. *Applied and Environmental Microbiology*, *77*(5), 1854–1861.

To, A., Joubes, J., Barthole, G., Lecureuil, A., Scagnelli, A., Jasinski, S., & Baud, S. (2012). WRINKLED transcription factors orchestrate tissue-specific regulation of fatty acid biosynthesis in *Arabidopsis*. *The Plant Cell*, *24*(12), 5007–5023.

Uauy, R., Mena, P., & Rojas, C. (2000). Essential fatty acids in early life: structural and

functional role. *The Proceedings of the Nutrition Society*, *59*(1), 3–15.

Ujihara, T., Nagano, M., Wada, H., & Mitsuhashi, S. (2014). Identification of a novel type of polyunsaturated fatty acid synthase involved in arachidonic acid biosynthesis. *FEBS Letters*, *588*(21), 4032–4036.

Usher, S., Haslam, R. P, Ruiz-Lopez, N., Sayanova, O., & Napier, J. A. (2015). Field trial evaluation of the accumulation of omega-3 long chain polyunsaturated fatty acids in transgenic *Camelina sativa*: Making fish oil substitutes in plants. *Metabolic Engineering Communications*, *9*(2), 93–98.

Vandongen, R., Mori, T. A., Burke, V., Beilin, L. J., Morris, J., & Ritchie, J. (1993). Effects on blood pressure of ω3 fats in subjects at increased risk of cardiovascular disease. *Hypertension*, *22*(3), 371–379.

Vigeolas, H., Waldeck, P., Zank, T., & Geigenberger, P. (2007). Increasing seed oil content in oil-seed rape (*Brassica napus L.*) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnology Journal*, *5*(3), 431–441.

Vrinten, P., Mavraganis, I., Qiu, X. & Senger, T. (2013). Biosynthesis of long chain polyunsaturated fatty acids in the marine ichthyosporean *sphaeroforma arctica*. *Lipids*, *48*(3), 263–274.

Walsh, T. A., Bevan, S. A., Gachotte, D. J., Larsen, C. M., Moskal, W. A., Merlo, P. A. O., & Metz, J. G. (2016). Canola engineered with a microalgal polyketide synthase-like system produces oil enriched in docosahexaenoic acid. *Nature Biotechnology*, *34*(8), 852–858.

Wang, H., Wang, J., Jiang, J., Chen, S., Guan, Z., Liao, Y., & Chen, F. (2014). Reference genes for normalizing transcription in diploid and tetraploid *Arabidopsis*. *Scientific Reports*, *4*, 1–8.

Wang, W. C., Zhu, J. L., Lyu, F., Panigrahy, D., Ferrara, K. W., Hammock, B., & Zhang, G. D. (2014). Omega-3 polyunsaturated fatty acids-derived lipid metabolites on angiogenesis, inflammation and cancer. *Prostaglandins and Other Lipid Mediators*, *113–115*, 13–20.

Wang, Y., & Ma, S. T. (2013). Recent advances in inhibitors of bacterial fatty acid synthesis

type II (FASII) system enzymes as potential antibacterial agents. *ChemMedChem*, *8*(10), 1589–1608.

White, S. W., Zheng, J., Zhang, Y. M., & Rock, C. O. (2005). the Structural biology of type II fatty acid biosynthesis. *Annual Review of Biochemistry*, *74*(1), 791–831.

Witkowski, A., Joshi, A. K., & Smith, S. (2002). Mechanism of the β-ketoacyl synthase reaction catalyzed by the animal fatty acid synthase. *Biochemistry*, *41*(35), 10877–10887.

Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., & Qiu, X. (2005). Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nature Biotechnology*, *23*(8), 1013-1017.

Wu, G. Z., & Xue, H. W. (2010). *Arabidopsis* β-ketoacyl-[acyl carrier protein] synthase I is crucial for fatty acid synthesis and plays a role in chloroplast division and embryo development. *The Plant Cell*, *22*(11), 3726–3744.

Wu, J. H. Y., Micha, R., Imamura, F., Pan, A., Biggs, M. L., Ajaz, O., & Mozaffarian, D. (2012). Omega-3 fatty acids and incident type 2 diabetes: A systematic review and meta-analysis. *British Journal of Nutrition*, *107* (2), S214-227.

Xie, X., Ma, X. L., Zhu, Q. L., Zeng, D., Li, G., & Liu, Y. G. (2017). CRISPR-GE: A convenient software toolkit for CRISPR-based genome editing. *Molecular Plant*, *10*(9), 1246–1249.

Xie, X., Meesapyodsuk, D. & Qiu, X. (2018). Functional analysis of the dehydratase domains of a PUFA synthase from *Thraustochytrium* in *Escherichia coli*, *Applied Microbiology and Biotechnology*, *102*(2):847-856.

Xie, X., Meesapyodsuk, D., & Qiu, X. (2017). Ketoacyl synthase domains of a PUFA synthase in *Thraustochytrium* can effectively function as stand-alone enzymes in *Escherichia coli*. *Applied and Environmental Microbiology*, *83*(9) 1–13.

Xu, W., Qiao, K. J., & Tang, Y. (2013). Structural analysis of protein-protein interactions in type I polyketide synthases. *Critical Reviews in Biochemistry and Molecular Biology*, *48*(2), 98–122.

Yadav, N. S., Wierzbicki, A., Aegerter, M., Caster, C. S., Pérez-Grau, L., Kinney, A. J., &

Stecca, K. L. (1993). Cloning of higher plant omega-3 fatty acid desaturases. *Plant Physiology*, *103*(2), 467–476.

Yamazaki, Y., & Sugawara, H. (2009). National bioResource project information center. *Experimental Animals Japanese Association for Laboratory Animal Science*, *58*(2), 75–84.

Ye, C., Qiao, W. H., Yu, X. B., Ji, X. J., Huang, H., Collier, J. L., & Liu, L. M. (2015). Reconstruction and analysis of the genome-scale metabolic model of *schizochytrium limacinum* SR21 for docosahexaenoic acid production. *BMC Genomics*, *16*(1), 799.

Yin, Y., Gokhale, R., Khosla, C., & Cane, D. E. (2001). Erythromycin biosynthesis. The 4-pro-S hydride of NADPH is utilized for ketoreduction by both module 5 and module 6 of the 6-deoxyerythronolide B synthase. *Bioorganic and Medicinal Chemistry Letters*. *11*(12):1477–1479.

Yoshida, K., Hashimoto, M., Hori, R., Adachi, T., Okuyama, H., Orikasa, Y., & Morita, N. (2016). Bacterial long-chain polyunsaturated fatty acids: Their biosynthetic genes, functions, and practical use. *Marine Drugs*, *14*(5), 94.

Zhang, L., Joshi, A. K., Hofmann, J., Schweizer, E., & Smith, S. (2005). Cloning, expression, and characterization of the human mitochondrial β-ketoacyl synthase: complementation of the yeast cem1 knock-out strain. *Journal of Biological Chemistry*, *280*(13):12422-12429.

Zhang, S., Sakuradani, E., Ito, K., & Shimizu, S. (2007). Identification of a novel bifunctional Δ12/Δ15 fatty acid desaturase from a basidiomycete, *Coprinus cinereus* TD#822-2. *FEBS Letters*, *581*(2), 315–319.

Zhang, Y., Yu, L., Yung, K. F., Leung, D. Y., Sun, F., & Lim, B. L. (2012). Over-expression of *At*PAP2 in *Camelina sativa* leads to faster plant growth and higher seed yield. *Biotechnology for Biofuels*, *5*, 1–10.

Zhao, X. M., Meesapyodsuk, D., Qu, C. M. & Qiu, X. (2016). Genomic analysis of genes involved in the biosynthesis of very long chain polyunsaturated fatty acids in *Thraustochytrium sp*. 26185. *Lipids*, *51*(9), 1065–1075.