Structural features and functional residues important for the activity of an unusual membrane bound O-acyltransferase

by

Tam Nguyen Thu Tran

B.S., The University of Science, Ho Chi Minh City, Vietnam, 2006
M.S., Chonnam National University, South Korea, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Abstract

The membrane bound O-acyltransferase (MBOAT) family contains multi-pass membrane proteins that add fatty acids to different compounds. Despite their importance in economic activity and human health, little is known about the localization of the active site and regions important for determining substrate specificity of MBOATs. *Euonymus alatus* diacylglycerol acetyltransferase (*Ea*DAcT) is the only known MBOAT enzyme that exhibits a high preference for acetyl-CoA, the shortest possible acyl-CoA. *Ea*DAcT catalyzes the transfer of the acetate group from acetyl-CoA to the *sn*-3 position of diacylglycerol to form 3-acetyl-1,2-diacyl-*sn*-glycerol. Our goal was to investigate the structural features and the amino acid residues that define substrate specificity of *Ea*DAcT to provide insights into the mechanism by which MBOAT family controls substrate selection. By mapping the membrane topology of *Ea*DAcT we obtained the first experimentally determined topology model for a plant MBOAT. The *Ea*DAcT model contains four transmembrane domains with both the N- and C- termini oriented toward the endoplasmic reticulum lumen. The MBOAT signature region including the putative active site His-257 of the protein is embedded in the third transmembrane domain close to the interface between the membrane and the cytoplasm. In order to identify amino acid residues important for acetyltransferase activity, we isolated and characterized orthologs of *Ea*DAcT from other acetyl-TAG producing plants. Among them, the acetyltransferase from *Euonymus fortunei* possessed the highest activity in vivo and in vitro. Mutagenesis of conserved residues of DAcTs revealed that Ser-253, His-257 and Asp-258 are essential for enzyme activity of *Ea*DAcT, suggesting their involvement in the enzyme catalysis. Alteration of residues unique to acetyltransferases did not alter the acyl donor specificity of *Ea*DAcT, implying that multiple amino acids are important for substrate recognition. Together, this work identifies the structural features of *Ea*DAcT and offers an initial view of the amino acids important for activity of the enzyme.

Structural features and functional residues important for the activity of an unusual membrane bound O-acyltransferase

by

Tam Nguyen Thu Tran

B.S., The University of Science, Ho Chi Minh City, Vietnam, 2006
M.S., Chonnam National University, South Korea, 2009

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biophysics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Timothy P Durrett. PhD

Copyright

Abstract

The membrane bound O-acyltransferase (MBOAT) family contains multi-pass membrane proteins that add fatty acids to different compounds. Despite their importance in economic activity and human health, little is known about the localization of the active site and regions important for determining substrate specificity of MBOATs. *Euonymus alatus* diacylglycerol acetyltransferase (*Ea*DAcT) is the only known MBOAT enzyme that exhibits a high preference for acetyl-CoA, the shortest possible acyl-CoA. *Ea*DAcT catalyzes the transfer of the acetate group from acetyl-CoA to the *sn*-3 position of diacylglycerol to form 3-acetyl-1,2-diacyl-*sn*-glycerol. Our goal was to investigate the structural features and the amino acid residues that define substrate specificity of *Ea*DAcT to provide insights into the mechanism by which MBOAT family controls substrate selection. By mapping the membrane topology of *Ea*DAcT we obtained the first experimentally determined topology model for a plant MBOAT. The *Ea*DAcT model contains four transmembrane domains with both the N- and C- termini oriented toward the endoplasmic reticulum lumen. The MBOAT signature region including the putative active site His-257 of the protein is embedded in the third transmembrane domain close to the interface between the membrane and the cytoplasm. In order to identify amino acid residues important for acetyltransferase activity, we isolated and characterized orthologs of *Ea*DAcT from other acetyl-TAG producing plants. Among them, the acetyltransferase from *Euonymus fortunei* possessed the highest activity in vivo and in vitro. Mutagenesis of conserved residues of DAcTs revealed that Ser-253, His-257 and Asp-258 are essential for enzyme activity of *Ea*DAcT, suggesting their involvement in the enzyme catalysis. Alteration of residues unique to acetyltransferases did not alter the acyl donor specificity of *Ea*DAcT, implying that multiple amino acids are important for substrate recognition. Together, this work identifies the structural features of *Ea*DAcT and offers an initial view of the amino acids important for activity of the enzyme.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACAT	Acyl-CoA:cholesterol acyltransferase
ASAT	Acyl-CoA:sterol acyltransferase
DAcT	Diacylglycerol acetyltransferase
DAG	Diacylglycerol
DGAT	Acyl-CoA:diacylglycerol acyltransferase
ER	Endoplasmic reticulum
ESI-MS	Electro spray ionization-mass spectrometry
FAME	Fatty acid methyl ester
GC-FID	Gas chromatography-flame ionization detector
GOAT	Ghrelin O-acyltranferase
HA	Hemagglutinin
HHAT	Hedgehog acyltransferase
lcTAG	Long chain TAG
LPCAT	Acyl-CoA:lysophospholipid acyltransferase
MBOAT	Membrane bound O-acyltransferase
mPEG-mal	Methoxypolyethylene glycol maleimide
NEM	N-ethylmaleimide
PORC	Porcupine

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Timothy Durrett, for research ideas, funding and guidance, and for editing my publications and dissertation. I want to thank other past and current members of the Durrett lab, including Dr. Sunil Bansal, Dr. Jose Azna-Moreno, Karanbir Aulakh, Catherine Kornacki, James Houghton, Garrison Olds, Nathan Henderson, for their friendship, support and their help with my research.

I would like to thank Dr. Ruth Welti and Mary Roth, Libin Yao for providing me with crucial mass spectral data and teaching me mass spectrometry. I specially thank Dr. Ruth Welti for her support for me during my dissertation writing.

I would like to thank Dr. Kathrin Schrick and Dr. Larry Davis and their former graduate students including Aashima Khosla, Dan Stuckey, Janet Paper and Rohit Kamat for research materials, research ideas and meaningful discussions via weekly joint lab meeting.

I would like to thank Jennifer Shelton and Dr. Susan Brown for helping me with the processing of RNA sequencing data.

I would like to thank Dr. Jeroen Roelofs and Dr. Stella Lee for research materials.

I would like to thank the prep lab coordinator Mrs. Sue-Yi Huang for research materials and friendship.

I would like to thank my graduate committee members, namely, Dr. Timothy Durrett, Dr. Ruth Welti, Dr. Michael Kanost, Dr. Michal Zolkiewski, and Dr. Om Prakash for their helpful guidance through my graduate training. I would also want to thank Dr. Weiqun Wang for serving as the outside chair on my committee and for his helpful critiques of my dissertation.

I would like to thank my mother, Nguyen Thi Mai Anh, my late father, Tran Trong Viet, and my sister for their support over the years. I specifically thank my mother for taking care of my son during my dissertation writing. I thank my beloved husband, Hieu Sy Vu, for being supportive and understanding.

Chapter 1 - Literature review

1.1 The membrane bound O-acyltransferase family contains important enzymes

The membrane bound O-acyltransferase (MBOAT) enzyme family consists of multipass membrane proteins that acylate a variety of lipid and non-lipid substrates (Figure 1.1). The enzyme family is characterized by the presence of a conserved MBOAT domain embedded in a highly hydrophobic segment (Hofmann, 2000). MBOAT enzymes play key roles in many important biological processes. For instance, hedgehog acyltransferase (HHAT) palmitoylates sonic hedgehog protein, a morphogen, whose signaling is essential for proper differentiation and partitioning of tissues during vertebrate development (Buglino and Resh, 2008). Cholesterol acyltransferase type 1 and type 2 (ACAT1 and ACAT2) catalyze a cholesterol esterification reaction important in maintaining cholesterol homeostasis in vertebrates (Chang et al., 2009). Ghrelin O-acyltranferase (GOAT) octanoylates ghrelin, a weight- and glucose-modulating peptide hormone. Lipid modification of ghrelin is critical for stimulating the release of its receptor, the growth hormone secretagogue receptor 1a and promoting food intake, carbohydrate utilization, and adiposity (Yang et al., 2008; Ozawa et al., 2009). Porcupine (PORC) attaches a palmitate fatty acid to a serine residue of Wingless proteins Wnts, secreted signaling proteins that control embryogenesis and tissue homeostasis (Logan and Nusse, 2004). Acylation of Wnt by PORC is essential for the proper transport of the proteins during secretion processes (Takada et al., 2006). Acyl-CoA:lysophospholipid acyltransferases (LPCAT) are involved in the remodeling of membrane lipids. The enzymes are important in generating membrane asymmetry and diversity (Shindou and Shimizu, 2009). In addition, acyl-CoA:diacylglycerol acyltransferase type 1 (DGAT1) enzymes transfer a long chain acyl group from acyl-CoA to the *sn-3* position of a diacylglycerol (DAG) synthesizing triacylglycerol (TAG), the major lipid storage in eukaryotes (Liu et al., 2012). Mammalian DGAT1 enzymes have been implicated in body weight homeostasis (Smith et al., 2000). In plants, TAG is the main component of vegetable oil used for human consumption, industrial applications, and biofuels (Durrett et al., 2008; Dyer et al., 2008). *Ea*DAcT, a diacylglycerol acetyltransferase enzyme recently isolated from *Euonymus alatus* (burning bush) is also a member of the MBOAT family. The enzyme catalyzes the synthesis of an unusual TAG molecule, 3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG), by acetylating a DAG

molecule (Durrett et al., 2010). The enzyme is the only known MBOAT enzyme that can use acetyl-CoA as an acyl donor substrate (Bansal and Durrett, 2016a).

1.2 The molecular mechanism of MBOATs is largely unknown

The highly hydrophobic nature of MBOATs has prevented structural studies of the proteins. Attempts to purify recombinant tung tree DGAT1 in bacteria were not successful. Most of the recombinant tung tree DGAT1 produced in bacteria remained in inclusion bodies, and the purified proteins were prone to degradation (Cao et al., 2011). Purification of *Brassica napus* DGAT1 produced in *Saccharomyces cerevisiae* suffered from a low yield insufficient for protein structural studies using conventional approaches (Caldo et al., 2015). Therefore, despite the biological importance of MBOATs, the molecular mechanism of the enzymes is largely unknown.

An invariant histidine residue present within the MBOAT domain and an asparagine amino acid positioned in a hydrophilic region were proposed to be the putative active sites of the acyltransferase enzymes (Hofmann, 2000). However, the importance of these amino acids in several MBOATs remains unclear. For instance, mutation at the invariant histidine residue eliminates enzyme activity of almost all MBOATs tested but reduces enzyme activity of Hhat by only 50%. Similarly, the asparagine residue is essential for enzyme acitivity of HHAT and GOAT but is not required for PORC function (Yang et al., 2008; Buglino and Resh, 2010; Rios-Esteves et al., 2014). Additionally, site-directed mutagenesis at other amino acid residues also affected enzymatic activities of some MBOATs at different levels, suggesting the involvement of multiple functional amino acid residues in the catalytic activity of the enzymes. A summary of relative acyltransferase activity of MBOAT family member mutants is presented in Table 1.1.

While amino acids important for activity of MBOATs that acylate peptides and proteins have been studied in detail, similar studies have not been performed on acyltransferase enzymes that acylate lipids. The one exception is ACAT1 whose functional amino acid residues have been intensively studied (Das et al., 2008). The lack of such information hinders the understanding of the substrate selection mechanism of MBOAT enzymes.

1.3 Membrane topology-a structural feature determinant of integral membrane proteins

Approximately 20%-30% of genes in most genomes encode intergral membrane proteins that are involved in various biological processes (Krogh et al., 2001). However, typical methods

such as X ray crystallography and nuclear magnetic resonance used for solving three dimensional structures of water soluble proteins cannot be successfully applied for most intergral membrane proteins due to their highly hydrophobic nature.

The alpha-helical bundle is the predominant structural motif observed in the membrane-spanning portion of integral membrane proteins. Beta-barrel transmembrane domains are found less frequently (White and Wimley, 1999). Regardless of their size and shape, membrane proteins possess a common structural feature influenced by the lipid bilayer environment in which they are embedded: hydrophobic transmembrane segments are buried in the lipid bilayer membrane at a certain angle and hydrophilic loops are oriented toward the cytoplasm or the non-cytoplasmic compartment (van Geest and Lolkema, 2000; von Heijne, 2006). The number of transmembrane segments and their orientations relative to the lipid bilayer membrane is defined as the membrane topology of a protein. This topology can be predicted by multiple methods that employ the following principles: the middle of the membrane is enriched with hydrophobic amino acids, tyrosine and tryptophan tend to reside in the water-lipid interface regions, positively charged residues are likely located in cytoplasmic parts of the protein, and charged and polar residues are rarely present in the membrane embedded segment (von Heijne, 2006). A summary of commonly used membrane topology prediction algorithms is presented in Table 1.2.

1.4 Biochemical approaches to verify predicted membrane topology of a protein

Multiple biochemical approaches are employed to verify predicted protein membrane topologies. The approaches require introduction of target sites including reporter enzymes, antibody epitope tags, proteolytic sites, N-glycosylation sites, or chemically modifiable amino acid residues at specific positions of a protein (van Geest and Lolkema, 2000) without exception. The localization of modified positions in relative to the membrane is determined using biochemical techniques. Thus, by inserting reporter sites in different hydrophilic loops, a predicted membrane topology of a protein can be verified. Table 1.3 presents commonly selected reporter sites and biochemical approaches used to determine their localizations. Typically, a thorough scanning for reporter sites whose introduction has a minimal effect on proper folding of a protein is required prior to verification of a protein membrane topology.

1.5 Experimentally verified MBOAT membrane topologies

The structural features that determine the substrate specificity of MBOAT enzymes are largely unknown. How the enzymes recognize chemically and structurally different acyl acceptor substrates remains to be elucidated. Mapping the membrane topology of an integral membrane protein is crucial for understanding the mechanism by which a protein selects its substrates. For instance, the membrane orientations of the substrate binding sites and the active sites are important in determining the supply of substrates for an enzyme. Although the membrane topologies of some important mammalian MBOAT proteins have been mapped, similar work has not been done for plant MBOATs. The lack of such structural information hinders understanding the substrate selection mechanism of MBOAT family.

1.5.1 Membrane topology of ACAT type1

ACAT1 was the first MBOAT to have its membrane topology experimentally determined. The topography of human ACAT1 was initially defined by employing protein epitope tagging approach in conjunction with selective permeabilization of cellular membranes followed by immunofluorescence microscopy (Lin et al., 1999). The proposed membrane topology of hACAT1 included seven TMDs with a long cytosolic N-terminus stretch that consists of 147 amino acids and a 31 amino acid C-terminus oriented toward the ER lumen. In this model, the MBOAT homeodomain that included the putative histidine active site of the protein resided in the largest luminal loop (Lin et al., 1999). The model was later revised using modification of a single cysteine residue approach which further confirmed the original localization of the N- and C-termini (Guo et al., 2005). However, the later study demonstrated the presence of two additional TMDs that directed the C-terminal region of the largest luminal loop proposed in the earlier study to the cytoplasm, creating an additional short luminal loop. In the new model, the MBOAT domain is located in one of the additional TMD (Guo et al., 2005). The rest of the hACAT1 topology is similar between the two studies (Figure1. 2).

1.5.2 Membrane topology of HHAT

The membrane topology of Sonic hedgehog acyltranferase has been recently mapped (Matevossian and Resh, 2015). The deduced protein topology was obtained by monitoring localization of FLAG and HA epitope tags introduced to selective internal regions of HHAT protein using protease protection assays and indirect immunofluorescence of cells permeabilized with selective reagents. The enzyme contains ten TMDs and two re-entrant loops. The majority

of the protein is oriented toward the cytoplasm including the short N- and C-termini (Figure 1.2) (Matevossian and Resh, 2015). The invariant putative histidine residue of HHAT was determined to be embedded in a TMD.

1.5.3 Membrane topology of GOAT

Grelin O-acytransferase was confirmed tobe an ER membrane protein that contains 11 TMDs separated by relatively short hydrophilic loops. The N- and C-termini of the protein are localized on opposite sides of the lipid bilayer membrane. A re-entrant loop was experimentally identified approximately in the middle of the protein from amino acid 165 to amino acid 180. The invariant putative histidine residue is in a luminal loop (Figure 1.2). The topology model was experimentally determined by employing selective permeabilization followed by indirect immunofluorescence microscopy in combination with glycosylation shift immunoblotting (Taylor et al., 2013).

1.5.4 Membrane topology of murine DGAT1

The topology of the murine DGAT1 was experimentally determined in microsomal vesicles isolated from HEK-293T cell using multiple approaches including a protease protection assay and indirect immunofluorescence microscopy (McFie et al., 2010). According to the proposed model, murine DGAT1 spans the lipid bilayer membrane 3 times with 2 large luminal loops and a small cytosolic loop residing approximately in the middle of the protein. The N-terminus of murine DGAT1 is cytosolic while its C-terminus is oriented toward the ER lumen (Figure 1.2). In the proposed model, the putative histidine active site of the protein was mapped to the ER lumen.

A more recent study on topology of murine DGAT1 isolated from rat liver, however, demonstrated that the protein has dual topologies within the ER (Wurie et al., 2011). The study took advantage of measuring DGAT-mediated activity in conjunction with selective permeabilization of membranes by different reagents in hepatocytes to infer orientation of the active site in respect to the ER membrane. Based on the observation that DGAT activity was equally detected on the cytosolic and luminal aspects of the ER membrane of HEpG2 cells, Werie et al. concluded that murine DGAT1 has dual membrane topology. This study, however, did not provide direct evidence for the proposed dual topology of murine DGAT1.

1.6 *Euonymus alatus* diacylglycerol acetyltranferase – an MBOAT that synthesizes unusual TAG

3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG) accumulates in seeds of various plants including those belong to the Celastraceae, Balsaminaceae, Lardizabalaceae, Ranunculaceae and Rosaceae families (Bagby and Smith, 1967; Kleiman et al., 1967). It is also found in small quantities in *Eurosta solidaginis*, a freeze-tolerant insect and in deer antlers (Yang et al., 2004; Marshall et al., 2014). Acetyl-TAG contains an acetyl group at the *sn-3* position of diacylglycerol (Figure 1.3) instead of a longer acyl group found at the same position in regular TAG (herein referred to as lcTAG) (Durrett et al., 2010).

The unusual structure of acetyl-TAG gives the molecule altered physical and chemical properties, such as reduced kinematic viscosity and lower melting temperature compared to lcTAG (Durrett et al., 2010; Marshall et al., 2014). The advantageous properties of acetyl-TAG suggest potential uses as biofuels and for other industrial applications. In fact, a mixture of acetic acid esters of mono- and diglycerides of fatty acids that contains acetyl-TAG is being used to formulate emulsifiers, food coating agents, lubricants, anti-dusting agents and plasticizers (Gaupp and Adams, 2004).

The enzyme *Euonymus alatus* diacylglycerol acetyltransferase (*Ea*DAcT) catalyzes the transfer of the shortest possible acyl group, acetate, to the *sn-3* hydroxyl group of 1,2-DAG to form acetyl-TAG. The enzyme was isolated from the endosperm tissue of *Euonymus alatus* seeds that contain 94.5 mole% acetyl-TAG (Durrett et al., 2010). Phylogenetic analysis reveals that *Ea*DAcT belongs to a plant-specific subfamily of the MBOAT family (Figure 1.1) that includes an Arabidopsis acyl-CoA:sterol acyltranferase (*At*ASAT1) and the jojoba (*Simmondsia chinensis*) wax synthase (*Sc*WS) (Durrett et al., 2010; Liu et al., 2015).

1.7 *Ea*DAcT clusters with distinct substrate preference enzymes

*Ea*DAcT clusters with the *At*ASAT enzyme that catalyzes the transfer of a long acyl group to sterols forming sterol esters (Chen et al., 2007)(Figure 1.1). Although structurally related to acyl-CoA: sterol acyltransferases of yeast and animal origins, the *At*ASAT protein is a shorter polypeptide and lacks two conserved motifs, FYxDWWN and H/YSF proposed to be important for ACAT activity (Cao et al., 1996; Oelkers et al., 1998; Chen et al., 2007) and thus deviates from other known sterol acyltransferases in the MBOAT family (Yang et al., 1996; Yu et al., 1996).

*Ea*DAcT also clusters with *Sc*WS (Figure1.1) which catalyzes the transfer of a long acyl chain from acyl-CoA to fatty alcohols. The enzyme synthesizes the abundant wax esters in seeds of the desert shrub jojoba (Lardizabal et al., 2000). *Sc*WS bears no similarity to the mammalian wax ester synthase (Cheng and Russell, 2004) and the bifunctional wax synthase/acyl-CoA:diacylglycerol acyl transferases (WS/DGAT) isolated from *Arabidopsis thaliana* and bacteria (Li et al., 2008).

Although *Ea*DAcT exhibits a high degree of sequence similarity and identity to *At*ASAT1 and *Sc*WS, it possesses a strong preference for acetyl-CoA as the acyl donor substrate. It is capable of using acyl-CoAs with acyl chain lengths shorter than eight carbon atoms, albeit with lower efficiency. The enzyme can acylate aliphatic chain fatty alcohol as well as DAG containing a range of acyl chain lengths. In addition, the acetylation activity of the *Ea*DAcT is strictly to the *sn-3* position of glycerolipids when the other hydroxyl groups on the glycerol backbone are already esterified by acyl groups (Bansal and Durrett, 2016a).

We hypothesized that the distinct acyl donor substrate specificity of *Ea*DAcT compared to *At*ASAT1 and *Sc*WS, coupled with their high degree of sequence similarity, would facilitate the identification of amino acid residues important for the substrate specificity of the proteins. Such work in conjunction with the determination of the membrane topology of *Ea*DAcT protein would allow us to elucidate the substrate selection mechanism of *Ea*DAcT. For example, the determination of the substrate pool accessible to *Ea*DAcT might rely on the membrane orientations of the active site and substrate binding regions of the enzyme. Given the extremely short length of the acetyl-CoA substrate, an understanding of how *Ea*DAcT controls the selection of the substrate will provide insights into the mechanism by which the MBOAT family controls the length of the acyl donor chain. Such knowledge is useful in better manipulating acyltransferase reactions with the aim of improving human health and enhancing agricultural production.

In Chapter 2, we employed multiple approaches to map the membrane topology of *Ea*DAcT. To identify amino acid residues important for enzyme activity of *Ea*DAcT, we isolated functional diacylglycerol acetyltransferase enzymes from acetyl-TAG producing plant species to obtain additional DAcT sequences. The diverse DAcT sequences allowed us to better compare MBOAT sequences to identify amino acid residues conserved among MBOAT enzymes or within DAcT enzymes. To study the role of these amino acids, we introduced site-directed

mutations to these residues in *Ea*DAcT and monitored effects of the mutations on enzyme activity and stability of the protein. The isolation of plant orthologs of *Ea*DAcT and the mutagenesis studies are detailed in Chapter 3 of this dissertation.

Additionally, during the isolation of DAcT orthogs, we discovered a class of novel bifunctional WS/DGAT members of the MBOAT family. The enzymes are distinct from the WS/DGAT enzymes isolated from *A. thaliana* and bacteria. The isolation and initial characterization of these enzymes are described in Chapter 4.

Tables and figures

*Ea*DAcT	*Ca*ACAT1	*Ca*ACAT2	*Hs*HHAT	*Mm*GOAT	*Mm*PORC	*Mm*DGAT1
	C92A-85%					
	S127A-100%	S109A-100%				
W15	Y142F-100%	Y124F-100%				
	S194A-80%	S176A-100%				
		H224A-0%				
I105	H268N-80%	Y124A-100%				
T106	S269A-34%	S249A-100%	S182A-100%			
F132			Y207A-2%			
	Y308F-70%					
			G217A-20%			
F135	Y312F-60%	Y289F-100%		C181A-100%		
P136			S221A-87%			
L140		Y290F-50%				
I145	Y322F-25%	Y300F-80%				
L161	C333A-74%					
A183		H304A-100%				
E184	C387A-54%					
	D400N-0%	D387E-100%				
F295	Y404F-20%	Y382F-25%				
E297	D406A-100%	D384A-100%				
A201	S410A-100%					
S203	S412A-100%	S390A-100%				
Q205	S414L-50%	S392L-10%				
W212					W305A-50%	
N213			D339A-10%	N307A-0%	N306A-80%	
S217	H425A-35%	H403A-20%				
L220			L346A-100%			
T223		Y411F-80%	Y207A-2%			
Y225	Y433A-40%		Y351A-100%			
F250			F371A-100%		Y334A-40%	
S253	S456A-0%	S434A-35%			S347A-13%	
M256			W378A-20%		L340A-20%	
H257	H460A-0%		H379A-50%	H338A-0%	H341A-4%	H426A-0%
	C516A-80%					
I283		Y496A-70%				
C293	C528A-42%					
M311	C546A-45%					
Reference	Lu et al., 2002 Das et al., 2008	Das et al., 2008	Buglino et al., 2010	Yang et al., 2008	Rios-Esteves et al., 2014	McFie et al., 2010

Table 1.1 Relative enzyme activities of MBOAT family member mutants

Positions of the mutations on other proteins were mapped to analogous residues on the *Ea*DAcT protein. Amino acids are presented using the single letter nomenclature system. A mutant is denoted as XnX' in which X is the native amino acid residue, X' is the replacement amino acid, and n is the position of the amino acid on the protein. Relative activities of the mutants compared to the wild-type protein are indicated. Enzyme abbreviations: ACAT, Acyl-CoA:Cholesterol acyltransferase; DAcT, Diacylglycerol acetyltransferase; DGAT, Acyl-CoA:Diacylglycerol acyltransferase ; HHAT, Sonic hedgehog acyltransferase; GOAT, Ghrelin O-Acyltransferase; PORC, Porcupine. Species abbreviations: *Ca*, *Chlorocebus aethiops*; *Ea*, *Euonymus alatus; Hs*, *Homo sapiens*; Mm, *Mus musculus*.

Membrane topology prediction program	Reference	Principle and features
TOPPRED	(von Heijne, 1992)	Combining hydrophobicity plots with the 'positive-inside rule'. TOPPRED is capable of generating a set of possible topologies ranked by the 'positive-inside rule'.
TMHMM	(Krogh et al., 2001)	TMHMM was developed based on the Hidden Markov Model. TMHMM is able to model helix length.
TMFinder	(Deber et al., 2001)	Combination of hydrophobicity and nonpolar phase helicity scales. TMFinder can reliably distinguish between membrane proteins and globular proteins.
MEMSAT-3	(Jones, 2007)	Uses scores compiled from membrane protein data and a dynamic programming algorithm in combination with neural network trained on sequence profiles. MEMSAT-3 is able to calculate the most probable length, location and topological orientation of each transmembrane segment.
OCTOPUS	(Viklund and Elofsson, 2008)	Combines the Hidden Markov Model and artificial neural networks. Octopus is able to integrate modeling of re-entrant/membrane-dipping regions and transmembrane hairpins in the topological grammar.
MemBrain	(Shen and Chou, 2008)	Integrates the sequence representation by multiple sequence alignment matrix with the optimized evidence-theoretic K-nearest neighbor prediction algorithm and the fusion of multiple prediction window sizes. MemBrain showed improvement in predicting the ends of TMDs shorter than 15 residues and is capable of detecting N-terminal signal peptides.
TOPCONS	(Hennerdal and Elofsson, 2011)	TOPCONS is a consensus based topology prediction that computes consensus predictions of membrane protein topology using a Hidden Markov Model and input from several topology predictors
MEMSAT-SVM	(Nugent and Jones, 2012)	Employing Pore-Walker method to label pore-lining residues in a training set of transmembrane protein structures, MEMSAT-SVM is capable of automatically identifying pore-lining regions
SCAMPI	(Peters et al., 2016)	Accounting for the higher hydrophobicity of the N- and C-terminal helices, SCAMPI shows improved performance in predicting the topology of membrane proteins that contain large non-membrane domains.

Table 1.2 Summary of commonly used membrane topology prediction algorithms

Target site	Biochemical approaches to localize the reporter sites	Reference	Rationale
Antibodies epitope tags	Proteolysis in conjunction with immunoblotting.	(Romano and Michaelis, 2001; McFie et al., 2010)	Epitope tags fused luminal loops are not accessible by externally added protease. The protected epitope tags can be detected by immunoblotting with antibodies against epitope tags.
	Selective permeabilization of cellular membranes in conjunction with immunofluorescence	(Kast et al., 1996; Matevossian and Resh, 2015)	Cellular membranes can be selective permeabilized with different reagents thus allowing accessibility of epitope tag to fluorescent conjugated antibodies which can be detected by fluorescence microscopy
Proteolytic sites	Proteolysis in conjunction with immunoblotting.	(Sahin-Toth et al., 1995)	Proteolytic sites located in cytoplasmic loops but not luminal loops are accessible to externally added proteases.
Glycosylation tag	Electrophoresis separation of glycosylated proteins and the non-glycosylated counter parts.	(Taylor et al., 2013)	The glycosylation of a protein occurs exclusively in the ER lumen. Thus only glycosylation sites located in a luminal loop are accessible to the glycosylation machinery. Glycosylation at a single site in a protein results in an increase in the apparent molecular mass of approximately 2.5 kDa.
Enzyme tags: alkaline phosphatase, ß-galactosidase, and ß-lactamase	Enzymatic assays	(Broome-Smith et al., 1990; Bibi and Beja, 1994; Geller et al., 1996; Silhavy et al., 1977)	Active conformations of several enzymes are formed within a selective cellular compartment. Enzymatic activity of correctly folded proteins can be assayed.
Biotin acceptor domain	Immunoblotting of the biotin-avidin protein complex	(Jander et al., 1996)	The biotinylation of the biotin acceptor domain occurs exclusively in the cytoplasm. Biotin forms a stable complex with avidin.
Cysteine residues	Chemical modification of cysteine residues in combination with immunoblotting	(Long et al., 1998; Guo et al., 2005; Wang et al., 2008; Liu et al., 2011)	A cysteine residue can be modified specifically by chemical reagents differing in membrane permeability. Modification at the cysteine residue in a protein results in detectable change in the protein.

Table 1.3 Biochemical approaches to verify predicted membrane topology of integral membrane proteins

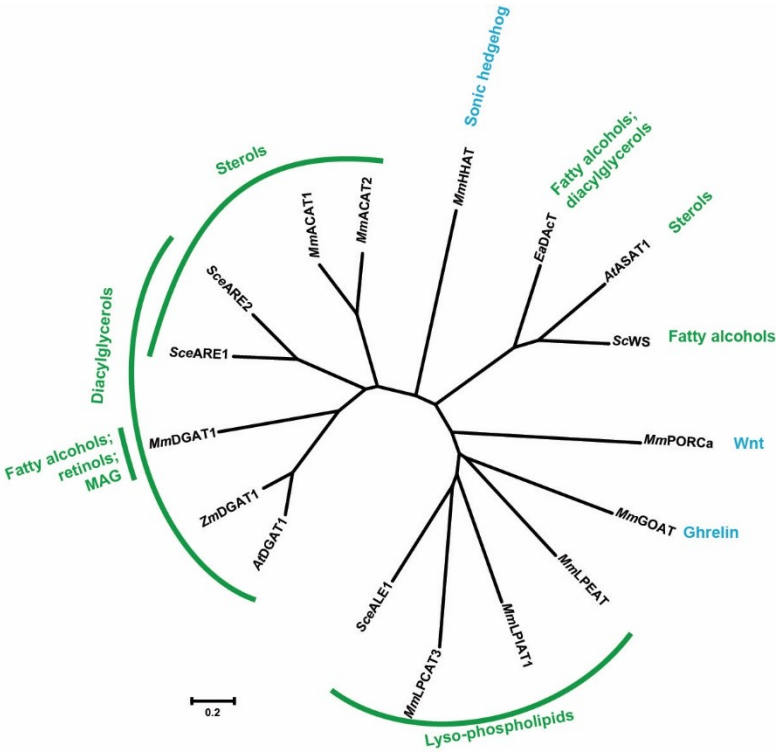

Figure 1.1 A phylogenetic tree of representative MBOATs with known substrates

Phylogenetic analyses were conducted using MEGA6. The evolutionary history was inferred using the neighbor-joining method and the evolutionary distances were computed using the Poisson correction method. The tree is drawn to scale with branch lengths proportional to evolutionary distances. The 0.2 scale represents 20% change. Lipid substrates are indicated in green and protein/peptide substrates in blue. MAG, monoacylglycerol. Enzyme abbreviations: ACAT, acyl-CoA cholesterol acyltransferase; ALE, acyltransferase for lysophosphatidylethanolamine; ARE, ACAT-related enzyme; ASAT, acyl-CoA sterol acyltransferase; DAcT, diacylglycerol acetyltransferase; DGAT, diacylglycerol acyltransferase; GOAT, ghrelin O-acyltransferase; HHAT, hedgehog acyltransferase; LPCAT, lysophosphatidylcholine acyltransferase; LPEAT, lysophosphatidylethanolamine acyltransferase; LPIAT, lysophosphatidylinositol acyltransferase; PORC, porcupine. Species abbreviations: *At*, *Arabidopsis thaliana*; *Ea*, *Euonymus alatus*; *Mm*, *Mus musculus*; *Sce*, *Saccharomyces cerevisiae*; *Sc*, *Simmondsia chinensis*; Zm, *Zea mays*. Protein accession numbers: AtDGAT1, NP_179535.1; AtASAT1, NP_190765.1; EaDAcT, ADF57327.1; MmACAT1, NP_033256.2; MmACAT2, AAC64057.1; MmDGAT1, NP_034176.1; MmGOAT, NP_001119786.1; MmHHAT, NP_659130.2; MmLPCAT3, NP_660112.1; MmLPEAT, NP_705774.1; MmLPIAT1, NP_084210.2; MmPORCa, NP_058609.1; SceALE1, NP_014818.1; SceARE1,NP_009978.1; SceARE2, NP_014416; ScWS, Q9XGY6.1; ZmDGAT1, NP_001288553.1.

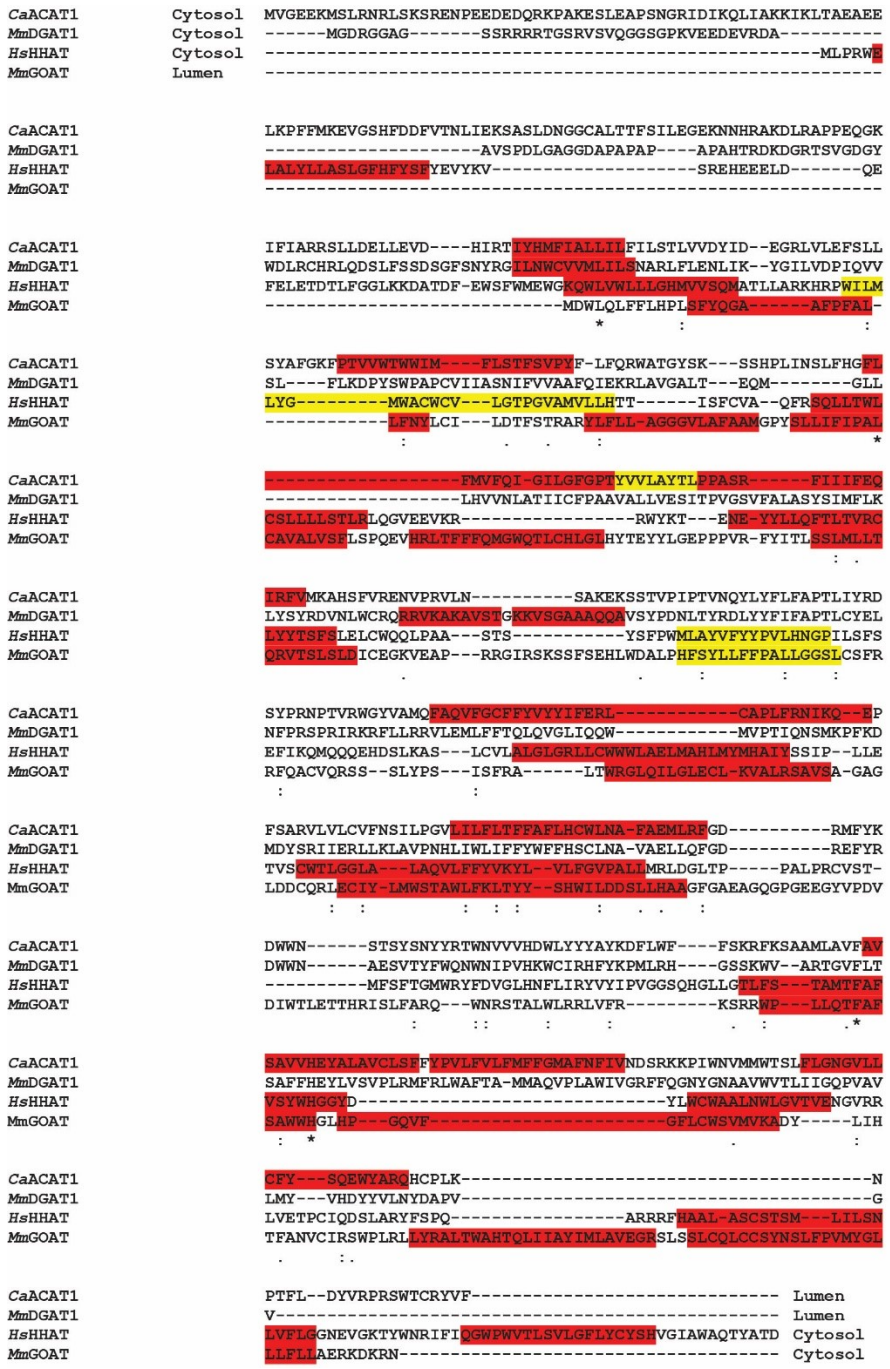

Figure 1.2 Schematic presentation of experimentally verified membrane topologies of MBOAT enzymes

Transmembrane segments are highlighted in red. Re-entrant loops are colored in yellow. The orientations of the N- and C-termini are indicated. Enzyme abbreviations: ACAT, Acyl-CoA:Cholesterol acyltransferase; ; DGAT, Acyl-CoA:Diacylglycerol acyltransferase; GOAT, Ghrelin O-acyltransferase; HHAT, Sonic hedgehog acyltransferase. Species abbreviations: *Ca*, *Chlorocebus aethiops*; *Hs*, *Homo sapiens*; Mm, *Mus musculus.*

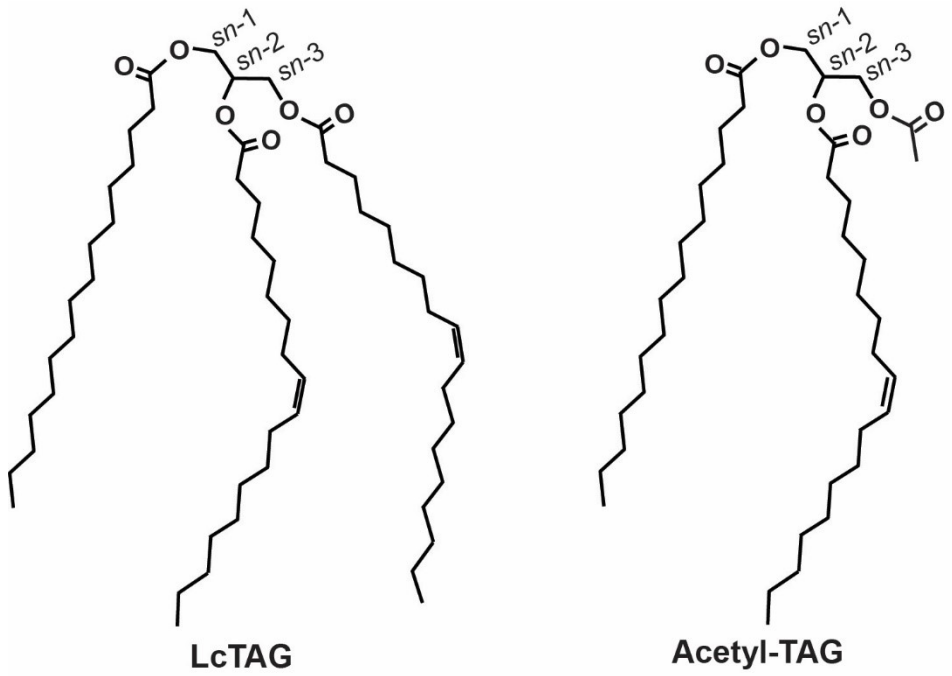

Figure 1.3 Structures of a lcTAG and an acetyl-TAG molecule

The acetyl-TAG molecule contains an acetate group at the *sn-3* position of the glycerol backbone while a long chain acyl group is present at the same position in the case of lcTAG.

Chapter 2 - Membrane topology of *Ea*DAcT, a plant MBOAT with unusual substrate specificity

2.1 SUMMARY

Euonymus alatus diacylglycerol acetyltransferase (*Ea*DAcT) catalyzes the transfer of an acetyl group from acetyl-CoA to the *sn*-3 position of diacylglycerol (DAG) to form 3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG). *Ea*DAcT belongs to a plant-specific clade of the membrane bound O-acyltransferases (MBOAT) that acylate different lipid substrates. Sucrose gradient density centrifugation revealed that *Ea*DAcT colocalizes to the same fractions as an endoplasmic reticulum specific marker. By mapping the membrane topology of *Ea*DAcT, we obtained the first experimentally determined topology model for a plant MBOAT. The *Ea*DAcT model contains four transmembrane domains with both the N- and C- termini oriented toward the lumen of the endoplasmic reticulum. In addition, there is a large cytoplasmic loop between the first and second transmembrane domains. The MBOAT signature region of *Ea*DAcT is embedded in the third transmembrane domain close to the interface between the membrane and the cytoplasm. During the topology mapping, we discovered two cysteine residues (C187 and C293) located on opposite sides of the membrane that are important for enzyme activity.

2.2 INTRODUCTION

The *Euonymus alatus* diacylglycerol acetyltransferase (*Ea*DAcT) catalyzes the acetylation of the *sn*-3 hydroxyl group of DAG to form 3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG) (Durrett et al., 2010). The presence of acetyl group instead of a longer fatty acyl group gives acetyl-TAG altered physical properties compared to regular triacylglycerols (here referred to as lcTAG), including reduced viscosity and lower melting temperatures (Durrett et al., 2010; Liu et al., 2015). *Ea*DAcT's preferred substrate is DAG but it is also capable of acetylating fatty alcohols to form alkyl acetates, albeit with reduced efficiency (Bansal and Durrett, 2016a; Ding et al., 2016). This weak wax synthase activity is consistent with the enzyme's sequence similarity to the jojoba wax synthase, and distances *Ea*DAcT from the DGAT1 enzymes with which it shares a common DAG substrate (Durrett et al., 2010; Liu et al., 2015).

*Ea*DAcT belongs to the membrane bound O-acyltransferase (MBOAT) family which consists of integral membrane enzymes that catalyze the transfer of the fatty acid moiety from an

acyl-CoA to a variety of lipid and protein substrates (Hofmann, 2000). Most characterized MBOATs acylate lipids. For example, acyl-CoA: cholesterol acyltransferases (ACAT) attach fatty acids to cholesterol (Chang et al., 1993a) and type 1 diacylglycerol acyltransferases (DGAT1) acylate the *sn*-3 position of diacylglycerol (DAG) to form triacylglycerols (TAG) (Cases et al., 1998). Other enzymes are involved in the remodeling of the fatty acid composition of membranes through their activity towards lysophospholipids (Shindou and Shimizu, 2009). Additionally, a few MBOATs transfer fatty acids to protein and peptide substrates, typically to activate biological functions. This group includes the mouse enzyme Porcupine that adds the monounsaturated fatty acid palmitoleate to a key serine of Wnt-3a in order for this signaling protein to be secreted (Takada et al., 2006). Similarly, Hedgehog acyltransferase (HHAT) palmitoylates Sonic hedgehog as part of the signaling protein's post-translational processing (Buglino and Resh, 2008). Ghrelin O-acyltransferase (GOAT) adds octanoic acid to the hormone peptide ghrelin, thus allowing it to stimulate weight gain and modulate glucose metabolism (Gutierrez et al., 2008; Yang et al., 2008).

Despite the biological importance of MBOATs, knowledge of molecular mechanisms controlling substrate selection and acyltransferase reaction remains ambiguous due to the lack of a high-resolution protein structure. The highly hydrophobic nature of MBOATs hinders structural study of the enzymes using conventional approaches such as crystallography and nuclear magnetic resonance. Therefore, current knowledge of MBOAT structural information is based on the membrane topology of the proteins, including the number of transmembrane segments and their orientations relative to the membrane. While such topologies for several mammalian MBOATs have been mapped (Guo et al., 2005; McFie et al., 2010; Rios-Esteves et al., 2014; Matevossian and Resh, 2015), similar studies for plant MBOATs have not yet been performed.

In this study, we employed multiple experimental approaches to determine the membrane topology of *Ea*DAcT and generate the first topology model for a plant MBOAT.

2.3 EXPERIMENTAL PROCEDURES

2.3.1 Cloning and mutagenesis of *Ea*DAcT mutants

Sequences encoding the cysteine-less *Ea*DAcT variants were synthesized by Genewiz (South Plainfield, NJ). Site directed mutations were introduced using the Phusion site-directed mutagenesis kit (New England Biolabs, MA, USA). Primers used for cloning and mutagenesis

are listed in Table 2.1. All acetyltransferase genes were expressed in the TAG-deficient, quadruple knockout *S. cerevisiae* strain H1246 (Sandager et al., 2002).

2.3.2 Sucrose gradient fractionation

Cell lysate from yeast expressing *Ea*DAcT was loaded onto a 10%-70% discontinuous sucrose gradient in the presence or absence of 20 mM $MgCl_2$ and centrifuged at 100,000 g for 18 h. 1 ml fractions were carefully collected from the top and 20 µl volumes of each fraction were separated on a 12% SDS-PAGE (GenScript, NJ, USA). *Ea*DAcT-HA, Pma1, and Sec61 proteins were detected using immunoblotting.

2.3.3 Microsome DGAT assay and lipid analysis

Microsomal protein isolation and in vitro DGAT assays were performed as described previously (Durrett et al., 2010), except that galactose-mediated induction of protein expression occurred for 24 h and the DGAT assay reactions contained 15 µg of microsomal protein and were incubated at 30 ˚C for 15 min. Neutral lipid extraction from plant tissues and dried yeast cells and subsequent mass spectrometry analysis were done as described previously (Durrett et al., 2010).

2.3.4 Protease protection assays

Right-side-out microsomes were isolated as described previously (Liu et al., 2011) using Zymolyase-100T (Sunrise Science Products, San Diego, CA, USA). The microsomes were resuspended in microsome storage buffer containing 250 mM sorbitol, 50 mM potassium acetate, 20 mM Tris-HCl, pH 7.4. Protease protection assays were performed as previously described (Abeliovich et al., 1998). Microsomal proteins (40 µg) were treated with protease K (0.3 mg/ml) in the presence or absence of 0.02% Triton X100 on ice for 1 h. Reactions were quenched by the addition of phenylmethylsulfonyl fluoride to a final concentration of 1 mM. Proteins were resolved with 10% SDS-PAGE and detected by immunoblotting for HA, cMyc, and Kar2.

2.3.5 Thiol-specific chemical modifications

*Ea*DAcT cysteine residues were chemically modified by thiol-specific reagents as described previously (Liu et al., 2011) with the following modifications. Each reaction contained 20 µg right-side-out microsomal protein, 20 mM N-ethylmaleimide (NEM; Sigma Aldrich), 5 mM methoxypolyethylene glycol maleimide (mPEG-mal; 5 kDa; SunBio, Orinda, CA) or 0.3% Triton X100 (all final concentrations) as appropriate and were incubated on ice for 1 h. Reactions

were terminated by the addition of 100 mM dithiolthreitol (final concentration) to quench thiol-specific reagents.

2.3.6 Immunoblotting and protein quantification

Resolved proteins from SDS-PAGE were transferred to a nitrocellulose membrane and incubated with the appropriate primary antibodies at the following dilutions: anti-HA antibody (1:10,000; clone 2-2.2.2.14, Thermo Scientific), anti-Kar2 (1:1000; clone sc-33630, Santa Cruz Biotechnology), anti Pma1 antibody (1:500), anti Sec61 antibody (1:1,000), and anti cMyc antibody (1:3000; Santa Cruz Biotechnology) followed by incubation with secondary antibodies. The anti-Sec61 polycolonal antibody and the anti-Pma1 polycolonal antibody were kind gifts from Dr. Randy Schekman (University of California, Berkeley) and Dr. Carolyn Slayman (Yale School of Medicine), respectively. For protease protection assays, membranes were incubated with goat anti-rabbit IgG-HRP and rabbit anti-mouse IgG horse radish peroxidase conjugated secondary antibodies and the antibodies were detected using SuperSignal West Pico Chemiluminescent substrate (Thermo Scientific). For other experiments involving protein quantification, membranes were probed with goat anti-rabbit IgG Dylight^{TM}488 and goat anti-mouse IgG Dylight^{TM}633 antibodies (Thermo Scientific). The blots were scanned on a Typhoon scanner (model 9410; Amersham, with the signal for *Ea*DAcT-HA obtained using the 633 nm laser and signals for Sec61 and Pma1 obtained using the 488 nm laser. Protein abundance was quantified using densitometry analysis with ImageQuantTL software (version 8.1; GE Healthcare Life Science).

2.4 RESULTS

2.4.1 *Ea*DAcT localizes to the endoplasmic reticulum

As the synthesis of acetyl-TAG occurs in a manner analogous to that of regular TAG (Milcamps et al., 2005), we hypothesized that *Ea*DAcT is also localized to the endoplasmic reticulum (ER), the site of TAG biosynthesis (Huang, 1992). To confirm the sub-cellular localization of *Ea*DAcT in yeast, we separated cellular extracts from yeast expressing a hemagglutinin (HA) epitope tagged version of *Ea*DAcT using sucrose gradient fractionation. In the absence of Mg^{2+}, *Ea*DAcT abundance mirrored that of the ER protein Sec61 with both proteins most abundant in the fifth fraction from the top of the gradient (Figure 2.1). In the presence of Mg^{2+}, the accumulation of both proteins shifted to a heavier fraction (Figure 2.1). In contrast, the plasma membrane protein Pma1 was most abundant in the ninth fraction regardless

of the presence or absence of Mg^{2+}. These results suggest that when expressed in yeast, *Ea*DAcT associates with the ER membrane but not the plasma membrane.

2.4.2 Both termini of *Ea*DAcT are exposed to the lumen

*Ea*DAcT was previously predicted to be a multipass transmembrane protein (Durrett et al., 2010). Different transmembrane domain (TMD) prediction algorithms previously employed to predict topology of other MBOATs also indicated multiple TMDs present in *Ea*DAcT (Table 2.2). According to these predictions, *Ea*DAcT possesses either an even or odd number of transmembrane segments, resulting in different orientations of the N- and C- termini relative to the membrane bilayer. However, four TMDs corresponding to amino acids 10-30, 157-177, 240-260, 309-333 were consistently predicted with high scores, even by different algorithms.

We used protease protection assays to experimentally determine the orientation of the N- and C-termini of *Ea*DAcT. Similar methods have been successfully used to map the topology of several multipass membrane proteins (Romano and Michaelis, 2001; Paul et al., 2007; Matevossian and Resh, 2015). To implement this approach, we fused the N- and C-termini of *Ea*DAcT with Myc and HA epitope tags, respectively. The in vitro acetyltransferase activity of the tagged *Ea*DAcT protein (Myc-*Ea*DAcT-HA) was comparable to that of native *Ea*DAcT (Figure 2.2a). In addition, expression of the epitope tagged protein in the yeast H1246 quadruple mutants, which is unable to synthesize TAG (Sandager et al., 2002), resulted in accumulation of acetyl-TAG but not lcTAG (Fig 2.2b). Together, these results indicate that fusion of epitope tags to the N- and C- termini of *Ea*DAcT does not alter *Ea*DAcT activity and suggest that the tags do not dramatically perturb *Ea*DAcT's native structure.

To map the orientation of the N- and C- termini of *Ea*DAcT, we isolated intact right-side out microsomes from yeast expressing Myc-*Ea*DAcT-HA and monitored accessibility of the epitope tags to protease K. In the absence of protease K, *Ea*DAcT migrated as single 40-kDa band as predicted for *Ea*DAcT in a monomeric form (Figure 2.3). Exposure to protease K followed by immunoblotting with anti-HA antibodies resulted in multiple reactive bands: full-length *Ea*DAcT and 3 smaller segments. Similarly, immunoblotting to detect the Myc-epitope after protease K treatment of microsomes resulted in 2 protein bands: full-length *Ea*DAcT at a significantly reduced level and a smaller fragment with a strong signal (Figure 2.3). These results indicate that both the Myc and HA epitope tags at the N- and C-termini of *Ea*DAcT were located in the ER lumen where they were protected by the lipid bilayer membrane and thus inaccessible

to protease K. The multiple bands are due to truncated *Ea*DAcT proteins resulting from protease K digestion of cytosolic loops. Treatment with both detergent and protease K resulted in undetectable signal of Myc-*Ea*DAcT-HA and a significant reduced signal of Kar2 suggesting that the treatment duration was sufficient for complete digestion of Myc-*Ea*DAcT-HA when not protected by the membrane. In the absence of detergent, the luminal chaperone protein Kar2 was not accessible to protease K and its levels remained the same as in untreated microsomes, confirming that microsomes remained intact during the isolation process.

2.4.3 Multiple cysteine residues are required for *Ea*DAcT activity

The luminal nature of both the N and C termini of *Ea*DAcT suggested that the protein possesses an even number of transmembrane segments. We employed an approach that relies on the chemical modification of cysteine residues (Guo et al., 2005; Liu et al., 2012) to determine the topology of the remainder of the protein. Substituting all nine endogenous cysteine residues with alanine (*Ea*DAcT-C_0A) resulted in a mutant with only 10% activity of the native protein (Figure 2.4b), suggesting that at least some of the cysteine residues are important for *Ea*DAcT activity. This extreme reduction in activity was unexpected as the removal of cysteine residues from other MBOAT proteins, such as ACAT1, allowed the enzymes to retain 40% of wild-type activity (Lu et al., 2002).

An attempt to restore activity by substituting the structurally similar amino acid serine in place of cysteine was also unsuccessful; in fact, lower levels of protein accumulation suggested the *Ea*DAcT-C_0S mutant was more unstable (Figure 2.4a). Therefore, to determine which particular cysteine(s) are important for *Ea*DAcT activity, we re-introduced individual residues to the *Ea*DAcT-C_0A protein. Among these variants, the *Ea*DAcT-C187 (containing a single cysteine at position 187) showed the highest in vitro activity with 45% that of the native protein (Figure 2.4b). Replacement of C187 by alanine in the wildtype protein reduced *Ea*DAcT enzyme activity by 90% confirming the importance of C187 (Figure 2.5). As the reintroduction of C187 to the cysteineless protein did not fully restore all enzyme activity, we hypothesized that *Ea*DAcT must rely on multiple cysteine residues for full activity. Indeed, re-introduction of both cysteine 187 and cysteine 293 restored activity to 85% of the wild-type protein (Figure 2.4b), indicating that these two residues are particularly important for *Ea*DAcT activity. Most of the other *Ea*DAcT mutants containing two cysteine residues showed activity comparable to that of

*Ea*DAcT-C187 but some combinations, such as *Ea*DAcT-C102-C187, *Ea*DAcT-C284-C187 and *Ea*DAcT-C361-C187, possessed lower activity (Figure 2.4b).

2.4.4 *Ea*DAcT possesses four transmembrane domains

We initially determined the membrane orientation of C187 and C293 in 3 cysteine mutants with different activities to test whether the changes in activity reflect alterations in membrane topology. The membrane orientation of a cysteine residue relative to the lipid bilayer membrane can be determined by its accessibility to thiol-specific reagents with different membrane permeabilities and molecular weights. NEM and mPEG-mal are thiol-specific reagents commonly used for the purpose of determining the localization of cysteine residues in relative to the membrane bilayer (Bogdanov et al., 2005; Guo et al., 2005; Liu et al., 2012). They form stable thioether linkages with the sulfhydryl group of a cysteine molecule. NEM is a small membrane permeable reagent (125 Da) that serves as a cysteine blocking reagent whose crosslinking to *Ea*DAcT does not significantly increase the apparent molecular weight of the protein. It is however mostly unreactive to cysteine residues embedded in the membrane bilayer or at the protein-protein interface due to the lack of a water molecule used as a proton acceptor in such environments (Bogdanov et al., 2005). Methoxypolyethylene glycol maleimide, on the other hand is, at 5 kDa, large and membrane impermeable; thus it is only accessible to cytosol exposed cysteine residues. The reagent is accessible to lumen localized cysteine residues only when the membrane bilayer is solubilized by a detergent. Crosslinking of mPEG-mal to a single cysteine residue of *Ea*DAcT protein resulted in a detectable band shift of 5 kDa on a SDS-PAGE. Figure 2.7 depicts predicted immunoblot patterns of different *Ea*DAcT proteins that contain single cysteine residues treated with combinations of NEM and mPEG-mal under solubilizing and non-solubilizing conditions.

We isolated right-side-out microsomes from yeast expressing *Ea*DAcT-C187 (45% WT activity), *Ea*DAcT-C293 (18% activity), and *Ea*DAcT-C187-C293 (85% activity). The right-side-out microsomes were subjected to chemical modification of cysteine residues with NEM and mPEG-mal under intact or solubilized conditions. Modification of Kar2, which contains one cysteine, by NEM and mPEG-mal was used as a control for the intactness of the microsome vesicles. In the absence of detergent, a Kar2 band shift was not observed suggesting that Kar2 was protected from mPEG-mal modification by the membrane (Figure 2.7, lanes 2 and 3). However, when the vesicles were disrupted and NEM was absent, mPEG-mal crosslinked to

Kar2 as evidenced by a band shift (Figure 2.7, lane 4). A light band shift was observed in treatment with detergent in the presence of both and NEM and mPEG-mal (Figure 2.7, lane 5), but was not as evident as in lane 4, suggesting that a small portion of Kar2 was not completely blocked by NEM. *Ea*DAcT-C187 showed the same pattern as Kar2 suggesting that C187 was luminal (Figure 2.7). In contrast, when *Ea*DAcT-C293 was exposed to PEG-mal, a band shift was readily observed in the absence of detergent, indicating that C293 is exposed to the cytosol (Figure 2.7). When *Ea*DAcT-C187-C293 was analyzed, a band shift was observed in the absence of detergent (Figure 2.6, lane 2) and is attributable to one mPEG-mal modification, and is consistent with modification of C293. In contrast, when the microsomes were disrupted with detergent, a band with a greater shift was observed without NEM blocking consistent with modification of both C187 and C293 (Figure 2.6, lane 4). This result confirms that C187 and C293 are located on opposite sides of the membrane bilayer, with C293 cytosolic and C187 luminal. Importantly, we were able to derive the same conclusions about the topology of *Ea*DAcT using variants with different activity levels, suggesting the topology of *Ea*DAcT protein structure was likely reflective of the native protein.

The ability to differentiate between one and two modified cysteines via different band shifts (Figure 2.7) allowed us to map the location of the soluble loops between predicted TMDs using additional *Ea*DAcT mutants, each containing C187 and one other appropriately placed cysteine. We decided to include C187 in these two-cysteine mutants in order to maintain the activity of the mutants at a level at least equal to that of the *Ea*DAcT-C187 mutant (Figure 2.4b).

In order to localize some segments lacking endogenous cysteine, we constructed *Ea*DAcT-A52C-C187and *Ea*DAcT-S122C-C187, which contained novel cysteine residues introduced by amino acid replacement and were expressed at levels comparable to wild-type *Ea*DAcT (Figure 2.4a). In these experiments, we assumed that C187 was luminal and the positions of the other cysteine residue were determined by looking for additional band shifts. Based on the mPEG-mal modification patterns of various *Ea*DAcT two-cysteine variants (Figure 2.8), we conclude that luminal cysteine residues include C179 and C187. Cytosolic cysteine residues include A52C, C102, S122C, C284 and C293. As NEM was unable to block C165 and C166, these residues are mapped to a transmembrane segment.

We integrated the localization of the N- and C- termini of *Ea*DAcT determined by protease protection assay (Figure 2.3), the orientations of cysteine residues relative to the ER

membrane (Figure 2.7, Figure 2.8), and the predicted TMD of *Ea*DAcT (Table 2.2) to propose a model for *Ea*DAcT topology consistent with these results (Figure 2.9).

2.5 DISCUSSION

The ER localization of yeast-expressed *Ea*DAcT (Figure 2.1) is consistent with the similar localization of DGAT enzymes which also synthesize TAG (Shockey et al., 2006; Stone et al., 2009). Instead of a C-terminal dilysine motif present in many other ER resident proteins (Benghezal et al., 2000), *Ea*DAcT possesses two potential pentapeptide ER retrieval motifs of the type -ϕ-X-X-K/R/D/E-ϕ-, where ϕ is any large hydrophobic amino acid residue (McCartney et al., 2004). Such motifs have previously been shown to be important for the ER retention of tung tree (*Vernicia fordii*) DGAT enzymes which also lack a C-terminal dilysine motif (Shockey et al., 2006). However, the role of such motifs in subcellular localization of *Ea*DAcT needs to be further studied.

We tested the validity of different membrane topology predictions for *Ea*DAcT (Table 2.2) with a combination of protease protection (Figure 2.3) and cysteine-modification assays (Figure 2.7 and 2.8), to generate the first experimentally verified topology model for a plant MBOAT (Figure 2.9). In this model, *Ea*DAcT spans the membrane lipid bilayer four times with both the N- and C-termini located in the ER lumen. Where these are known, the orientations of the termini of other MBOAT proteins do not fall into a predictable pattern. Both N- and C-termini can be cytoplasmic, as for tung tree DGAT1 and HHAT (Shockey et al., 2006; Matevossian and Resh, 2015; Konitsiotis et al., 2015), or they can be located on opposite sides of the membrane with either terminal located toward the cytoplasm, such as murine DGAT1, ACAT1, Porcupine, and Goat (Guo et al., 2005; McFie et al., 2010; Taylor et al., 2013; Rios-Esteves et al., 2014).

A particularly noticeable feature of *Ea*DAcT's topology is the relatively large cytosolic loop of 125 amino acids between the first and second TMDs (Figure 2.9). Any potential TMDs in this region were only weakly predicted by some algorithms (Table 2.2). Further, experimental evidence demonstrating the cytoplasmic localization of A52, C102 and S122 (Fig 2.8) argues for the absence of any TMDs from residues 31 to 156.

The first TMD from amino acids 10 to 30 was predicted by all algorithms with high probability (Table 2.2) and was confirmed by the luminal orientation of a Myc epitope tag fused to the N-terminus of *Ea*DAcT (Figure 2.3) and the cytosolic orientation of a cysteine residue

substituted in place of A52 (Figure 2.8). The absence of a canonical N-terminal signal peptide suggests that this region may also serve as an uncleaved signal-anchor sequence that directs the binding of nascent protein to the ER and initiates cotranslational insertion (Wessels and Spiess, 1988). The placement of the second, third and fourth TMDs is supported by experimental evidence that locates C179 and C187 in the lumen, followed by C284 and C293 in the cytosol and the C-terminus in the lumen (Fig 2.8). The hydrophobic MBOAT signature region which contains the putative histidine active site is situated in the third TMD, consistent with ACAT1 and PORC topology models (Guo et al., 2005; Rios-Esteves et al., 2014). This TMD consists of 35 amino acid residues and is comparable to the size of the homologous TMD in ACAT1, PORC, and HHAT (Guo et al., 2005; Rios-Esteves et al., 2014; Matevossian and Resh, 2015). Interestingly, the N-terminal region of this TMD contains a high proportion of aromatic residues including a histidine and three tyrosine residues known to preferentially reside at the lipid bilayer/water interface in α-helical integral membrane proteins (Ulmschneider and Sansom, 2001). In contrast, MBOATs that use long chain acyl-CoA as acyl donor substrates do not possess similar sequences. This aromatic-rich signature suggests that this region, including the proposed histidine active site, may be in close proximity with the cytoplasm and thus readily accessible to the cytosolic acetyl-CoA substrate, possibly explaining *Ea*DAcT's preference for acetyl-CoA. Ultimately however, a high resolution structure of the enzyme is needed to confidently identify substrate binding regions of *Ea*DAcT.

The role of cysteine in the enzymatic activity of MBOATs has not been studied intensively. A study on human ACAT1 showed that although thiol-specific modification severely inhibits the enzyme activity, the cysteine-free enzyme is still catalytically active, possessing 40% of the activity levels of the wild type protein (Lu et al., 2002). Similarly, the enzyme activity of *Ea*DAcT was significantly reduced when all nine endogenous cysteine residues of the protein were mutated by alanine. The two residues most important for activity, C187 and C293 (Figure 2.4b) are unlikely to form a structurally important disulfide bond together as they are located on opposite sides of the lipid bilayer membrane (Figure 2.7). A further investigation on whether C187 and C293 are involved in coordinating substrate binding of *Ea*DAcT will provide insight into catalytic mechanism of the enzyme.

Acknowledgement

I would like to thank Dr. Timothy Durrett for providing research ideas and guidance for this work. I am grateful to Dr. Randy Schekman (University of California, Berkeley) and Dr. Carolyn Slayman (Yale School of Medicine) for kindly providing the anti-Sec61 and anti-Pma1 antibodies, respectively. I appreciate Dr. Kathrin Schrick and Dr. Larry Davis for their critical advice on this research during our joint lab meetings. ESI-MS was performed at the Kansas Lipidomics Research Center Analytical Laboratory where instrument acquisition and lipidomics method development was supported by National Science Foundation (EPS 0236913, MCB 1413036, DBI 0521587, DBI1228622), Kansas Technology Enterprise Corporation, K-IDeA Networks of Biomedical Research Excellence (INBRE) of National Institute of Health (P20GM103418), and Kansas State University. This work was supported by the National Science Foundation under Award No. EPS-0903806 and matching support from the State of Kansas through the Kansas Board of Regents.

Tables and figures

Name	Purpose	Sequence (5' to 3')
5'-EaDAcT-S122C	Generating *Ea*DAcT that has a serine residue at position 122 replaced by a cysteine residue	CAT CTT TTG AAA AAG GCA TTC ATG
3'-EaDAcT-S122C		TGA AAT TGG AcA TGG GTT TGT TG
5'-EaDAcT-A52C	Generating *Ea*DAcT that has an alanine at position 52 replaced by a cysteine residue	CTC CTA TTA ATC tgT CCC TTG AAC
3'-EaDAcT-A52C		ATA AG GAC AGG AAGTAA AGA GAG
5'-EaDAcT-C21	Generating *Ea*DAcT that has only one cysteine residue at position 21. Cysteine-less *Ea*DAcT was used as PCR template	CAA GCC ATG GTA tgt CTA TCT TAC GC
3'-EaDAcT-C21		TAC CCA AAC CTT GAT GAA GTT CTT G
5'-EaDAcT-C361	Generating *Ea*DAcT that has only one cysteine residue at position 361. Cysteine-less *Ea*DAcT was used as PCR template	ACA AGG TTT GTG tgt GGA AAT ATG TAC
3'-EaDAcT-C361		TCC AAG CAT GAC AAT CTT CTC C
5'-EaDAcT-C102	Generating *Ea*DAcT that has only one cysteine residue at position 102. Cysteine-less *Ea*DAcT was used as PCR template	ATC TCA ATA GCT tgt CTT CCC ATC ACA
3'-EaDAcT-C102		GAA ATG TAG GAG GTT TTG TGG GAG
5'-EaDAcT-C165-C166	Generating *Ea*DAcT that has 2 cysteine residues at position 165-166. Cysteine-less *Ea*DAcT was used as PCR template	TTA GTG ATC TAT tgt tgt CAT GTG TAC GTT
3'-EaDAcT-C165-C166		GAC CAC GTT TGG ATC CAT GTA C
5'-EaDAcT-C179	Generating *Ea*DAcT that has only one cysteine residue at position 179. Cysteine-less *Ea*DAcT was used as PCR template	TCA CTC TCT CTA tgt GCA ACC CTA GCT
3'-EaDAcT-C179		GAT ATC CAA CAT AAC GTA CAC ATG ACA AC
5'-EaDAcT-C187	Generating *Ea*DAcT that has only one cysteine residue at position 187. Cysteine-less *Ea*DAcT was used as PCR template	CT GAA TTC CTT tgt GGG TTT GAT G
3'-EaDAcT-C187		CTA GGG TTG CAG CTA GAG AGA GTG AG
5'-EaDAcT-C284	Generating *Ea*DAcT that has only one cysteine residue at position 284. Cysteine-less *Ea*DAcT was used as PCR template	TA CAT GGG ATT tgt GAA GCT TTG GA
3'-EaDAcT-C284		GGA CGA AGT GCC CTG TCA TAT C
5'-EaDAcT-C293	Generating *Ea*DAcT that has only one cysteine residue at position 293. Cysteine-less *Ea*DAcT was used as PCR template	GTG GAG ATG AAG tgt AAG AGA TCA AGG
3'-EaDAcT-C293		CTC CAA AGC TTC AGC AAT CCC A

Table 2.1 Primers used for mutagenesis of *Ea*DAcT

Nucleotides in upper case denote sequences complementary to the DNA template; lower case nucleotides denote restriction enzyme sequences or mutated sequences.

Prediction program	Reference	TMD1	TMD2	TMD3	TMD4	TMD5	TMD6	TMD7	TMD8	TMD9	TMD10
TOPPRED	(von Heijne, 1992)	10-30		55-75		130-150	157-177	240-260		309-329	
TOPCONS	(Hennerdal and Elofsson, 2011)	11-32	36-57	58-79			159-180	241-262		306-327	
OCTOPUS	(Viklund and Elofsson, 2008)	12-33	36-57	58-79		129-150	159-180	241-262		304-325	
SCAMPI	(Peters et al., 2016)	11-31	34-54	56-76	87-107		159-179	249-259		306-326	
TMFinder	(Deber et al., 2001)	13-27	37-82				159-186	235-265		309-333	340-358
MEMSAT-SVM	(Nugent and Jones, 2012)	12-27	37-56	60-81		133-148	157-188	244-261	274-289	307-327	
MEMSAT-3	(Jones, 2007)	10-29	36-59				166-189	243-262	275-294	307-326	
MemBrain	(Shen and Chou, 2008)	15-26	38-56	62-77			158-177	240-266		308-324	

Table 2.2 Predicted transmembrane domains of *Ea*DAcT

Numbers denote amino acid residue positions.

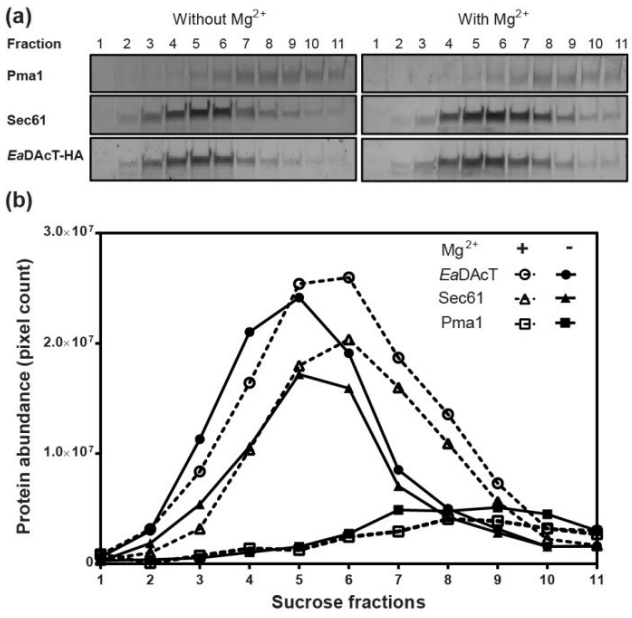

Figure 2.1 *Ea*DAcT associates with the endoplasmic reticulum

Lysates from yeast cells expressing *Ea*DAcT-HA were fractionated by centrifugation in a sucrose step gradient in the absence or presence of Mg^{2+}. (a) Immunoblots to detect the plasma membrane marker Pma1, the ER marker Sec61 and *Ea*DAcT-HA in different fractions. (b) Quantification of immunoblot signal for *Ea*DAcT, Sec61 and Pma1 in different fractions.

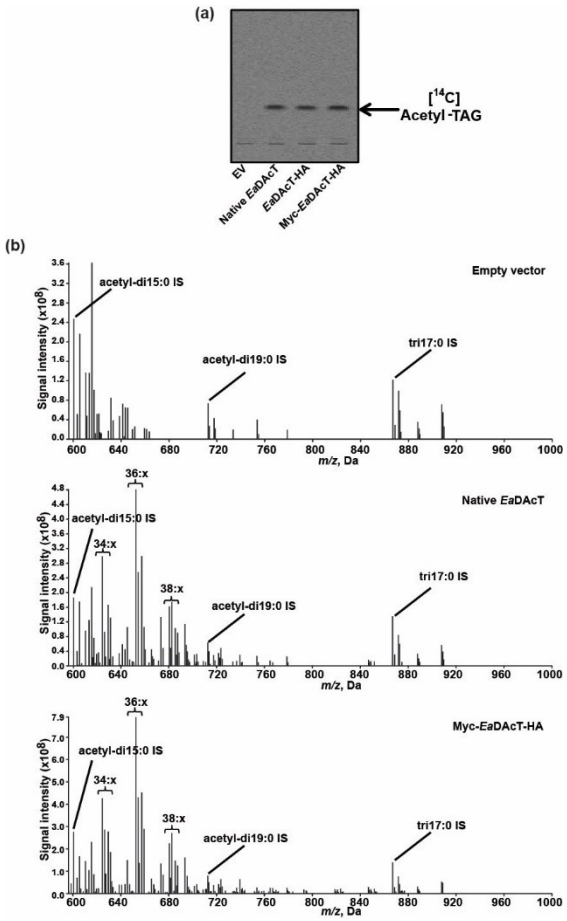

Figure 2.2 N- and C-termini epitope tags do not alter activity of *Ea*DAcT

(a) Autoradiogram of TLC separated lipid extracts from H1246 yeast microsomes expressing empty vector, native *Ea*DAcT, *Ea*DAcT-HA and Myc-*Ea*DAcT-HA incubated with [^{14}C]acetyl-CoA demonstrating the in vitro synthesis of [^{14}C]acetyl-TAG. (b) Positive-ion electrospray ionization mass spectra of neutral lipid extracts from H1246 yeast expressing empty vector, native *Ea*DAcT, and Myc-*Ea*DAcT-HA. The internal standard 3-acetyl-1,2-dinonadecanoyl-*sn*-glycerol (acetyl-di19:0) was added during lipid extraction and triheptadecanoin (tri17:0) and 3-acetyl-1,2-dipentadecanoyl-sn-glycerol (acetyl-di15:0) were added to the lipid extracts prior to ESI-MS. Peaks correspond to the m/z values of the $[M+NH_4]^+$ adducts of the intact acetyl-TAG molecules. The number of acyl carbons in each cluster of peaks is indicated; for the clarity, the number of double bonds is not defined (x).

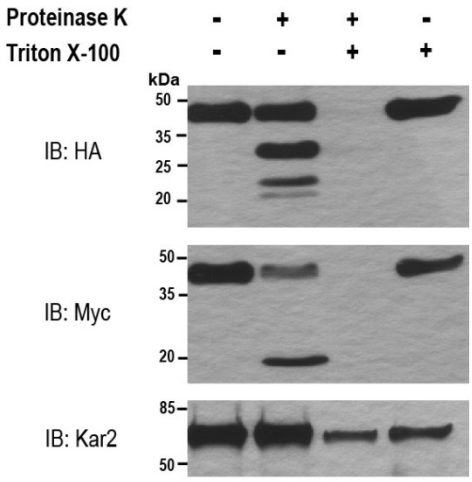

Figure 2.3 Both *Ea*DAcT termini are localized in the lumen

Protease protection assays on right-side-out microsomes isolated from yeast cells expressing Myc-*Ea*DAcT-HA. The protection of the N- or C- terminus was determined by immunoblotting for the Myc- or HA-epitopes respectively. Kar2 served as a luminal ER marker.

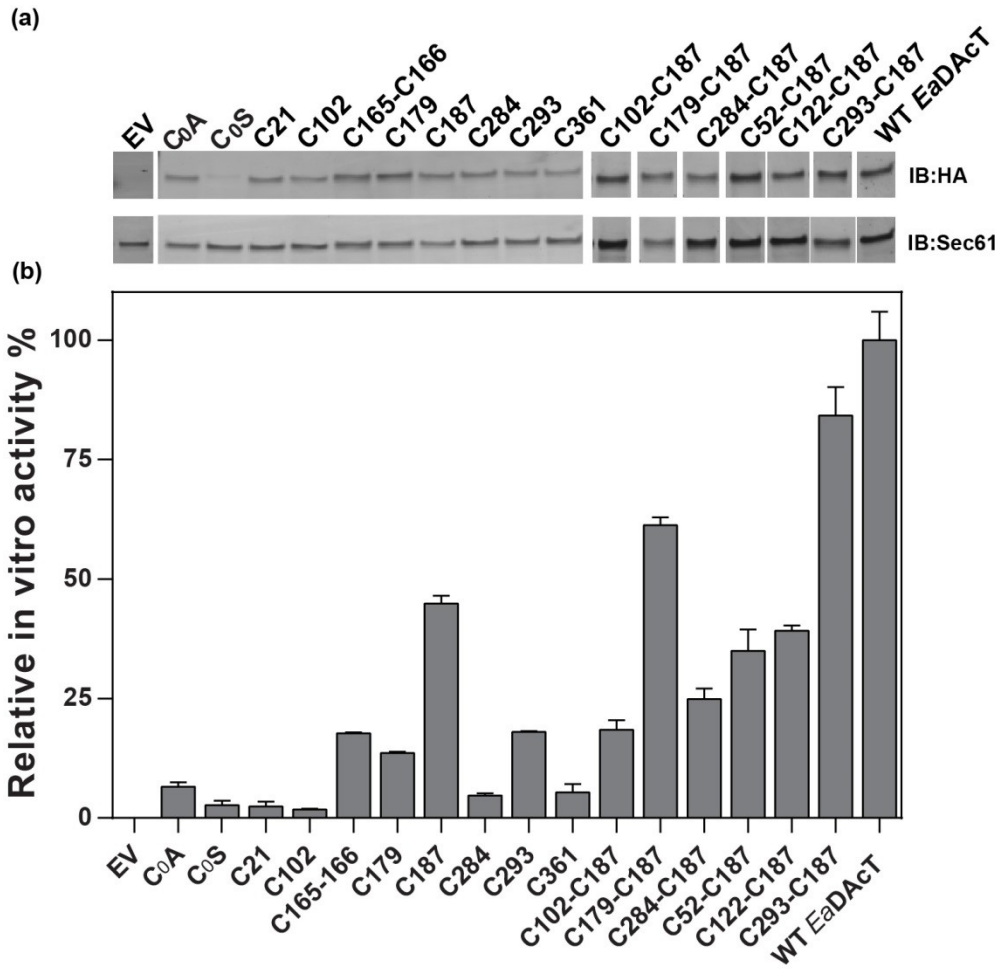

Figure 2.4 In vitro acetyl-TAG synthesis activity of *Ea*DAcT cysteine mutants

(a) Accumulation of *Ea*DAcT-HA proteins containing zero, one or two cysteine residues. Levels of different *Ea*DAcT-HA mutant proteins in yeast microsome fractions were determined by immunoblotting for the HA epitope. Sec61 was used as a loading control. EaDAcT proteins containing one or two cysteine residues were derived by modifying EaDAcT-C_0A. (b) Relative in vitro acetyltransferase activity of *Ea*DAcT cysteine mutants. Activities of the mutants were normalized to that of the wild-type *Ea*DAcT control included with each set of assays. Error bars represent the standard deviation of at least 2 technical replicates. Results shown are from one of at least two independent experiments.

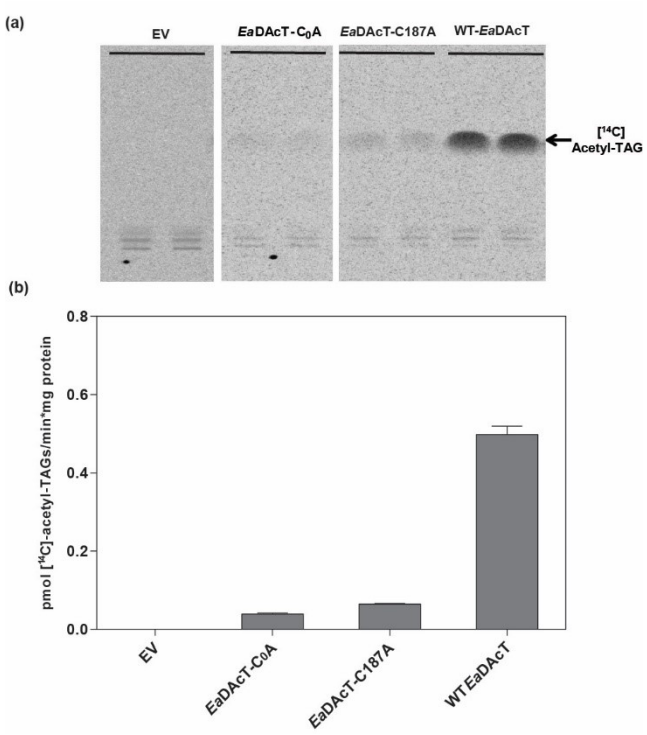

Figure 2.5 C187 is required for *Ea*DAcT activity

(a) Autoradiogram of TLC separated lipid extracts from yeast microsomes expressing the empty vector (EV), *Ea*DAcT lacking all endogenous cysteine (*Ea*DAcT-C_0A), *Ea*DAcT-C187A and wild-type *Ea*DAcT incubated with [^{14}C]acetyl-TAG. Samples were separated on the same TLC; lanes not relevant for this experiment were removed from the image. (b) Quantification of [^{14}C]acetyl-TAG synthesized by in vitro by *Ea*DAcT-C_0A, *Ea*DAcT-C187A or wild-type *Ea*DAcT. Error bars represent standard deviation of 2 technical replicates. The result is representative of at least two independent experiments.

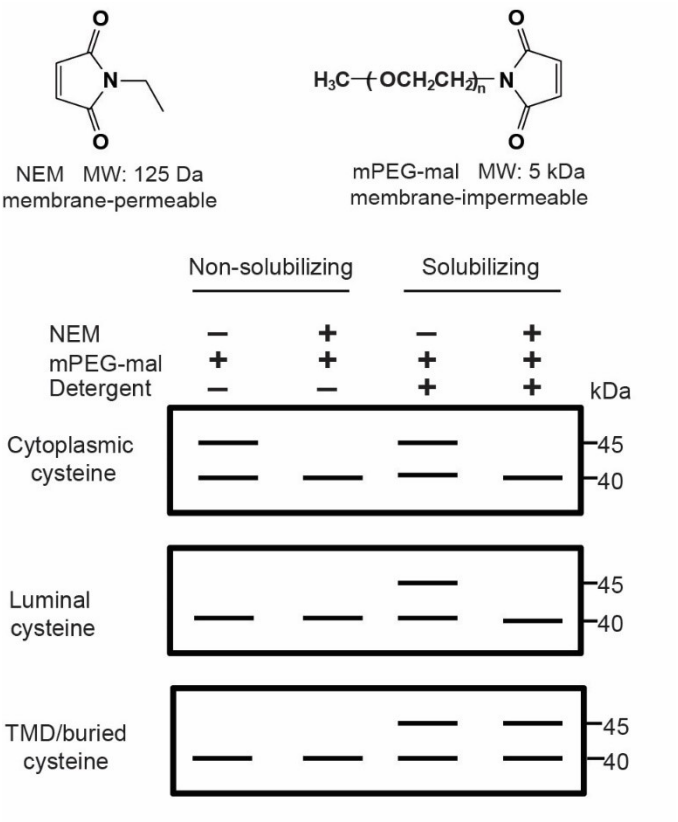

Figure 2.6 Predicted immunoblot patterns of a single cysteine containing *Ea*DAcT protein modified by NEM and mPEG-mal

Schematic depiction of immunoblot patterns of NEM and mPEG-mal treatment of a single cysteine containing *Ea*DAcT protein under solubilizing and non-solubilizing conditions. Th e figure is adapted from Liu et al., 2011.

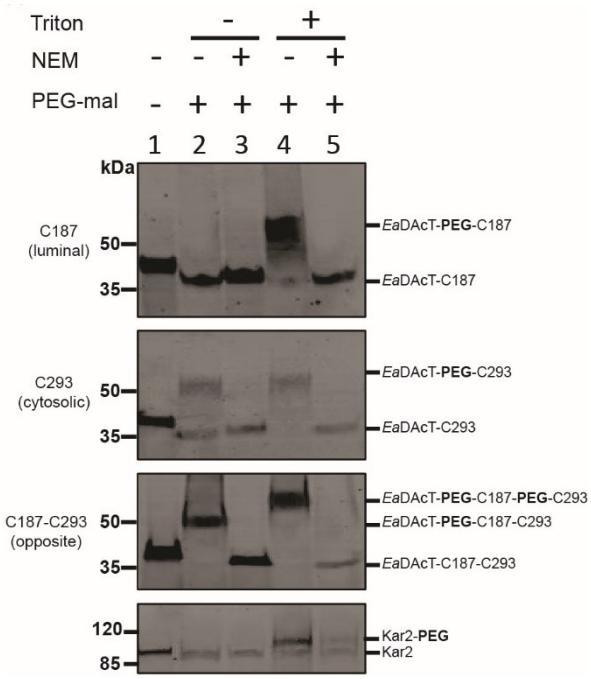

Figure 2.7 C187 and C293 are located on opposite sides of the membrane bilayer

Accessibility of C187 and C293 to thiol-specific modifications. Right-side-out microsomal membranes were prepared from yeast expressing single cysteine variants of *Ea*DAcT-HA and incubated with different combinations of mPEG-mal, NEM and Triton X-100. Thiol-specific modifications to *Ea*DAcT-HA and the luminal control protein Kar2 were detected with immunoblotting. The results shown are representative of at least 2 independent experiments.

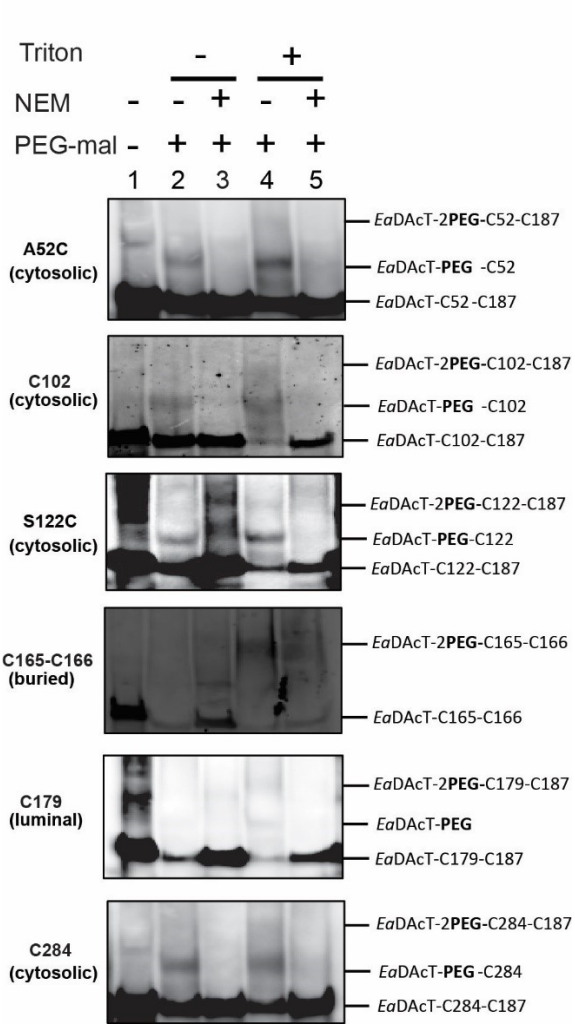

Figure 2.8 Orientation of *Ea*DAcT cysteine residues relative to the ER membrane

Right-side-out microsomal membranes were prepared from yeast expressing variants of *Ea*DAcT-HA containing two cysteine residues. Thiol-specific modification of these double cysteine *Ea*DAcT mutants was performed and changes to *Ea*DAcT-HA and Kar2 were detected using immunoblotting. The results shown are representative of at least 2 independent experiments.

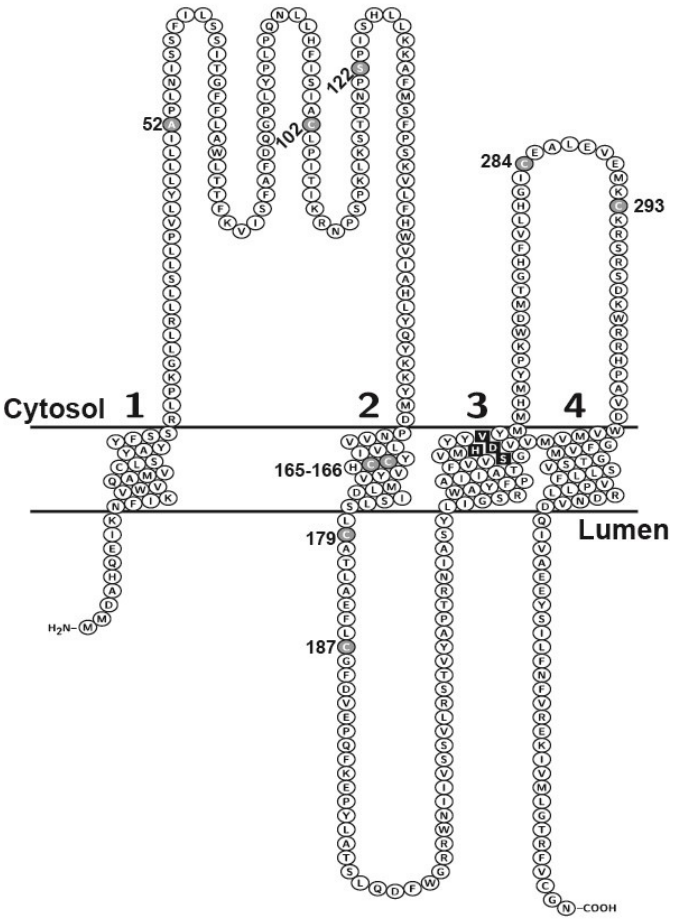

Figure 2.9 Experimentally verified *Ea*DAcT topology model

The model was created using Protter software to incorporate experimental results. Endogenous cysteines and residues converted to cysteine for topology mapping are represented by grey-filled circles. Residues for which membrane-sidedness was determined are numbered. Black-filled squares denote other amino acid residues important for activity that were mutated in this study (Chapter 3).

Chapter 3 - Identification of residues important for enzyme activity and substrate selection of *Ea*DAcT

3.1 SUMMARY

Euonymus alatus diacylglycerol acetyltransferase (*Ea*DAcT) catalyzes the transfer of an acetyl group from acetyl-CoA to the *sn*-3 position of diacylglycerol (DAG) to form 3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG). *Ea*DAcT belongs to a plant-specific clade of the membrane bound O-acyltransferases (MBOAT) that acylate different lipid substrates. In order to identify amino acid residues important for acetyltransferase activity, we isolated and characterized orthologs of *Ea*DAcT from other acetyl-TAG producing plants. Among them, the acetyltransferase from *Euonymus fortunei* possessed the highest activity in vivo and in vitro. Mutagenesis of conserved residues revealed that serine 253, histidine 257, aspartate 258 and valine 263 are essential for *Ea*DAcT activity. Alteration of individual residues unique to acetyltransferases did not alter the acyl donor specificity of *Ea*DAcT, suggesting that multiple amino acids are important for substrate recognition.

3.2 INTRODUCTION

MBOAT enzymes that acylate a variety of lipid and non-lipid substrates play key roles in many biological processes (Masumoto et al., 2015). However, the amino acid residues in the active site of the enzymes are largely unknown. While functional amino acids of some mammalian MBOATs that acylate proteins have been studied in detail, similar works have not yet been performed for plant MBOATs that mostly acylate lipids. Previous studies proposed the involvement of an invariant histidine and an asparagine residue in the active site of some mammalian MBOAT (Yang et al., 2008; Buglino and Resh, 2010; Rios-Esteves et al., 2014). However, mutation at the invariant histidine of Sonic hedgehog acyltransferase (HHAT) reduced the enzyme activity by only 50% (Buglino and Resh, 2010). Additionally, a Porcupine (PORC) mutant that has the invariant asparagine mutated to alanine retained 80% activity of the wild-type enzyme (Rios-Esteves et al., 2014). These results argue for the importance of the conserved histidine and asparagine in HHAT and PORC.

*Ea*DAcT is an unusual MBOAT that catalyzes the transfer of the shortest possible acyl group, acetate, to the *sn*-3 hydroxyl group of DAG to form 3-acetyl-1,2-diacyl-*sn*-glycerol

(acetyl-TAG; (Durrett et al., 2010). *Ea*DAcT shares high sequence similarity and identity with *Arabidopsis thaliana* sterol acyltransferase (*At*ASAT1) and *Simmondsia chinensis* (jojoba) wax synthase (*Sc*WS) that use long chain acyl-CoA as acyl donor substrates (Lardizabal et al., 2000; Chen et al., 2007). We hypothesized that the distinct acyl donor substrate specificity of *Ea*DAcT compared to *At*ASAT1 and *Sc*WS coupled with their high degree of sequence similarity, will facilitate the identification of amino acid residues important for the substrate specificity of the proteins. Such work will allow the understanding of functional amino acid residues involved in substrate selection of *Ea*DAcT. Given the extreme nature of the acetyl-CoA substrate, the understanding of how *Ea*DAcT selects its substrate will provide insight into the mechanism by which MBOATs determine the chain length of acyl donor substrates.

To obtain diverse DAcT sequences, we isolated additional DAcTs from other acetyl-TAG producing plant species.The sequence diversity of the newly isolated acetyl-TAG synthesizing enzymes provided insights into functional residues important for acetyltransferase activity. Through site directed mutagenesis, we confirmed the role of various amino acids important for the acetyltransferase activity of *Ea*DAcT. Together, this work provides an initial view of the amino acids important for enzyme activity of *Ea*DAcT.

3.3 EXPERIMENTAL PROCEDURES

3.3.1 Seed collection and RNA isolation

Seeds were collected from local plants or purchased from commercial seed suppliers. Total RNA was isolated either from endosperm tissue (Celastraceae species) or from whole seeds (*Adonis aestivalis, Akebia quinata, Sorbus alnifolia,* and *Sorbus aucuparia*) using an established procedure (Chang et al., 1993b) with the following modifications. Ground tissue was incubated in extraction buffer (2% hexadecyltrimethylammonium bromide, 2% polyvinylpryrrolidone K 30, 100 mM Tris-HCl pH 8.0, 25 mM EDTA, 2 M NaCl, 0.5 g/L spermidine, 2% ß-mercaptoethanol) at 65 ˚C for 5 min. Protein and DNA were removed by extracting with 1:1 phenol:chloroform (v/v) twice. The final RNA-containing aqueous phase was mixed with LiCl to a final concentration of 3 M and incubated at -20 ˚C overnight. RNA was precipitated by centrifuging at 11,000 g for 2 h and then washed with 70% ethanol.

3.3.2 Illumina sequencing and computational analysis

cDNA libraries were prepared from total RNA using Illumina Truseq RNA kits, and sequenced using an Illumina HiSeq 2000 sequencing system (www.illumina.com). Reads were

stringently pre-cleaned with PRINSEQ (Schmieder and Edwards, 2011) version 0.20.3. Single k-mer assemblies were performed for k=27 up to k=57 with a step size 10 with Velvet (Zerbino and Birney, 2008) version 1.2.08 and Oases (Schulz et al., 2012) version 0.2.08. Single k-mer assemblies were merged Oases using k=27. Merged assembly transcripts were clustered with MIRA (Chevreux et al., 1999) version 3.4 and CDHit (Fu et al., 2012) version 4.6.1. BLASTX version 2.2.26 (Altschul et al., 1997) was run to search assembled contigs for putative homologs to the *Arabidopsis thaliana* genome. The script OrthologHitRatio.pl was used to estimate percent of the full protein coding sequence represented in the contig, or the ortholog hit ratio (OHR). An OHR of 1 indicates a potential full-length transcript. MIRA clusters have higher OHRs therefore MIRA Clusters were used in expression profiling. Assemblies were filtered for contigs > 400 bp and aligned used Bowtie2 version 2.1.0. Alignment counts were summarized using Count_reads_denovo.pl version 1.0. BLASTP was used to search the assembled contigs for sequences with high similarity to *Ea*DAcT, which were selected for cloning. Protein sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004). Phylogenetic analysis was performed with MEGA6 software (Tamura et al., 2011) using the neighbor-joining method with the Poisson model. Divergence times for all branching points in the topology were calculated with the RelTime method (Tamura et al., 2012) using the branch lengths contained in the inferred tree. The topological organization of *Ea*DAcT was predicted using the following algorithms: TOPPRED 2 (von Heijne, 1992), TOPCONS (Tsirigos et al., 2015), OCTOPUS (Viklund and Elofsson, 2008), SCAMPI (Bernsel et al., 2008), TMfinder (Deber et al., 2001), MEMSAT-SWM (Nugent and Jones, 2012), MEMSAT-3 (Jones, 2007).

3.3.3 Cloning, expression, and mutagenesis of acyltransferase genes

Primers specific to *Ea*DAcT (5'- GGAAGAAGCCGGTGATTGATGAAAG – 3' and 5' - TTGGAGGTGGAGATGAAGTGTAAG - 3') were used to amplify the coding regions of DAcT genes from cDNA isolated from *C. scandens* and different *Euonymus* species. The 5' and 3' regions of these genes were then cloned with Rapid Amplification of cDNA Ends (RACE) kit (Epicentre, Madison, WI, USA). Transcription initiation sites were predicted using the NetStart 1.0 Prediction server (Pedersen and Nielsen, 1997). First strand cDNAs were synthesized using oligo dT primer and reverse transcriptase (SuperScript III, Invitrogen). Gene sequences containing full coding regions of DAcT orthologs were amplified from cDNA using Phusion polymerase (Thermo Fisher Scientific, MA, USA) and primers corresponding to the 5' and 3'

end regions, and were cloned into the yeast expression vector pYES2/CT (Invitrogen), with the hemagglutinin epitope coding sequence inserted to the 3' end of the gene sequences prior to the stop codon. Site directed mutations were introduced using the Phusion site-directed mutagenesis kit (New England Biolabs, MA, USA). Primers used for cloning and mutagenesis are listed in Table 3.1. All acetyltransferase genes were expressed in the TAG-deficient, quadruple knockout *S. cerevisiae* strain H1246 (Sandager et al., 2002).

3.3.4 Microsome DGAT assay and lipid analysis

Microsome DGAT assay and lipid analysis were described previously in the experimental procedure of Chapter 2. TAG was quantified using ESI-MS as described previously (Bansal and Durrett, 2016b).

3.3.5 Immunoblotting and protein quantification

Immunoblotting and protein quantification was done as described previously in the experimental procedure of Chapter 2.

3.3.6 Promoter turn-off protein stability assay

Transformed yeast expressing different acyltransferases were grown in galactose containing medium to induce protein expression. After 24 h, cultures were washed once with synthetic defined medium lacking a carbon source and switched to glucose containing medium. Cell lysates were extracted at different time points and lysate protein concentrations were quantified by incubating with BCA reagent (Thermo Fisher Scientific, MA, USA). An equal amount of protein lysate (typically 50 µg) was separated using SDS-PAGE and Sec61 loading control and HA-tagged acyltransferase proteins were detected by immunoblotting.

3.4 RESULTS

3.4.1 Seeds from divergent plant species produce acetyl-TAG

We were interested in identifying residues important for the substrate specificity of *Ea*DAcT, which preferentially uses the shortest acyl-CoA, acetyl-CoA. We hypothesized that residues important for acetyltransferase activity are conserved among DAG acetyltransferase enzymes present in the phylogenetically diverse species that produce acetyl-TAG. Most of these species belong to the Celestraceae plant family (*Euonymus* and *Celastrus* species); others are members of the Ranunculales and Ericales orders (Bagby and Smith, 1967; Kleiman et al., 1967) and therefore are distantly related to the Celestraceae family. The presence of acetyl-TAG in seeds of *Celastrus scandens*, *E. bungeanus*, *E. atropurpureus*, *E. fortunei*, *E. kiautschovicus*, *A.*

aestivalis, A. quinata, S. alnifolia, and *S. aucuparia* was confirmed by electrospray ionization mass spectrometry (ESI-MS). In Euonymus species and *C. scandens*, acetyl-TAG are the major storage lipids present while lcTAG are present at relatively low levels in endosperm tissues (Figure 3.1). Interestingly, while lcTAG predominate in the aril layer of *E. bungeanus* and *E. atropurpureus*, the aril layer of *E. fortunei*, *E. kiautschovicus,* and *C. scandens* contained mostly acetyl-TAG (Figure 3.1). Similar results were previously observed in *E. europaeus, E. verrucosus*, and *E. maximowiczianus,* where the abundance of acetyl-TAGs in seeds and arils varied from species to species (Sidorov et al., 2013). In contrast, lcTAG and acetyl-TAG were present at relatively comparable levels in *A. aestivalis, A. quinata, S. alnifolia,* and *S. aucuparia* seeds (Figure 3.1).

3.4.2 Novel orthologs of *Ea*DAcT possess acetyltransferase activity

We isolated genes homologous to *Ea*DAcT from plants that accumulate acetyl-TAG in their seeds. We then expressed these genes in H1246 yeast and analyzed lipid extracts to demonstrate that some of the enzymes were capable of synthesizing acetyl-TAG (Figure 3.2, Figure 3.3a). Similar to *Ea*DAcT (Durrett et al., 2010), no lcTAG was detected in the lipids extracted from yeast expressing these DAcT enzymes (Figure 3.2). These *Ea*DAcT orthologs varied in their ability to produce acetyl-TAG. For example, yeast expressing *Ef*DAcT accumulated the highest amount of acetyl-TAG with 28.7 nmoles/ per mg DW, 6.7 times higher than the levels achieved through expression of *Ea*DAcT (4.24 nmoles/mg DW; Figure 3.3b).Yeast expressing *Cs*DAcT accumulated the lowest amount of acetyl-TAGs (1.1 nmoles/mg DW).We isolated microsomes from yeast cells expressing the proteins, and incubated them with [^{14}C] acetyl-CoA to further confirm the acetyltransferase activity of the orthologs. The formation of [^{14}C] acetyl-TAG demonstrated that all DAcT enzymes possess acetyltransferase activity in vitro (Figure 3.3c). In general, the in vivo accumulation of acetyl-TAG reflected the in vitro activity levels, with *Ef*DAcT and *Ek*DacT showing higher activity compared to the other tested DAcT orthologs (Figure 3.3c). We quantified acetyltransferase protein levels present in the microsomes by immunoblotting against the HA epitope fused to the C-terminus of all proteins and normalizing to the abundance of Sec61 present in the yeast microsomes (Figure 3.3d). In some cases, the variation in in vitro enzyme activity of the orthologs can be partially attributed to differences in protein accumulation in the microsomes, but other factors likely also have an effect.

The higher DAG acetyl-transferase activity of *Ef*DAcT compared to *Ea*DAcT is of interest as *Ef*DAcT and *Ea*DAcT are highly similar to each other with 99% similarity and 94% identity (Figure 3.4). Among 22 amino acids different between *Ef*DAcT and *Ea*DAcT, 7 of them are weakly similar amino acids, 13 of them are strongly similar amino acids and the remaining 2 are very different amino acids. The *Ea*DAcT mutant (*Ea*DAcT-6M-HA) in which 6 amino acid residues of *Ea*DAcT, A52P, P121T, K130S, A226T, S298R, and Y341C, were mutated to make it more like *Ef*DAcT did not show increased activity compared to the wild-type *Ea*DAcT (Figure 3.5).

3.4.3 DAcT proteins are distantly related to DGAT1 proteins

Consistent with their sequence similarity to *Ea*DAcT, phylogenetic analysis indicated that the newly isolated DAcTs are more closely related to the *Simmondsia chinensis* (jojoba) wax synthase (*Sc*WS) and Arabidopsis sterol acyltransferase (*At*ASAT1) than to DGAT1 acyltransferases, which also acylate DAG (Figure 3.6). All of the newly isolated DAcTs except for *Aa*DAcT cluster together. *Aa*DAcT appears to have diverged longer ago than *Sc*WS and *At*ASAT1 from the DAcT cluster. A protein sequence alignment of these proteins revealed that while the C-terminal regions of all sequences show a high degree of similarity and identity, the N-terminal region present in DGAT1 is absent from DAcTs, *Sc*WS, and *At*ASAT1 (Figure 3.4). In addition, the MBOAT homeodomain is highly conserved among the sequences with the invariant histidine embedded in a long stretch of hydrophobic residues (Figure 3.4).

3.4.4 Effects of mutagenesis on conserved residues

We further analyzed the protein sequence alignment of DAcTs and jojoba wax synthase with the Conserved Property Difference Locator program (Mayer and Shanklin, 2007) to identify functionally different amino acids that distinguish the acetyltransferases from the long chain acyltransferases. We hypothesized that these amino acid residues may be responsible for the difference in acyl-CoA substrate usage between these different groups of MBOATs. Two potential amino acid positions conserved in all DAcT enzymes were identified: glutamine 150 (Q150) is replaced by a negatively charged glutamate in *Sc*Ws and *At*SAT1, and serine 222 (S222) which is substituted by non-polar residues (Figure 3.4). However, replacement of Q150 and S222 by a serine and a leucine residues, respectively, reduced acetyltransferase activity of *Ea*DAcT (Figure 3.7a) but did not switch *Ea*DAcT to being able to use long chain acyl-CoA as substrate as observed for jojoba wax synthase (Figure 3.7b).

Additional comparisons of the MBOAT homology domains in the different acyltransferases identified four other residues of interest: S253, H257, D258 and V263 (Figure 3.4). The first two of these, S253 and H257, are highly conserved in the MBOAT family. In addition, the homologous serine residues in ACAT1 and PORC were shown to be essential for activity (Das et al., 2008; Rios-Esteves et al., 2014). Similarly, the conserved histidine residue is necessary for enzyme activity in many other MBOATs (Guo et al., 2005; Buglino and Resh, 2010; McFie et al., 2010; Rios-Esteves et al., 2014). Consistent with these earlier results, replacement of S253 and H257 with alanine abolished *Ea*DAcT enzyme activity, revealing the essential role of these residues in catalytic activity of the protein (Figure 3.8). In addition, the mutations caused a decrease in the stability of *Ea*DAcT (Figure 3.9).

The aspartate residue (D258) adjacent to the highly conserved H257 was intriguing given the conserved negative charge at this location and the potential role of such residues in a proposed catalytic triad serine-histidine-aspartate of cholesterol acyltransferase 1 (Das et al., 2008). Substitution of D258 with alanine eliminated acetyltransferase activity and reduced the stability of *Ea*DAcT (Figure 3.8, Figure 3.9). Replacing D258 with asparagine to maintain a similar sized residue also greatly reduced in vitro activity, further demonstrating the importance of a negative charge at this position. Substituting D258 with glutamate to maintain a negative charge reduced acetyltransferase activity by 38%, indicating that the charge spacing is also important for maximal activity (Figure 3.8b). This D258E mutation makes *Ea*DAcT more similar to the sequences of acyltransferases that transfer long chain acyl-CoA; however the altered enzyme still does not possess the ability to synthesize lcTAG (Figure 3.10).

Finally, the MBOAT homology domain in the acetyltransferases contains an additional amino acid, V263 (Figure 3.4). Deletion of this residue to make the enzyme more like the acyltransferases that utilize long chain acyl-CoAs diminished in vitro acetyltransferase activity and had a negative effect on the stability of the mutant protein (Figure 3.8a, Figure 3.9). The ΔV263 mutant did not possess long chain acyltransferase activity (Figure 3.10). We constructed the D258E ΔV263 double mutant to make *Ea*DAcT even more similar to long chain acyltransferases; however, this mutant did not restore protein stability or result in long chain acyltransferase activity (Figure 3.8a, Figure 3.10).

3.5 DISCUSSION

3.5.1 Isolation of DAcT orthologs from other acetyl-TAG producing species

In order to obtain more information about the residues important for the extreme acyl donor substrate specificity of *Ea*DAcT, we isolated orthologs of the enzyme from plant species that accumulate acetyl-TAG. Six newly identified DAcT enzymes were able to synthesize acetyl-TAG in vitro and in vivo, with some possessing higher activity compared to others (Figure 3.3). Like *Ea*DAcT, these new acetyltransferases (DAcTs) are only distantly related to DGAT1 proteins, but are closely related to enzymes that acylate fatty alcohols and sterols (Figure 3.6). *Ef*DAcT possesses the highest DAG acetyl-transferase activity among the isolates (Figure 3.3) and might be a useful enzyme for the biotechnological production of acetyl-TAG in transgenic oilseed crops.

Proteins of *S. aucuparia*, *S. alnifolia* and *A. quinata* homologous to *Ea*DAcT did not show acetyl-TAG synthesis activity, indicating that the enzymes responsible for acetyl-TAG synthesis in these species may not belong to the MBOAT family. The characterization of the MBOATs isolated from these species is discussed in more detail in chapter 4.

3.5.2 Residues important for acetyltransferase activity

The acyltransferase mechanism of the MBOAT family is largely unknown due to the lack of structural information. One attractive hypothesis is the involvement of a serine-histidine-aspartate catalytic triad proposed to catalyze the acyltransferase reaction in ACAT1 (Das et al., 2008) and present in other enzymes involved in lipid metabolism, including lipases and acyltransferases (Winkler et al., 1990; Brumlik and Buckley, 1996; Peelman et al., 1998). For some enzymes, glutamate functions in place of aspartate as the acidic group (Schrag et al., 1991). Our mutagenesis data show that S253, H257 and D258 are essential for the enzyme activity of *Ea*DAcT (Figure 3.8a). Consistent with these observations, S253 and H257 in *Ea*DAcT are homologous to the serine and histidine residues in the putative catalytic triad of ACAT1 (Das et al., 2008). In many acyltransferases, including ACAT1, a glutamate replaces D258 but the role of this residue was not tested in ACAT1. Instead, an aspartic residue corresponding to E197 in *Ea*DAcT was proposed to participate in ACAT1's catalytic triad (Das et al., 2008). However, an acidic group is not present in this position in *At*ASAT1 (Figure 3.4), casting some doubt on whether this residue participates in a conserved catalytic mechanism. Instead, the low activity of D258A and D258N (Figure 3.8b) and the high conservation of aspartate or glutamate in analogous positions in acyltransferases (Figure 3.4) suggest the importance of a negative charge at D258. Further, the D258E mutants which retained that negative charge possessed

approximately 50% lower activity, indicating that the spacing of a negative charge is important for activity in *Ea*DAcT (Figure 3.8b). Consistent with this result, *Aa*DAcT, which is more similar to long-chain utilizing acyltransferases (Figure 3.6), possesses a glutamate instead of aspartate at position 258, possibly explaining this enzyme's weak acetyltransferase activity (Figure 3.3).

Another unique and highly conserved residue among the DAcTs is the additional valine at position 263 in *Ea*DAcT. Deletion of this valine residue significantly reduces the stability of *Ea*DAcT (Figure 3.8a). This result is consistent with the low stability of *Aa*DAcT when expressed in yeast (Fig 3.3) and suggests that this valine residue is important for DAcT protein stability. However, further mutagenesis on *Aa*DAcT is needed to justify this assumption. Mutations at different positions of the MBOAT domain reduced stability of *Ea*DAcT (Figure 3.9), suggesting the importance of the domain in maintaining the stable structure of the protein.

In summary, we have identified a number of amino acid residues important for the activity of *Ea*DAcT, a plant MBOAT with extreme substrate specificity. Some of the residues are highly conserved in the MBOAT family, suggesting they play roles in catalysis. Ultimately however, a structure of the enzyme is needed to confidently identify the active site. Mutation of individual residues unique to DAcT enzymes that synthesize acetyl-TAG failed to alter the specificity of *Ea*DAcT and enable it to utilize longer acyl-CoA substrates. Future studies, aided by the elucidation of the membrane organization of the enzyme (Chapter 2), will focus on altering multiple residues or regions of the protein that might be important for controlling substrate specificity.

Acknowledgement

I would like to thank Dr. Timothy Durrett for providing research ideas and guidance for this work. I would like to thank Jennifer Shelton for her help in processing the RNA sequencing data. I would like to thank Dr. Kathrin Schick and Dr. Larry Davis for their helpful advice. This work was supported by the National Science Foundation under Award No. EPS-0903806 and matching support from the State of Kansas through the Kansas Board of Regents.

Tables and figures

Name	Purpose	Sequence (5' to 3')
5'-EtDacT	Cloning *Et*DAcT to pYES2/CT vector at *Sac*I and *Xho*I site	ttggagctc ATG ATG GAT GCT CAT CGA GAG
3'-EtDAcT		tagactcgagTCA ATT TCC ACA CAT AAA ACT TG
5'-EbDacT	Cloning *Eb*DAcT to pYES2/CT vector at *Sac*I and *Xho*I site	ttggagctc ATG ATG GAT GCT CAT CGA GAG
3'-EbDAcT		tagactcgag TCA ATT TCC ACA CAT AAA CCT TG
5'-EkDAcT	Cloning *Ek*DAcT to pYES2/CT vector at *Sac*I and *Xho*I site	ttggagctc ATG ATG GAT GTT CAT CAA GAG
3'-EkDAcT		tagactcgag TCA ATT TCC ACA GAT AAA CCT TG
5'-EfDAcT	Cloning *Ef*DAcT to pYES2/CT vector at *Sac*I and *Xho*I site	ttggagctc ATG ATG GAT GTT CAT CAA GAG
3'-EfDAcT		tagactcgag TCA ATT TCC ACA GAT AAA CCT TG
5'-CsDAcT	Cloning *Cs*DAcT to pYES2/CT vector at *Sac*I and *Bst*I site	ttggagctc ATG ATG GAT TTC AAT CAA G
3'-CsDAcT		gaaccactgtgctgg CGT CAA TCC TGT TCC AA
5'-AaDAcT	Cloning *Aa*DAcT to pYES2/CT vector at *Kpn*I and *Xma*I site	caggtacc ATGGGAGGTGAACTGAGG
3'-AaDAcT		aattcccggg GCTTGTGGCAAAGATTG
5'-EaDAcT-S253A	Generating *Ea*DAcT that has the serine at position 253 replaced by an alanine	ACA TTT GTC GTA gct GGA GTT ATG CAT
3'-EaDAcT-S253A		TGC AAT TAT AGC TGG AAA ATA AGC CCA
5'-EaDAcT-H257A	Generating *Ea*DAcT that has the histidine at position 257 replaced by an alanine	TCA GGA GTT ATG gct GAT GTA GTG TAC T
3'-EaDAcT-H257A		TAC GAC AAA TGT TGC AAT TAT AGC TGG
5'-EaDAcT-D258A	Generating *Ea*DAcT that has the aspartate at position 258 replaced by an alanine	TCA GGA GTT ATG CAT gct GTA GTG TAC TAT
3'-EaDAcT-D258A		TAC GAC AAA TGT TGC AAT TAT AGC TGG
5'-EaDAcT-ΔV263	Generating *Ea*DAcT that has the valine at position 263 deleted	TAC ATG ATG CAT ATG TAT CCC AAG TGG
3'-EaDAcT-ΔV263		ATA GTA CAC TAC ATC ATG CAT AAC TCC T
3'-EaDAcT-D258N	Reverse primer to generate *Ea*DAcT that has the aspartate at position 258 replaced by an asparagine	TCA GGA GTT ATG CAT aat GTA GTG TAC TAT
3'-EaDAcT-D258E	Reverse primer to generate *Ea*DAcT that has the aspartate	TCA GGA GTT ATG CAT gaa GTA GTG TAC TAT

	at position 258 replaced by a glutamate	
5'-EaDAcT-A52P	Generating *Ea*DAcT that has the alanine at position 52 replaced by a proline	CTC CTA TTA ATC cCT CCC TTG AAC
3'-EaDAcT-A52P		ATA AAG GAC AGG AAG TAA AGA GAG
5'-EaDAcT-P121T	Generating *Ea*DAcT that has the proline at position 52 replaced by a threonine	CAT CTT TTG AAA AAG GCA TTC ATG
3'-EaDAcT-P121T		TGA AAT TGG AGA TGt GTT TGT TG
5'-EaDAcT-K130S	Generating *Ea*DAcT that has the lysine at position 52 replaced by a serine	CTT TTG AAA Agt GCA TTC ATG TCA
3'-EaDAcT-K130S		ATG TGA AAT TGG AGA TGG GTT
5'-EaDAcT-A226T	Generating *Ea*DAcT that has the alanine at position 52 replaced by a threonine	TCG ACT GTC TAT aCC CCA ACT AGA
3'-EaDAcT-A226T		ACG TAG GAC ACT TGA CAC AAT TAT
5'-EaDAcT-S298R	Generating *Ea*DAcT that has the serine at position 52 replaced by an arginine	G AGA TCA AGG cGT GAC AAG TGG
3'-EaDAcT-S298R		TT ACA CTT CAT CTC CAC CTC CAA
5'-EaDAcT-Y341C	Generating *Ea*DAcT that has the tyrosine at position 52 replaced by a cysteine	CA GAA GAG TAC Tgt ATT TTG TTT AAT
3'-EaDAcT-Y341C		CTA CAA TTT GGT CCA CAT TAT CC

Table 3.1 Primer sequences used for cloning of acetyltransferase genes and mutagenesis of *Ea*DAcT

Nucleotides in upper case denote sequences complementary to the DNA template; lower case nucleotides denote restriction enzyme sequences or mutated sequences.

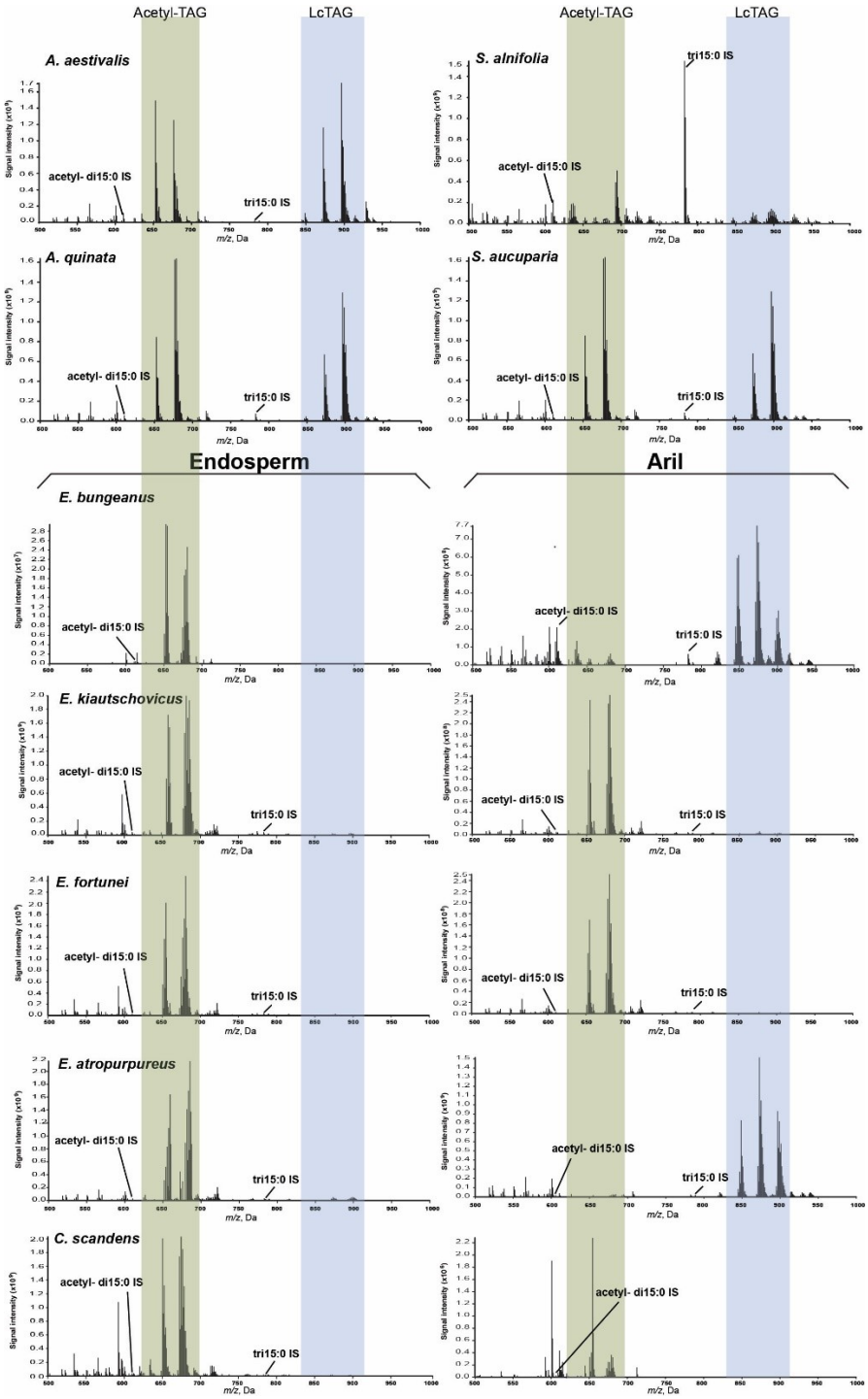

Figure 3.1 ESI-MS analysis of lipid extracts from seeds of acetyl-TAG producing plants

Positive-ion electrospray ionization mass spectra of neutral lipid extracts from aril and seed tissue of acetyl-TAG producing fruits indicating the presence of acetyl-TAG (highlighted in green) and long chain TAG (highlighted in blue). Tripentadecanoin (tri15:0) was added during lipid extraction and 3-acetyl-1,2-dipentadecanoyl-sn-glycerol standard (acetyl-di15:0) was added prior to ESI-MS.

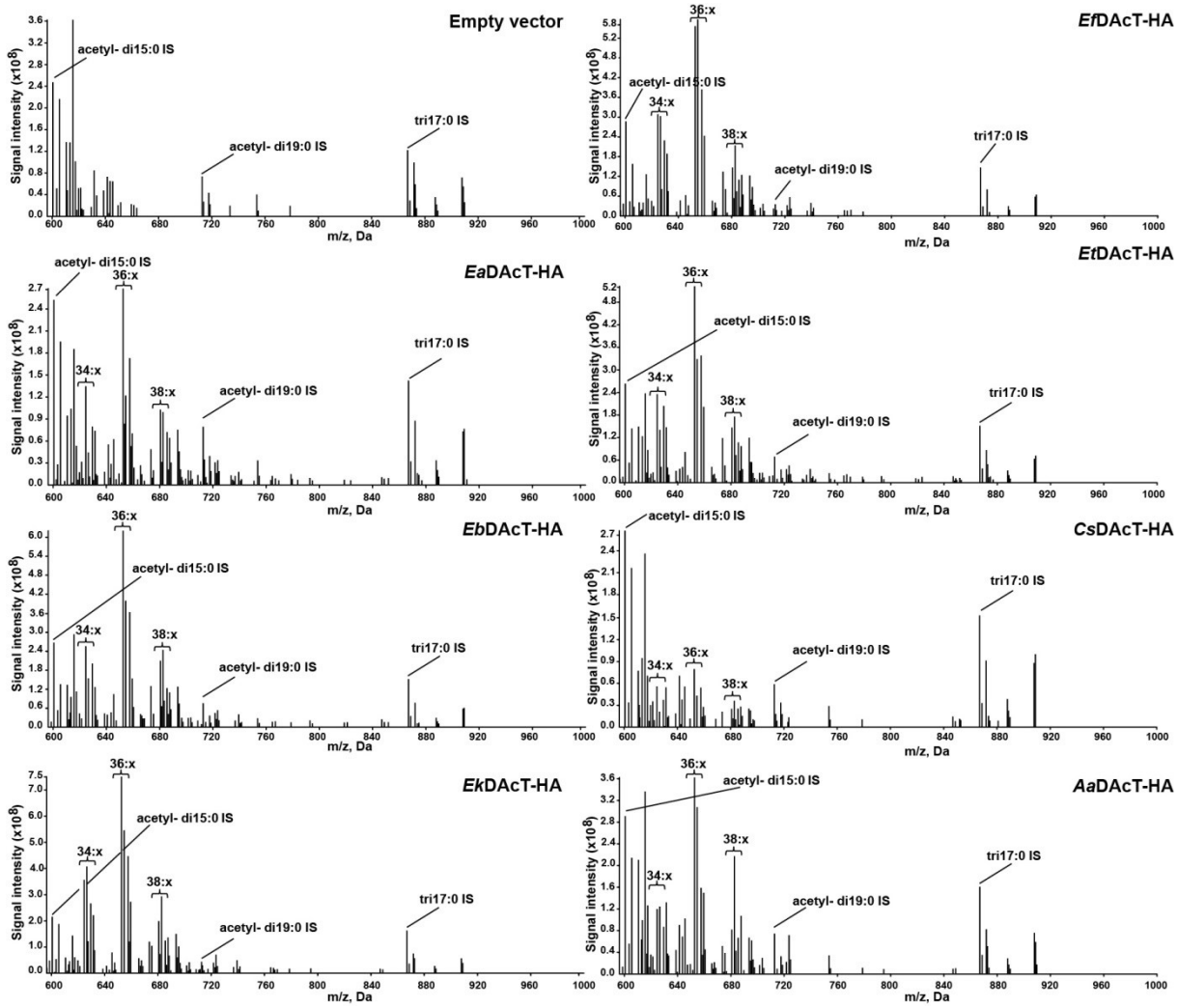

Figure 3.2 Yeast expressing *Ea*DAcT orthologs accumulate acetyl-TAG but not long chain TAG

Positive-ion electrospray ionization mass spectra of neutral lipid extracts from H1246 yeast expressing empty vector and HA-epitope tagged DAcT orthologs. Peaks correspond to m/z values of the $[M+NH_4]^+$ adducts. The 3-acetyl-1,2-dinonadecanoyl-sn-glycerol (acetyl-di19:0) internal standard was added during lipid extraction and triheptadecanoin (tri17:0) and 3-acetyl-1,2-dipentadecanoyl-sn-glycerol (acetyl-di15:0) internal standards were added prior to ESI-MS. Peaks correspond to the m/z values of the $[M+NH_4]^+$ adducts of the intact acetyl-TAG molecules. The number of acyl carbons in each cluster of peaks is indicated; for the clarity, the number of double bonds is not defined (x).

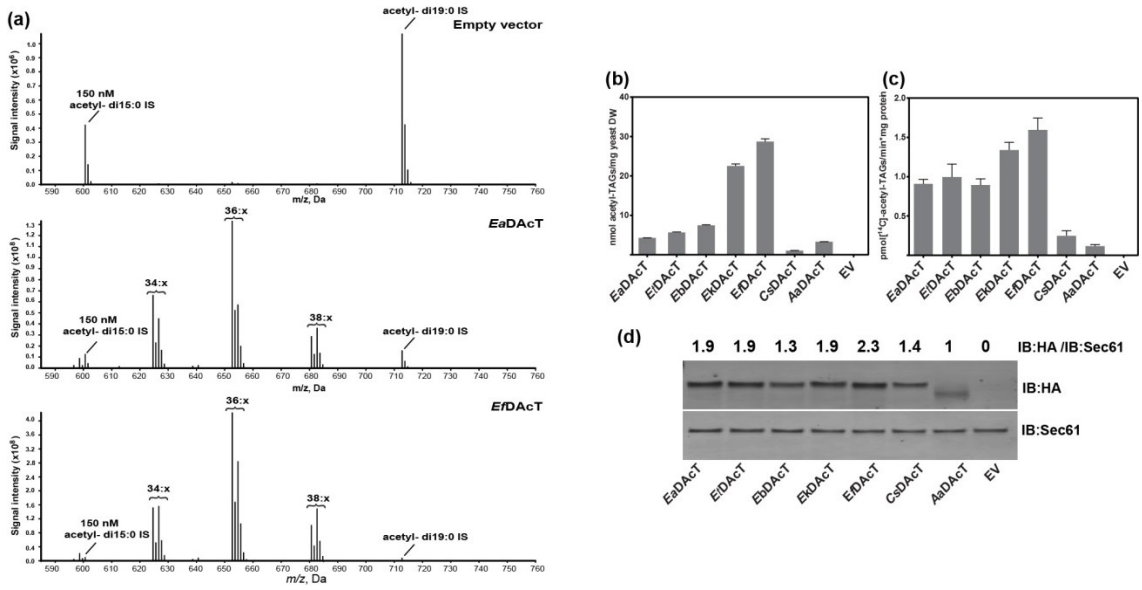

Figure 3.3 DAcT orthologs possess in vitro and in vivo acetyl-TAG synthesis activity

(a) Representative mass spectra derived from the neutral loss of ammonium acetate from lipid extracts of yeast expressing the empty vector pYES2/CT, *Ea*DAcT-HA or *Ef*DAcT-HA. The internal standard 3-acetyl-1,2-dinonadecanoyl-*sn*-glycerol (acetyl-di19:0) was added during lipid extraction and 3-acetyl-1,2-dipentadecanoyl-*sn*-glycerol (acetyl-di15:0) was added during ESI-MS analysis. Peaks correspond to the *m/z* values of the $[M+NH_4]^+$ adducts of the intact acetyl-TAG molecules. The number of acyl carbons in each cluster of peaks is indicated; the number of double bonds is not defined (x). (b) Acetyl-TAGs produced by yeast expressing DAcT orthologs were quantified using ESI-MS. Values represent the mean ± S.D. of acetyl-TAG content derived from three different ESI-MS analyses and are representative of at least two replicate cultures. (c) In vitro DAG acetyltransferase activity of DAcT enzymes from different plant species. Yeast microsomes expressing different DAcT orthologs were incubated with [^{14}C]acetyl-CoA and resulting [^{14}C]acetyl-TAG production quantified. Error bars indicate the standard deviation of 3 technical replicates. The results shown are representative of two independent experiments. (d) Expression levels of different DAcT orthologs in yeast microsomes. The abundance of different HA epitope-tagged DAcT enzymes was examined by immunoblotting for the HA epitope or for Sec61 as a loading control. Numbers indicate the ratio of DAcT proteins to Sec61 protein.

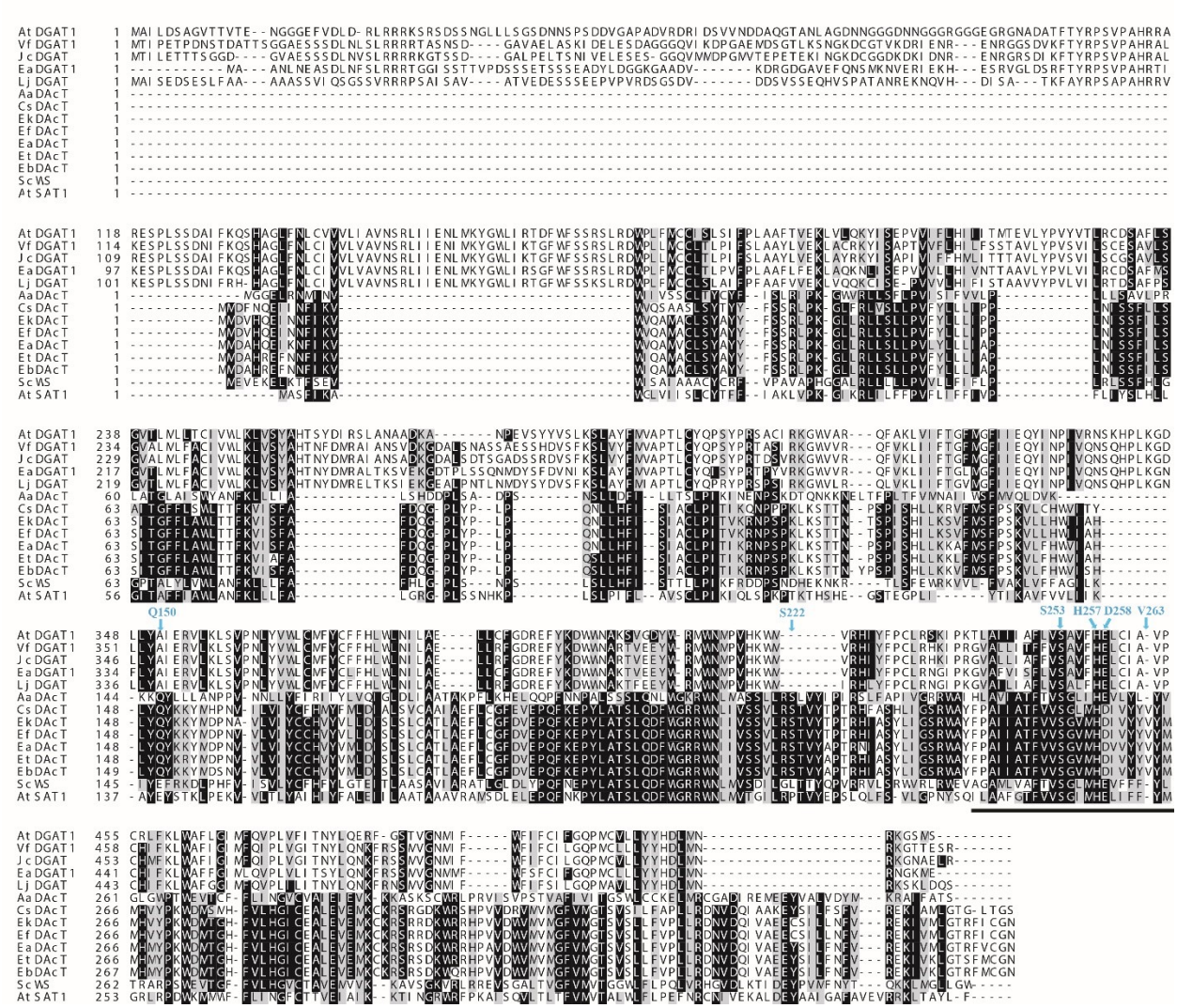

Figure 3.4 Multiple sequence alignment of plant MBOATs with different substrate specificities

Protein sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE). Residues similar to consensus are shaded in grey and residues identical to consensus are indicated by white letters on a black background. The MBOAT signature region is underlined. Arrows indicate residues at which mutagenesis was conducted. Proteins aligned were DAcT, DGAT, wax synthase (WS) and sterol O-acyltransferase 1 (ASAT1) sequences from *Adonis aestivalis* (*Aa*); *Arabidopsis thaliana* (*At*); *Celastrus scandens* (*Cs*); *Euonymus alatus* (*Ea*); *Euonymus atropurpureus* (*Et*) ; *Euonymus bungeanus* (*Eb*); *Euonymus kiautschovicus* (*Ek*); *Euonymus fortunei* (*Ef*); *Jatropha curcas* (*Jc*); *Lotus japonicas* (*Lj*); *Simmondsia chinensis* (*Sc*); and *Vernicia fordii* (*Vf*).

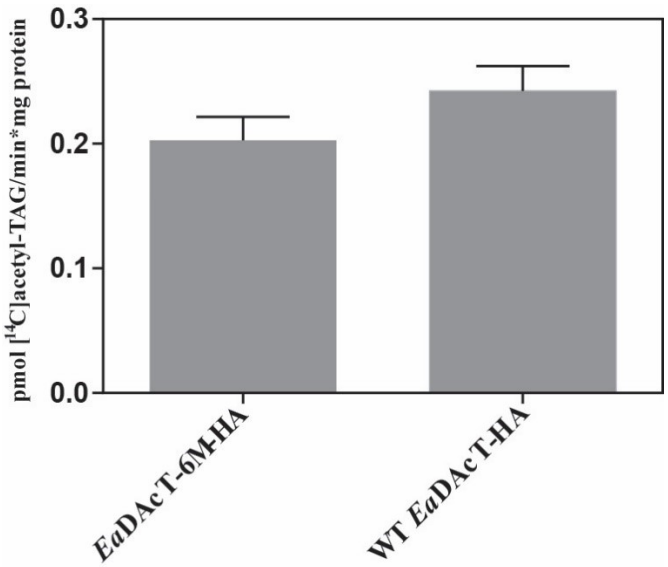

Figure 3.5 In vitro acetyltransferase activity of the *Ef*DAcT like *Ea*DAcT-6M-HA mutant

Microsomes were isolated from yeast expressing the *Ea*DAcT-6M-HA mutant and wild-type *Ea*DAcT-HA. In vitro acetyltransferase activity was quantified by incubating the microsome with [^{14}C]acetyl-CoA and quantifying the synthesis of [^{14}C]acetyl-TAG.

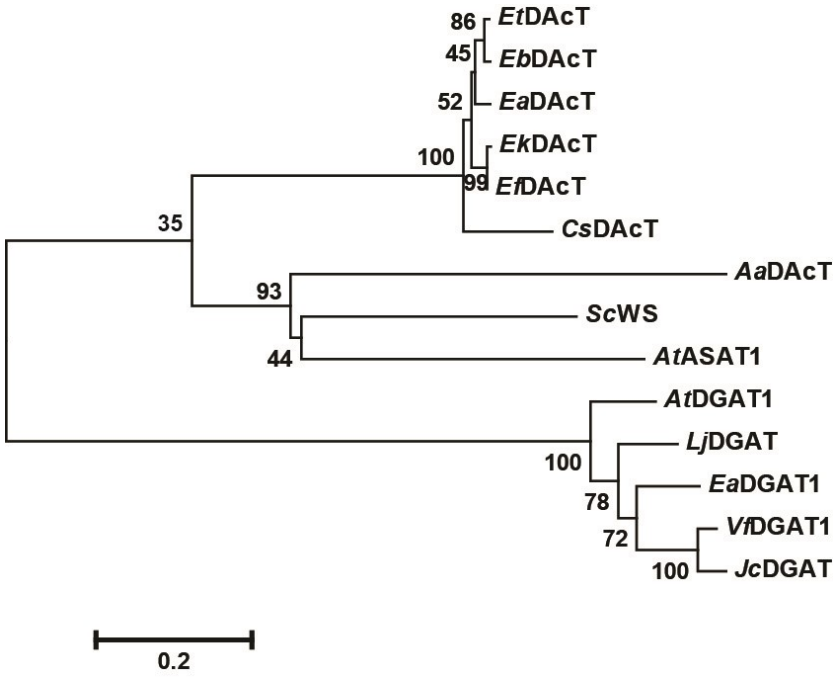

Figure 3.6 Phylogenetic tree of DAcT, *Sc*WS, *At*ASAT1 and DGAT1 enzymes

Evolution analyses were conducted in MEGA6. The evolutionary history was inferred using the neighbor-joining method and the evolutionary distances were computed using the Poisson correction method. Bootstrap values are shown in percentage at nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The 0.2 scale represents 20% change. Proteins used were DAcT, DGAT, wax synthase (WS) and sterol O-acyltransferase 1 (ASAT1) sequences from *Adonis aestivalis* (*Aa*)*; Arabidopsis thaliana* (*At*)*; Celastrus scandens* (*Cs*)*; Euonymus alatus* (*Ea*)*; Euonymus atropurpureus* (*Et*) *; Euonymus bungeanus* (*Eb*)*; Euonymus kiautschovicus* (*Ek*)*; Euonymus fortunei* (*Ef*)*; Jatropha curcas* (*Jc*)*; Lotus japonicas* (*Lj*)*; Simmondsia chinensis* (*Sc*)*;* and *Vernicia fordii* (*Vf*).

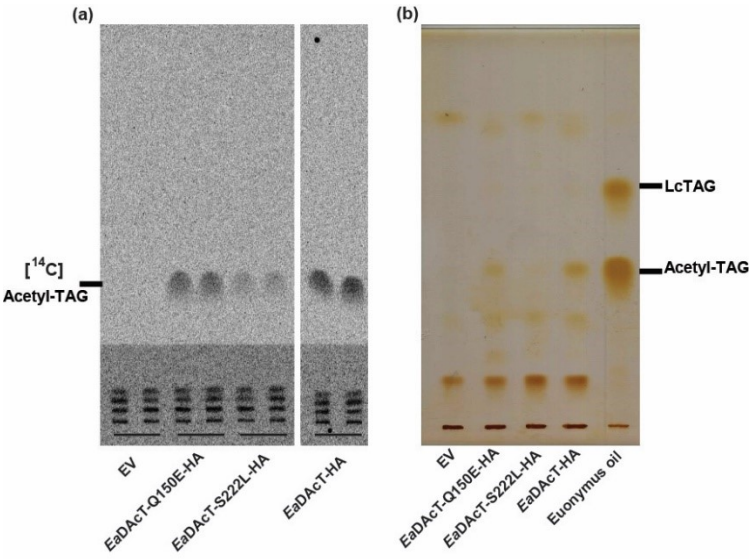

Figure 3.7 Q150 and S222 are required for *Ea*DAcT activity

(a) Autoradiogram of TLC separated lipid extract from H1246 yeast microsomes expressing empty vector, *Ea*DAcT-Q150E, *Ea*DAcT-S222L and the native *Ea*DAcT incubated with [^{14}C]acetyl-TAG. Samples were separated on the same TLC; lanes not relevant for this experiment were removed from the image. (b) TLC separation of lipid extracts from H1246 yeast expressing empty vector, *Ea*DAcT-Q150E, *Ea*DAcT-S222L, and native *Ea*DAcT demonstrating no long chain TAG were produced in the mutants. Loading with *Euonymus alatus* oil was used as TLC marker for acetyl-TAG and long chain TAG.

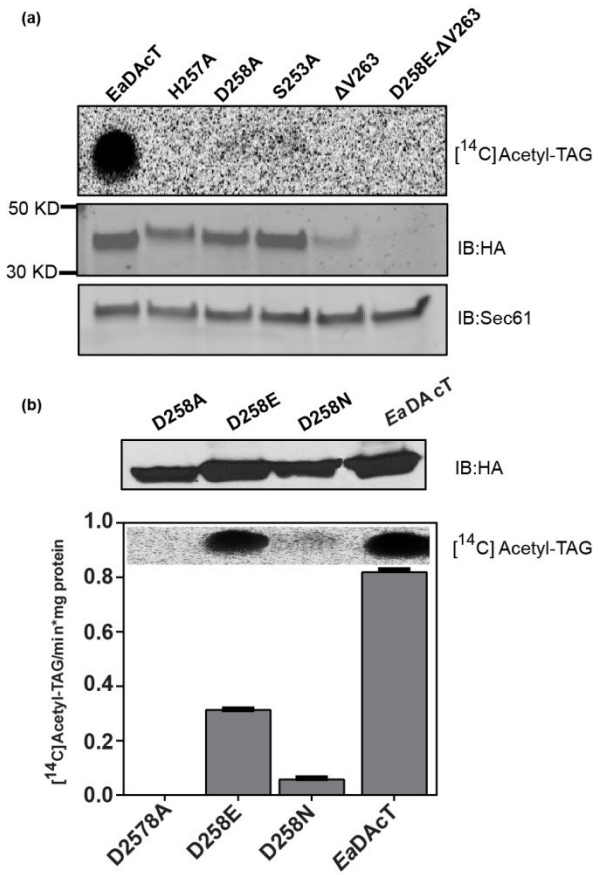

Figure 3.8 Identification of residues important for *Ea*DAcT activity

(a) Effect of H257A, D258A, S253A, and ΔV263 mutations on acetyltransferase activity. Microsomes were isolated from yeast expressing different mutants and wild-type *Ea*DAcT-HA. In vitro acetyltransferase activity was quantified by incubating with [^{14}C]acetyl-CoA and detecting [^{14}C]acetyl-TAG (upper panel). Protein levels of the *Ea*DAcT mutants in the microsome fraction were detected by immunoblotting for the HA epitope (middle panel). Sec61 was used as a loading control (lower panel). (b) Effect of different substitutions for D258 on *Ea*DAcT activity. Microsomes were isolated from yeast expressing different mutants and wild-type *Ea*DAcT-HA. In vitro acetyltransferase activity was quantified by incubating with [^{14}C]acetyl-CoA and detecting [^{14}C]acetyl-TAG (lower panel). Protein levels of the *Ea*DAcT variants in the microsome fraction were detected by immunoblotting for the HA epitope (upper panel).

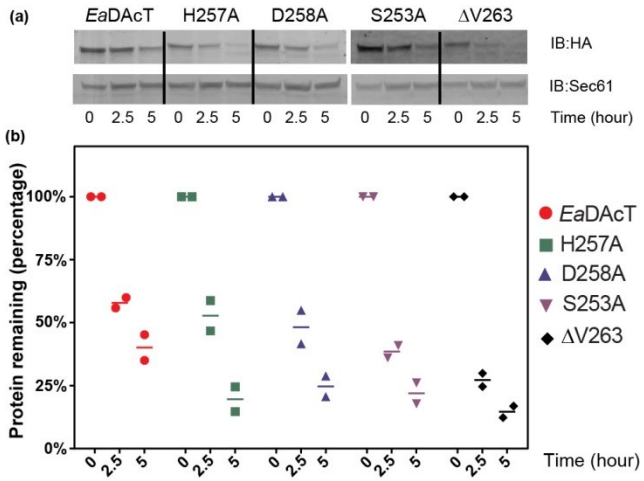

Figure 3.9 Mutations at the MBOAT domain affect stability of *Ea*DAcT

(a) Effects of H257A, D258A, S253A, and ΔV263 mutations on stability of *Ea*DAcT. Galactose induced expression of different mutants and wild-type *Ea*DAcT-HA in yeast was inhibited by switching the carbon source of the culture medium to glucose. Microsomal proteins were isolated at different time points after switching carbon source and protein levels of the mutants and wild-type *Ea*DAcT-HA in the microsome fraction were detected by immunoblotting for the HA epitope (top panel). Sec61 was used as a loading control (lower panel). Representative data of 2 independent cultures are shown. (b) Quantified protein remaining after switching carbon source. Protein levels at zero time point were set at 100%. Horizontal lines represent mean values derived from two independent cultures.

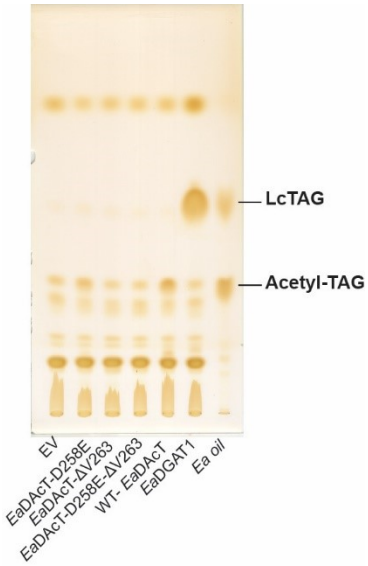

Figure 3.10 *Ea*DAcT-D258E, *Ea*DAcT-ΔV263 and *Ea*DAcT-D258E-ΔV263 do not synthesize long chain TAG

TLC separation of lipid extracts from H1246 yeast expressing empty vector, *Ea*DAcT-D258E, *Ea*DAcT-ΔV263, *Ea*DAcT-D258E-ΔV263, native *Ea*DAcT and *Ea*DGAT1 demonstrating no long chain TAG were produced by the mutant enzymes. Loading with *Euonymus alatus* oil was used as TLC marker for acetyl-TAG and long chain TAG.

Chapter 4 - Characterization of novel wax synthase/Acyl-CoA:diacylgycerol acyltransferase members of the MBOAT family

4.1 INTRODUCTION

Lipids are the major energy storage in angiosperm seeds. Most plants accumulate triacyglycerol (TAG), an ester of glycerol and 3 fatty acids, as their main lipid reserve. Due to the higher reduced state of TAG compared to carbohydrate and protein, the complete oxidation of TAG yields more energy than carbohydrate or protein. Indeed, TAG is highly anhydrous which allows compact storage of energy (Berg et al., 2002). The last committed step of TAG biosynthesis is catalyzed by acyl-CoA dependent and acyl-CoA independent enzymes including phospholipid:diacylglycerol acyltransferases, diacylglycerol transacylases, and diacylglycerol acyltransferases (Li-Beisson et al., 2013).

Diacylglycerol acyltransferases (DGATs) are acyl-CoA-dependent enzymes that transfer an acyl group from acyl-CoA to the *sn-3* position of the glycerol backbone on a diacylglycerol (DAG) molecule. Several classes of DGATs exist. DGAT1 enzymes belong to the membrane bound O-acyltranferase (MBOAT) enzyme family of multipass membrane proteins that acylate a variety of lipid and non-lipid substrates (Hofmann, 2000). DGAT2 proteins do not show homology to DGAT1 and different plant DGAT2 enzymes appear to have preference for the incorporation of unusual fatty acid containing substrates into DAG (Kroon et al., 2006; Li et al., 2010). Monoacylglycerol acyltransferases were identified by their high degree of similarity to DGAT2. They are mainly responsible for the synthesis of DAG but also possess TAG synthesis activity (Yen et al., 2002; Yen and Farese, 2003; Cao et al., 2007). Bifunctional WS/DGAT enzymes were found in some bacteria and *A. thaliana* with dual activity toward synthesizing wax ester (WE) and TAG. They are either membrane associated or cytoplasmic localized enzymes distinct from eukaryotic DGAT1 and DGAT2 (Kalscheuer and Steinbuchel, 2003; Daniel et al., 2004; Li et al., 2008). A group of DGATs that share little identity to DGAT1, DGAT2 and WS/DGAT were discovered in several plants (Saha et al., 2006; Panikashvili et al., 2009). The existence of these soluble enzymes demonstrates that, while TAG biosynthesis typically occurs in cellular membranes, the process can take place in the cytosol of some organisms.

The *Euonymus alatus* diacylglycerol acetyltransferase (*Ea*DAcT) is an unusual DGAT in that it catalyzes the transfer of the shortest possible acyl group from acetyl-CoA to the *sn*-3 position of a diacylglycerol (DAG) molecule to form 3-acetyl-1,2-diacyl-*sn*-glycerol (acetyl-TAG) (Durrett et al., 2010). The enzyme also possesses a weak WS activity by acetylating fatty alcohols to form alkyl acetates (Bansal and Durrett, 2016a; Ding et al., 2016). The presence of the acetyl group instead of a longer fatty acyl group gives acetyl-TAG reduced kinematic viscosity and lower melting temperature compared to regular triacylglycerols (herein referred to as lcTAG) (Durrett et al., 2010; Marshall et al., 2014). Thus, acetyl-TAG is suitable for many applications including food and industrial chemicals (Liu et al., 2015). Field production of acetyl-TAG by transforming *Ea*DAcT into the oilseed crop *Camelina sativa* has been successfully achieved. The transgenic lines produced 85 mol% acetyl-TAG in the oil, demonstrating that the unusual lipid can be attained at a high level (Liu et al., 2015). We hypothesized that the level of acetyl-TAG in transgenic plants can be further increased by transforming the plants with acetyltransferase enzymes possessing increased activities. Therefore, we conducted experiments to isolate DAcT enzymes capable of synthesizing higher level of acetyl-TAG compared to *Ea*DAcT.

In addition to *Euonymus alatus*, many different plant species have been reported to accumulate acetyl-TAG in their seeds (Bagby and Smith, 1967; Kleiman et al., 1967). For instance, seed lipids of *Adonis aestivalis*, *Sorbus aucuparia* and *Akebia quinata* contain both acetyl-TAG and lcTAG (Chapter 4). We sequenced RNA extracted from the seeds in order to identify enzymes responsible for the synthesis of acetyl-TAG in these plants. A BLAST search for proteins with high similarity to *Ea*DAcT in the transcriptomic data base was performed and genes that encode putative DAcT were cloned and further studied. We discovered 3 enzymes that, instead of acetylating DAG to form acetyl-TAG, catalyzed the transfer of a long chain acyl group from acyl-CoA to fatty alcohols and diacylglycerols to form wax esters (WE) and lcTAG, respectively. As MBOATs, these novel enzymes are distinct from bifunctional WS/DGAT enzymes described previously in bacteria and *A. thaliana*. Here, we report the isolation and biochemical characterization of these new enzymes.

4.2 MATERIALS AND METHODS

Seed collection and RNA isolation was previously described in Chapter 3.

Succrose gradient fractionation, immunoblotting and protein quantification procedures can be found in Chapter 2.

Details of Illumina sequencing and computational analysis were described in Chapter 3.

4.2.1 Synthesis of pentadecanoyl tridecanoate WE standard

78 mmoles of tridecanol alcohol (Nu Chek Prep, Inc., Waterville, MN) was mixed with 38 mmoles of pentadecanoyl chloride (Nu Chek Prep, Inc., Waterville, MN) at room temperature. A few particles of 4-dimethylaminopyridine (Sigma Aldrich, St. Louis, MO) were added to the mixture during mixing. After 7 h, 1 ml of water was added, the resultant lipids were extracted twice with 2 ml chloroform and dried under nitrogen gas. Dried lipids were suspended in chloroform and separated on Silica gel 60 TLC plate (Merck) using hexane:ether:acetic acid, 70:30:1 (v/v/v). Lipids were stained with 2',7'-dichlorofluorescein (Sigma Aldrich, St. Louis, MO) in methanol (0.075%:95%, m/v). The pentadecanoyl tridecanoate WE band was scraped and extracted with hexane:ether, 1:1 (v/v), and the concentration of the product was quantified by GC-FID.

4.2.2 Cloning and expression of WS/DGAT bifunctional proteins

Total RNA isolated from *A. aestivalis*, *S. aucuparia* and *A. quinata* seeds were used for first strand cDNA synthesis of *Aa*WS/DGAT, *Sa*WS/DGAT and *Aq*WS/DGAT, respectively, using oligo dT primer and reverse transcriptase (SuperScript III, Invitrogen). Primers specific to *Aa*WS/DGAT, *Sa*WS/DGAT and *Aq*WS/DGAT (Table 4.1) were used for amplification of the gene coding sequences from cDNAs using Phusion polymerase (Thermo Fisher Scientific, MA, USA). The genes were cloned into an engineered pYES2/CT vector that has the hemagglutinin (HA) epitope tag coding sequence inserted at the 3' end of the multiple cloning site region of the vector (pYES2/CT-HA) generating p*Aa*WS/DGAT-HA, p*Aq*WS/DGAT-HA and pS*a*WS/DGAT-HA. For co-expression of the bifunctional proteins with the fatty acid reductase from *M. aquaeolei* (*Ma*FAR), *Aa*WS/DGAT-HA and *Sa*WS/DGAT-HA coding sequences were amplified from p*Aa*WS/DGAT-HA and p*Sa*WS/DGAT-HA, respectively, and cloned to the bi-directional galactose-inducible GAL1-GAL10 promoter region of the pESC-Ura vector (Agilent Technologies). The *Ma*FAR kindly provided by Dr. Per Hofvander (Swedish University of Agricultural Sciences, Sweden) was then cloned to downstream of GAL1 promoter region of the vector. Recombinant vectors were transformed to the yeast H1246 strain, which is devoid of

TAG biosynthesis enzymes (Sandager et al., 2002) and expression of HA fusion recombinant proteins was induced by supplementing the yeast culture medium with galactose.

4.2.3 Yeast microsome activity assays

Yeast microsomal proteins were isolated as previously described (Durrett et al., 2010). Yeast microsome activity assays were carried out following the procedure described in Chapter 2 with modifications. [Oleyl-1-^{14}C] Coenzyme A (4.6 nmoles) (Perkin Elmer,Waltham, MA) was used as an acyl donor substrate. Endogenous DAG and oleyl alcohol (0.125 mM) (Nu Chek Prep Inc., Waterville, MN) were used as acyl acceptor substrates for long chain TAG sysnthesis assay and wax ester sysnthesis assay, respectively. Dimethyl-sulfoxide (Fisher Scientific, Waltham, MA) was used to solubilize oleyl alcohol into the microsome membrane.

4.2.4 Alcohol feeding assay

Yeast expressing proteins of interest were cultured overnight on the synthetic complete selective medium lacking uracil and tryptophan that contains 2% galactose. The cells were sub-cultured in 50 ml of the same medium with the addition of 0.7% tergitol and 5 mM fatty alcohol with a starting cell optical density of 0.2. After 48 h growth at 30^{o}C, yeast were harvested for lipid analysis. Alcohols used for the alcohol feeding assay are listed in Table 4.2.

4.2.5 Electrospray ionization-mass spectrometry analysis of yeast lipids

Neutral lipids were extracted from dried yeast cells as previously described (Durrett et al., 2010). 50 nmoles of tripentadecanoin extraction standard were added to each sample prior to extraction. For electrospray ionization-mass spectrometry (ESI-MS) analysis, 1 µg of neutral lipid in 400 µl chloroform was mixed with 700 µl mass spec solvent containing chloroform:methanol:300mM ammonium acetate in water (300:665:35, v/v/v). 50 nmoles tridecanoin internal standard were spiked in every sample prior ESI-MS analysis. Neutral lipids were analyzed in positive ion mode using a triple quadrupole mass spectrometer API 4000 (Applied Biosystems, Foster City, CA) equipped with an electrospray ionization (ESI) source. The samples were directly infused at 3 µl per minute. Instrumental settings for total ion scans were as follow: curtain gas, 20 (arbitrary units); ion source gases 1 and 2, 45 (arbitrary units), ion spray voltage, 5500 V; source temperature, 100^{o}C; declustering potential, 20 V; entrance potential, 10 V; and the interface heater, “on”. Spectra were acquired from 500 to 1000 *m/z* at 5 second/cycle for 40 cycles.

The product ion scan of 846.9 in positive mode was acquired with similar instrument parameters with the exceptions of: declustering potential, 50 V; collision energy, 25 V; and collision cell exit potential, 14 V. The spectrum of fragments from 846.9 was acquired from 100 to 900 *m/z* for 45 cycles at 5 second/cycle. Multi-channel analyzer (MCA) was active.

The product ion spectrum of 550.7 was acquired from 100 to 600 *m/z* with similar instrument parameters, with the exceptions of: declustering potential, 26 V; entrance potential, 4.5 V; collision energy, 26 V; and collision cell exit potential, 15 V.

ESI-MS spectra data were processed using a custom "MultiplePeriodProcessing" script for Analyst software provided by Applied Biosystems. Spectra were smoothed one time with option of 0.4 for previous and next point weight and 1 for current point weight, and baseline subtracted, with a window of 20 u. Peaks with intensity lower than 50 counts per sec were removed.

4.2.6 GC-FID quantification of yeast lipids

Lipid extracts from yeast expressing proteins of interest were separated on Silica gel 60 thin layer chromatography (TLC) plate (Merck) with hexane:ether:acetic acid, 70:30:1 (v/v/v) and 80:20:1 (v/v/v) for lcTAG and wax ester analysis, respectively. Lipids were stained with dichlorofluorescein and visualized under UV light (312 nm). Triheptadecanoin standard (50 nmoles) and pentadecanoyl tridecanoate (60 nmoles) were spotted onto lcTAG and WE bands, respectively, prior to scrapping. Acid catalyzed trans-esterification of lcTAG was carried out as previously described (Li et al., 2006). Derivatization of WEs was done in two steps. The first step involves converting WEs to fatty acid methyl esters (FAMEs) and fatty alcohols by trans-esterification. The fatty alcohols were then derivatized with 100 µl *N*,*O*-bis(trimethylsilyl) trifluoroacetamide (Sigma Aldrich, St. Louis, MO) and 100 µl pyridine (Sigma Aldrich, St. Louis, MO) at 110 °C for 10 min. FAMEs and derivatized fatty alcohols were quantified on a Shimadzu GC-2010 Plus gas chromatograph coupled with and flame ionization detector (GC-FID) (Shimadzu scientific instruments) equipped with a DB-23 (30.0 m length x 0.25 mm diameter x 0.25 µm film thickness) column (Agilent Technologies).

For wax ester analysis, the injector was operated in split mode with a split ratio of 5.0 and an injection volume of 1.0 µl. The flow rate of helium gas carrier was maintained at 6.6 $ml \cdot min^{-1}$. The oven temperature was kept at 150^{o}C for 1 min then ramped to 250^{o}C in 20 min

and held at 250ºC for 5 min. The detector sampling rate was set at 40 msec and data collection was stopped at 22 min.

For FAME analysis, 5 µl of each sample was injected and split at the ratio of 10:2. The flow rate of helium gas carrier was kept at 18.5 ml·min^{-1}.The oven temperature was maintained at 190ºC for 2 min and ramped to 240ºC at 5ºC per min and held at 240ºC for 5 min. Detector sampling rate was set at 40 msec and data collection was stopped at 17 min.

4.2.7 Analysis of total lipid of Camelina seeds

Analysis of total lipid of camelina seeds was carried out as previously described (Browse et al., 1986) with modifications. Preweighed seeds (10 seeds) were ground with a polytron in 1.5 ml toluene. Triheptadecanoin (Nu Chek Prep Inc., Waterville, MN) internal standard (5 µl of 10 mg/ml stock) was added to each sample prior to trans-methylation with 2 ml of 2.5% H_2SO_4 in methanol at 80ºC for 1 h. 400 µl toluene and 1 ml water were added to each tube and a vigorous voxtex was applied, followed by centrifugation. The upper layer was collected for FAME analysis.

4.2.8 Construction of plant transformation vectors and generation of *Camelina sativa* transgenic plants

The pBinGlyRed3 binary vector backbone used for constructing plant transformation vector was kindly provided by Dr. Edgar Cahoon (University of Nebraska-Lincoln). Sequences that encode for HA fusion *Aa*WS/DGAT and *Sa*WS/DGAT proteins were amplified by Phusion polymerase using pYES2/CT-*Aa*WS/DGAT-HA and pYES2/CT-*Sa*WS/DGAT-HA, respectively, and ligated downstream of the glycinin seed specific promoter on the pBinGlyRed3 binary vector. The binary vector contains the fluorescent protein DsRed, providing a visual selection marker for transgenic seeds. Sequence confirmed T-DNA constructs were introduced into *Agrobacterium tumefaciens* strain GV3101 kindly provided by Dr. Kathrin Schrick (Kansas State University). Agrobacterium-mediated tranformation of *Camelina sativa* 'Suneson' was carried out using a floral vaccuum infiltration procedure (Lu and Kang, 2008), and transgenic seeds were selected by their red fluorescence.

4.3 RESULTS

4.3.1 Novel bifunctional WS/DGAT enzymes synthesize lcTAG and wax ester.

We expressed 3 proteins that share high similarity to *Ea*DAcT isolated from *Adonis aestivalis*, *Sorbus aucuparia* and *Akebia quinata* in the H1246 yeast strain unable to synthesize

TAG. These proteins also exhibit a high degree of similarity to jojoba wax synthase (*Sc*WS). Microsomal vesicles from yeast expressing the proteins were extracted and used for in vitro assays. When the microsomal vesicles were incubated with radioactive labeled acetyl-CoA, we were not able to detect radioactive labeled acetyl-TAG, indicating that the proteins were unable to acetylate DAG (Figure 4.1a, left panel). When the protein-containing microsomal vesicles were incubated with radioactive labeled oleyl-CoA in the presence or absence of oleyl alcohol, we detected radioactive labeled oleyl oleate wax ester or lcTAG, respectively, in the resultant lipids (Figure 4.1a, right and middle panel). The results demonstrated that the proteins are able to catalyze the transfer of a long chain acyl group from acyl-CoA to a fatty alcohol or a diacylglycerol synthesizing wax esters (WE) or lcTAG, respectively. Thus, we annotated these novel proteins as bifunctional wax ester synthase/acyl-CoA: diacylglycerol acyltransferases. Accumulation of the bifunctional proteins in the yeast microsomes was varied. *Aa*WS/DGAT and *Sa*WS/DAT accumulated at higher levels than *Aq*WS/DGAT and *Sc*WS (Figure 4.1b). Therefore, we focused on *Aa*WS/DGAT and *Sa*WS/DGAT but not *Aq*WS/DGAT in further characterization experiments.

4.3.2 Yeast transformed with *Aa*WS/DGAT, *Sa*WS/DGAT accumulate long chain TAG.

We expressed *Aa*WS/DGAT and *Sa*WS/DGAT in H1246 yeast unable to synthesize TAG and analyzed lipids extracted from the yeast. TLC analysis indicated that yeast expressing *Aa*WS/DGAT or *Sa*WS/DGAT accumulate high levels of lcTAG but not acetyl-TAG (Figure 4.2). These results were confirmed by electrospray ionization mass spectrometry (ESI-MS). Neutral lipid fractions from yeast expressing the proteins revealed multiple peaks corresponding to ammonium adducts of different lcTAG molecular species (Figure 4.3a). ESI-MS^2 analysis of the 846.9 (*m/z*) peak resulted in characteristic fragments of the lcTAG: 575.7 *m/z* [$M+NH_4^+$ - C16:1 $RCOO^-$]; 547.6 *m/z* [$M+NH_4^+$- C18:1 $RCOO^-$] (Figure 4.3b). These results demonstrated that *Aa*WS/DGAT and *Sa*WS/DGAT can synthesize lcTAG in vivo. Quantification of accumulated lcTAG resulted in values of 3.6 nmoles/mg DW and 1.74 nmoles/mg DW for *Sa*WS/DGAT and *Aa*WS/DGAT, respectively (Figure 4.4a). The fatty acid composition of lcTAG from yeast expressing *Sa*WS/DGAT and *Aa*WS/DGAT is dominately palmitoleic acid (C16:1) and oleic acid (C18:1). Palmitic acid (C16:0) and stearic acid (C18:0) make up a smaller portion of fatty acid in the total fatty acid of lcTAG (Figure 4.4b)

4.3.3 Expression of *Aa*WS/DGAT and *Sa*WS/DGAT in yeast supplemented with fatty alcohols resulted in accumulation of wax esters.

To test whether the wax synthase activity of the proteins occurs in vivo, we co-expressed *Aa*WS/DGAT or *Sa*WS/DGAT with a fatty alcohol reductase from *M. aquaeolei* (*Ma*FAR) and analyzed yeast lipid extracts by TLC. The *Ma*FAR enzyme catalyzes the reduction of acyl-CoA or acyl-ACP to fatty alcohols (Hofvander et al., 2011) providing the acyl acceptor substrate for a wax ester synthesis reaction. We noticed that the yeast expressing *Ma*FAR showed slow growth and that this phenotype was partially rescued by coexpressing *Sa*WS/DGAT (Table 4.2). Yeast expressing *Aa*WS/DGAT or *Sc*WS with *Ma*FAR showed a slower growth than yeast expressing *Ma*FAR (Table 4.2). The co-expression resulted in additional lipid species not present in lipids of yeast expressing *Ma*FAR only (Figure 4.2). Using TLC, these unique lipid species co-migrated with an oleyl oleate wax ester standard suggesting that they might be mixtures of wax esters synthesized due to activity of *Aa*WS/DGAT or *Sa*WS/DGAT. We further verified the in vivo WE synthesis activity of *Aa*WS/DGAT and *Sa*WS/DGAT by growing yeast expressing the proteins in the presence of oleyl alcohol and analyzing neutral lipids extracted from the cells by ESI-MS. Peaks at 550.7, 552.6, 522.6, 524.6 *m/z*, corresponding to oleyl oleate, stearyl oleate, palmitoleyl oleate and palmityl oleate, respectively, were detected in yeast expressing *Aa*WS/DGAT or *Sa*WS/DGAT, but not yeast expressing empty vector (Figure 4.5a, Figure 4.6). Further ESI-MS^2 analysis of the 550.7 *m/z* peak yielded a fragmentation pattern (Figure 4.5b) similar to that described previously for oleyl oleate (Iven et al., 2013).

Quantification of the WE revealed that yeast expressing *Aa*WS/DGAT accumulated the highest amount of WE followed by yeast expressing *Sa*WS/DGAT and yeast expressing *Sc*WS (Figure 4.7a). WE comprised of a 16:0 fatty acyl moiety and a 18:0 alcohol moiety were the most abundant WE synthesized by *Aa*WS/DGAT. In contrast, WE synthesized by *Sa*WS/DGAT were predominately esters of 18:1 fatty acid and 18:0 alcohol (Figure 4.7b, 4.7c)

4.3.4 Fatty alcohol substrate characterization of *Aa*WS/DGAT and *Sa*WS/DGAT

WE extracted from jojoba seeds is the main natural replacement source for spermaceti oil used for industrial applications (Iven et al., 2013). Our in vitro and in vivo data demonstrated that the bifunctional proteins possess higher WE synthesis activity than jojoba WS. We were interested in further characterizing the fatty alcohol substrate of the proteins. Yeast lipids contain mostly saturated and monounsaturated fatty acids with 16 or 18 carbons and a small amount of

medium chain fatty acids (Moss et al., 1982). Thus there is a limited substrate diversity available to test the specificity of the bifunctional proteins. We therefore added fatty alcohols with different carbon chain lengths and numbers of double bonds to the yeast growth media to provide a wider range of substrates for the enzymes (Table 4. 3). Accumulated WEs were analyzed by TLC and quantified by GC-FID. Some fatty alcohols were not soluble in the culture medium resulting in undetectable WE (Table 4.2). *Aa*WS/DGAT and *Sa*WS/DGAT showed the highest activity with myristyl alcohol (C14:0). Of the tested fatty alcohols, both enzymes exhibited a preference for monounsaturated and diunsaturated fatty alchohols (Figure 4.8). Under most conditions, yeast expressing *Aa*WS/DGAT accumulated more wax ester than yeast expressing *Sa*WS/DGAT.

4.3.5 *Camelina sativa* lines transformed with bifunctional proteins do not synthesize more lcTAG

The DGAT activity of the bifunctional proteins might be employed to produce for many applications. We expressed the enzymes under the control of the strong seed-specific glycinin promoter in the oilseed crop *Camelina sativa.* The transformation cassette also contains the DsRed fluorescent protein as a visible reporter for transgenic lines. We obtained 2 and 4 transgenic lines that express *Sa*WS/DGAT-HA and *Aa*WS/DGAT-HA, respectively. These lines possess a single transgenic insertion based on their 3:1 segregation ratio of the DsRed marker (chi square test, $p<0.5$). Total fatty acid profiles of transgenic T_2 seed of Camelina transformed with *Aa*WS/DGAT-HA or *Sa*WS/DAT-HA were comparable to those of Camelina transformed with the empty vector and the wild-type plants (Figure 4.9). In addition, no difference in WE levels were obtained in seed oil of Camelina expressing the bifunctional proteins and the wild-type plants (Figure 4.10).

4.3.6 Bi-functional proteins belong to the MBOAT family

Consistent with their identification as homologs of *Ea*DAcT, *Aa*WS/DGAT, *Sa*WS/DGAT and *Aq*WS/DGAT belong to the MBOAT family. The proteins possess the characteristic MBOAT homeodomain embedded in a highly hydrophobic region (Figure 4.11). The putative histidine active site of MBOAT is also conserved in the bi-functional proteins (Figure 4.12). In the bifunctional proteins, adjacent to this histidine residue is a conserve glutamate residue. *Ea*DAcT has an aspartate at this position instead. The additional valine residue in the MBOAT domain of *Ea*DAcT is absent in the bifunctional proteins (Figure 4.12). While the C-terminal

region of the proteins shares a high similarity and identity to MBOATs that acylate fatty alcohols and diacylglycerols, the N-terminal region present in DGAT1 is absent from WS/DGATs, *Ea*DAcT and *Sc*WS (Figure 4.12).

Phylogenetic analysis indicates that *Aa*WS/DGAT, *Sa*WS/DGAT and *Aq*WS/DGAT cluster to a plant specific clade of the MBOAT family that includes the jojoba wax synthase, Arabidopsis sterol acyltransferase and diacylglycerol acetyltransferases. The bifunctional acyltransferases are distantly related to type 1 diacylglycerol acyltransferases and not related to the bifunctional WS/DGAT enzymes isolated from *A. thaliana* and bacteria (Figure 4.11).

4.3.7 *Sa*WS/DGAT localizes to the endoplasmic reticulum

Because the bifunctional acyltransferases cluster with the endoplasmic reticulum (ER) localized *Ea*DAcT (Chapter 2) and catalyze the synthesis of TAG occurring in the ER (Huang, 1992) we hypothesized that the enzymes are also localized to the ER. Given the similarity in catalysis of *Sa*WS/DGAT, *Aa*WS/DGAT and *Aq*WS/DGAT, we suspect that these proteins may localize to the same subcellular compartment. We subjected cellular extracts from yeast expressing an epitope tagged version of *Sa*WS/DGAT to a sucrose gradient to monitor sub-cellular localization of *Sa*WS/DGAT in yeast. Fractionation of the plasma membrane protein Pma1, the ER membrane protein Sec61 and *Sa*WS/DGAT in the gradient was monitored by immunoblotting. *Sa*WS/DGAT and Sec61 were most abundant in the fourth fraction in the presence of Mg^{2+} (Figure 4.13). The accumulation of both proteins was shifted to a heavier fraction when Mg^{2+} was added into the gradient (Figure 4.13). In contrast, the plasma membrane protein was most abundant in the ninth fraction regardless of the presence or absence of Mg^{2+} (Figure 4.13). These results suggest that *Sa*WS/DGAT associates with the ER membrane, but not the plasma membrane, when expressed in yeast.

4.4 DISCUSSION

Athough *Aa*WS/DGAT and *Sa*WS/DGAT share high identity and similarity to *Ea*DAcT, they are unable to catalyze the transfer of acetyl group from acetyl-CoA to diacylglycerol in vitro and in vivo (Figure 4.1). Instead, they transfer a long acyl chain from acyl-CoA to fatty alcohol and DAG synthesizing WE and lcTAG. Unlike the WS/DGAT bifunctional protein WSD1 from Arabidopsis that did not yield lcTAG when expressed in yeast (Li et al., 2008), expression of *Aa*WS/DGAT, *Sa*WS/DGAT and *Aq*WS/DGAT in yeast devoid of TAG biosynthesis genes resulted in accumulation of WE and lcTAG (Figure 4.2). In the presence of fatty alcohols, yeast

expressing the bifunctional proteins accumulated both wax ester and lcTAG. However, the latter was produced in a reduced amount compared to that in the absence of fatty alcohol (Figure 4.2) indicating competition between fatty alcohol and diacylglycerol substrates. The competition of diacylglycerol and fatty alcohol substrate was also observed for the *Ea*DAcT enzyme which catalyzes the transfer of acetyl group to both substrate (unpublished data).To better understand substrate selection of the bifunctional proteins, a substrate titration assay should be carried out.

*Aa*WS/DGAT exhibited higher WS activity than *Sa*WS/DGAT. In contrast, *Sa*WS/DGAT is more efficient as a DGAT than *Aa*WS/DGAT (Figure 4.4a, Figure 4.7a). The fatty acid composition of lcTAG synthesized by *Aa*WS/DGAT contains high levels of monounsaturated fatty acids (Figure 4.4b). In contrast, saturated fatty acids are preferred by the enzyme for synthesis of WE (Figure 4.7b), indicating altered preference for fatty acid in formation of different products. On the other hand, lcTAG and WE synthesized by *Sa*WS/DGAT are composed of mainly monounsaturated fatty acids (Figure 4.4b, Figure 4.7b) suggesting a difference in fatty acid substrate selection between *Aa*WS/DGAT and *Sa*WS/DGAT. Although *Ma*FAR possesses a stronger preference for 18:1-CoA than 18:0-CoA (Hofvander et al., 2011), WE extracted from yeast expressing both *Ma*FAR and *Aa*WS/DGAT or *Sa*WS/DGAT is comprised of only a small amount of 18:1 fatty alchohol but a high amount of 18:0 alcohol (Figure 4.7c). It is unclear whether only a trace amount of 16:1 fatty alcohol detected in the extracted WE (Fgure 4.7c) is due to a weak preference for the fatty alcohol of the bifunctional proteins or due to the limited 16:1 alcohol produced by *Ma*FAR (Hofvander et al., 2011). The later possibility may be the case as in the presence of 16:1 fatty alcohol, yeast expressing the bifunctional proteins accumulated 16:1 fatty alcohol containing WE at a high level (Figure 4.8). Our fatty alcohol feeding assay result is consistent with a previous report for fatty alcohol selection of *Sc*WS (Lardizabal et al., 2000) with the highest preference for 18:1 and 18:2 fatty alcohol of the tested fatty alcohols (Figure 4.8). We were unable to dissolve 18:0 and long chain fatty alcohols (more than 20 carbons) in the culture medium and thus no data were obtained for those fatty alcohols. In most assays, expression of *Aa*WS/DGAT or *Sa*WS/DGAT led to accumulation of WE at a higher level compared to that from expressing *Sc*WS. This result might have partially been due to the higher accumulation of *Aa*WS/DGAT and *Sa*WS/DGAT proteins compared to *Sc*WS protein when expressed in yeast (Figure 4.1c). Since all of the assays presented in this study were done using an HA fusion form of the proteins, it is neccessary to

determine whether fusion of the epitope tag to the proteins affected translation efficiency, enzymatic activities and stability of the proteins.

WS/DAT enzymes from different organisms were shown to have biological roles. For instance, the Arabidopsis WSD1 was demonstrated to be required for stem wax ester biosynthesis of the plant (Li et al., 2008) and the WS/DGAT isolated from *Acinetobacter* sp. Strain ADP1 is the key enzyme for the biosynthesis of storage lipid in the organism (Stoveken et al., 2005). Novel wax synthase/diacylglycerol acyltransferases *Sa*WS/DGAT, *Aa*WS/DGAT and *Aq*WS/DGAT represent a new group of MBOAT acyltransferases. It is unknown whether the novel bifunctional proteins are involved in the synthesis of WE and lcTAG in the native seeds. We compared read counts of the bifunctional proteins with those of the putative DGAT1 and PDAT1 proteins, two proteins responsible for seed TAG biosynthesis in most plants (Li-Beisson et al., 2013), in the same seed sample to provide insight into the expression level of the bifunctional proteins. Our transcriptomic data revealed that *Aa*WS/DGAT was expressed at a lower level than the putative *Aa*DGAT1 and *Aa*PDAT1 (Table 4.4). In contrast, the transcript level of *Sa*WS/DGAT was higher than that of the putative *Sa*DGAT1 and *Sa*PDAT (Table 4.4). Similarly, the expression level of *Aq*WS/DGAT was the highest among *Aq*WS/DGAT, *Aq*DGAT1 and *Aq*PDAT1 (Table 4.4). These data suggest that the involvement of bifunctional proteins in TAG and WE biosynthesis may be species specific.

The N-terminal region responsible for the positive cooperative interaction of DGAT1 with acyl-CoA substrate (Weselake et al., 2006; Roesler et al., 2016) is absent in the novel bifunctional proteins suggesting that this region is not essential for acyl-CoA substrate binding of the bifunctional proteins. The invariant putative histidine embedded in the highly hydrophobic MBOAT domain (Hofmann, 2000) and the conserved serine residue essential for enzymatic acitivity of *Ea*DAcT (Chapter 3), Porcupine acyltransferase and cholesterol O-acyl transferase type 1 (Das et al., 2008; Rios-Esteves et al., 2014) are also conserved through out the novel bifunctional proteins suggesting that these residues may also be important for the bifunctional protein catalysis (Figure 4.11). The conserved glutamate residue adjacent to the putative histidine active site of the bifunctional proteins resembles that of DGAT1 but not DAcTs which have an aspartate at the position (Figure 4.11). This feature, together with the absence of the additional valine residue found in the MBOAT domain of DAcTs, further distinguish the bifunctional proteins from DAcTs (Figure 4.11).

Several reports have demonstrated ER subdomain speficific localization of TAG biosynthesis enzymes including DGAT1 and DGAT2 (Shockey et al., 2006; Kim et al., 2016). Similarly, our sucrose gradient fractionation of the novel bifunctional proteins also suggests that the proteins were co-localized with the ER membrane protein Sec61 when expressed in yeast (Figure 4.13). However, the ER retrieval motif 'YYHDL' conserved in DGAT (McCartney et al., 2004) was absent in the novel bifunctional proteins thus localization of the proteins may rely on an unknown ER localization signal (Figure 4.11). The smaller size of the bifunctional proteins compared to DGAT1 proteins makes the proteins less hydrophobic than DGAT1 proteins. This advantageous characteristic together with the dual activity of the bifunctional proteins allow them to be used as protein models for characterizing the structual features that determine the selection of fatty alcohol and diacylglycerol substrate of an acyl-CoA dependent acyltransferase.

Camelina sativa transformed with *Aa*WS/DGAT-HA or *Sa*WS/DGAT-HA did not result in increased oil content in the seed of transgenic plants (Figure 4.9). Shockey et al. suggested that TAG biosynthesis occurs at specific ER subdomains resulted in different TAG pools. The lack of effect of *Aa*WS/DGAT and *Sa*WS/DGAT on lcTAG accumulation in transgenic *C. sativa* might due to improper localization to the necessary ER subdomain dedicated to TAG biosysnthesis or insufficicent expression of the proteins in the seeds. To evaluate these possibilities, a thorough study focusing on ER subdomain localization and expression level of the bifunctional proteins when expressed in *C. sativa* should be conducted. Additionally, more transgenic plants should be analyzed to confident in results.

Several studies have shown that accumulation of WE was only observed when a wax synthase was co-expressed with a fatty acid reductase in *C. sativa* seeds (Iven et al., 2013; Zhu et al., 2016). In our initial camelina work, we did not include a fatty acid reductase in the binary vectors transformed to camelina. This explains the absence of WE on TLC analysis of lipids extracted from T_2 seeds of transgenic *C. sativa* plants expressing *Aa*WS/DGAT-HA or *Sa*WS/DGAT-HA (Figure 4.10). Nevertheless, we will analyze seed lipid extracts from homozygous T_3 seeds of transgenic *C. sativa* plants expressing *Aa*WS/DGAT-HA or *Sa*WS/DGAT-HA using ESI-MS to detect the presence of low amounts of WE if any.

Acknowledgement

I would like to thank Dr. Timothy Durrett for providing research ideas and guidance for this work. I am grateful to Dr. Randy Schekman (University of California, Berkeley) and Dr. Carolyn

Slayman (Yale School of Medicine) for kindly providing the anti-Sec61 and anti-Pma1 antibodies, respectively. I appreciate Jennifer Shelton for her help in processing the RNA sequencing data. I would like to thank Dr. Kathrin Schrick (Kansas State University) for providing the GV3101 agrobacterium strain and her critical. Thanks to Dr. Per Hofvander (Swedish University of Agricultural Sciences, Sweden) for providing the *Ma*FAR gene. I also acknowledge James Houghton for his help in the fatty alcohol feeding experiment and in generating *C. sativa* transgenic plants. I would also like to thank Dr. Larry Davis for his helpful comments and Karanbir Aulakh for analyzing the map reads of putative proteins. ESI-MS was performed at the Kansas Lipidomics Research Center Analytical Laboratory where instrument acquisition and lipidomics method development was supported by National Science Foundation (EPS 0236913, MCB 1413036, DBI 0521587, DBI1228622), Kansas Technology Enterprise Corporation, K-IDeA Networks of Biomedical Research Excellence (INBRE) of National Institute of Health (P20GM103418), and Kansas State University. This work was supported by the National Science Foundation under Award No. EPS-0903806 and matching support from the State of Kansas through the Kansas Board of Regents.

Tables and figures

Name	Purpose	Sequence (5' to 3')
5'-*Kpn*I-*Aa*WS/DGAT	Cloning *Aa*WS/DGAT to pYES2/CT-HA vector at *Kpn*I and *Xma*I site	caggtaccATGGGAGGTGAACTGAGG
3'-*Xma*I-*Aa*WS/DGAT		aattcccgggGCTTGTGGCAAAGATTG
5'-*Kpn*I-*Sa*WS/DAT	Cloning *Sa*WS/DGAT to pYES2/CT-HA vector at *Kpn*I and *Xma*I site	caggtaccATGGAGGATGAAATCTAC
3'-*Xma*I- *Sa*WS/DAT		aattcccgggACTTTTCCACCAATAAAAG
5'-*Kpn*I- *Aq*WS/DGAT	Cloning *Aq*WS/DGAT to pYES2/CT-HA vector at *Kpn*I and *Xma*I site	caggtaccATGGAAGGGGAGATCAAT
3'-*Xma*I-*Aq*WS/DGAT		aattcccgggCAAGGGAGAAGAGGACC
5'-*Xho*I-*Ma*FAR	Cloning *Ma*FAR to pESC-URA vector at *Xho*I and *Kpn*I site	tgtctcgagATGGCAATCCAGCAGG
3-*Kpn*I-*Ma*FAR		caggtaccTCATGCCGCTTTTTTACG
5'-*Kpn*I-*Sc*WS	Cloning *ScWS* to pYES2/CT-HA vector at *Kpn*I and *Xma*I site	caggtaccATGGAGGTGGAGAAGGAG
3'-*Xma*I-*Sc*WS		aattcccgggCCACCCCAACAAACC C
5'-*Not*I- *Aa*WS/DGAT	5' primer used for cloning *Aa*WS/DGAT to pESC-Ura vector at *Not*I and *Bgl*II site.	atagcggccgcATGAAGGGAGTCAATGA ATTG
5'-*Not*I- *Sa*WS/DAT	5' primer used for cloning *Sa*WS/DGAT to pESC-Ura vector at *Not*I and *Bgl*II site.	atagcggccgcATGGAGGATGAAATCTA CAAC
5'-*Not*I-ScWS	5' primer used for cloning *Sc*WS to pESC-Ura vector at *Not*I and *Bgl*II site.	atagcggccgcATGGAGGTGGAGAAG
3'-*Bgl*II-HA-pYES2/CT	3' primer used for cloning *Aa*WS/DGAT and *Sa*WS/DGAT to pESC-URA vector at *Not*I and *Bgl*II site.	gctagatctTTAAGCGTAATCTGGAACAT C
5'-*EcoR*I-*Aa*WS/DGAT	5' primer used for cloning *Aa*WS/DGAT to pBinGlyRed3 vector at *EcoR*I and *Xba*I site.	atagaattcATGAAGGGAGTCAATGAAT TG
5'-*EcoR*I- *Sa*WS/DAT	5' primer used for cloning *Sa*WS/DGAT to pBinGlyRed3 vector at *EcoR*I and *Xba*I site..	atagaattcATGGAGGATGAAATCTACA AC
3'-*Xba*I-HA-pYES2/CT	3' primer used for cloning *Aa*WS/DGAT and *Sa*WS/DGAT to pBinGlyRed3 vector at *EcoR*I and *Xba*I site.	cctctagaCTCGAGTTAAGCGTAATC

Table 4.1 Primers used for cloning acyltransferase genes.

Nucleotides in upper case denote sequences complementary to the DNA template; in lower case nucleotides denote restriction enzyme sequences or additional sequences used for cloning.

Proteins expressed in yeast	Optical density at 600 nm after 48 h culture
Empty vector	**3.34 ± 0.12**
*Ma*FAR	**0.91 ± 0.06**
*Ma*FAR-*Aa*WS/DGAT	**0.46 ± 0.01**
*Ma*FAR-*Sa*WS/DGAT	**1.51 ± 0.05**
*Ma*FAR-*Sc*WS	**0.51 ± 0.02**

Table 4.2 Optical density of yeast cell after 48 h culture

Yeast expressing empty vector or proteins of interest were cultured in defined media containing galactose (for protein induction). After 48 h, the optical density of yeast cells in the culture was measured. Shown are the average optical density values and the standard deviation values derived from 3 replicate cultures.

Common name	Soluble in culture medium	Supplier
Benzyl alcohol	Yes	Sigma Aldrich
Cinnamyl alcohol	Yes	Sigma Aldrich
3-Phenyl-1-propanol	Yes	Sigma Aldrich
1-Octanol	No	Nu Chek Prep, Inc.
Decanol	Yes	Nu Chek Prep, Inc.
Lauryl alcohol	Yes	Nu Chek Prep, Inc.
Myristyl alcohol	Yes	Nu Chek Prep, Inc.
Pamityl alcohol	Yes	Sigma Aldrich
Stearyl alcohol	No	Sigma Aldrich
Arachidyl alcohol	No	Nu Chek Prep, Inc.
Behenyl alcohol	No	Nu Chek Prep, Inc.
Palmitoyl alcohol	Yes	Nu Chek Prep, Inc.
Oleyl alcohol	Yes	Nu Chek Prep, Inc.
11-Eicosenol	Yes	Nu Chek Prep, Inc.
Erucyl alcohol	Yes	Nu Chek Prep, Inc.

Table 4.3 List of fatty alcohols used for alcohol feeding assay

Plant species	Protein	Read count
Akebia quinata	Putative DGAT1	179
	Putative PDAT1	1027
	*Aq*WS/DGAT	1172
Sorbus aucuparia	Putative DGAT1	179
	Putative PDAT1	99
	*Sa*WS/DGAT	231
Adonis aestivalis	Putative DGAT1	985
	Putative PDAT1	819
	*Aa*WS/DGAT	70

Table 4.4 Transcript abundances of putative DGAT1 and PDAT1 proteins

Amino acid sequences of *At*DGAT1 and *At*PDAT1 were blasted against the RNA sequencing databases of *Akebia quinata*, *Sorbus aucuparia* and *Adonis aestivialis* species. Putative DGAT1 and PDAT1 proteins were selected based on their highest similarity and identity to *At*DGAT1 and *At*PDAT1, respectively, compared to other proteins. Shown are the numbers of individual reads that mapped to each gene normalized to gene lengths.

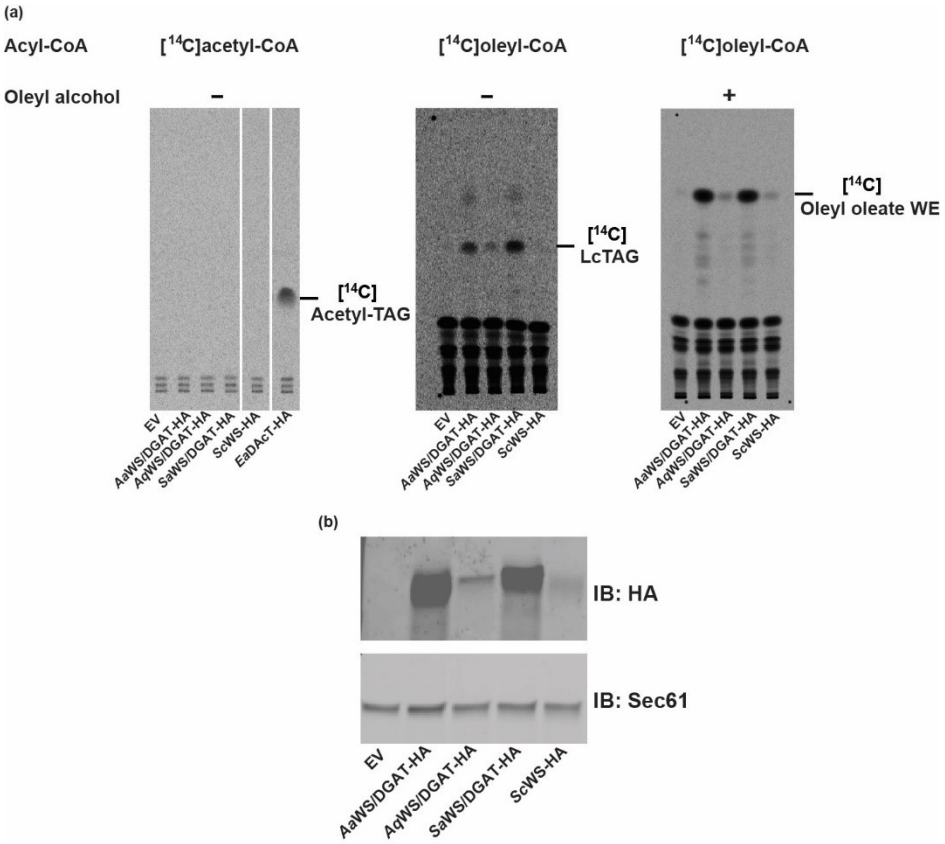

Figure 4.1 Bifunctional acyltransferases catalyze the transfer of a long acyl chain but not an acetyl group to fatty alcohols or diacylglycerols

(a) Autoradiogram of TLC separated lipid extracts from yeast microsomes expressing empty vector, *Aa*WS/DGAT, *Sa*WS/DGAT or jojoba wax synthase (*Sc*WS) incubated with different radioactive labeled acyl donor substrates in the presence or absence of oleyl alcohol. Samples not relevant for this experiment were removed from the image. (b) Expression levels of *Aa*WS/DGAT, *Sa*WS/DGAT and *Sc*WS in yeast microsomes. The abundance of different HA epitope-tagged enzymes was examined by immunoblotting for the HA epitope and for Sec61 as a loading control. Shown are representatives of at least 2 independent replicates.

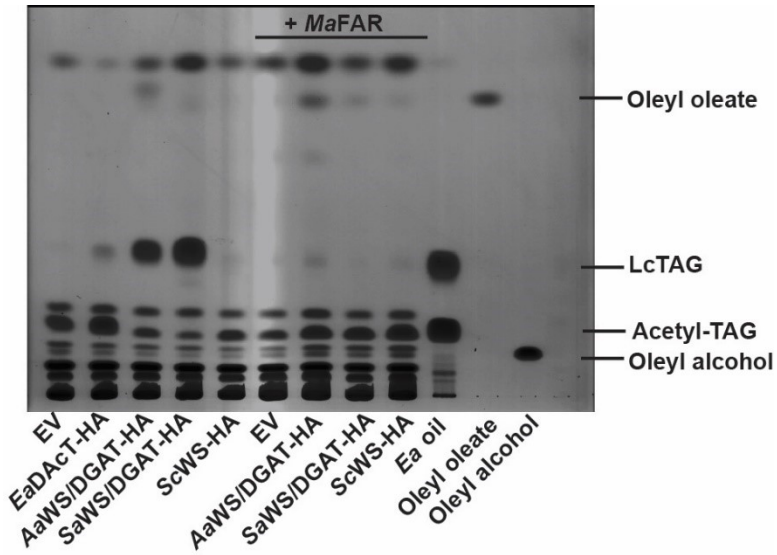

Figure 4.2 Bifunctional acytransferases synthesize wax esters and lcTAG in vivo

TLC separation of lipid extracts from yeast expressing empty vector, bifunctional acyltransferases or jojoba wax synthase (*Sc*WS) alone or with the fatty acid reductase from *M. aquaeolei* (*Ma*FAR). Oil extracted from *Euonymus alatus* seeds (*Ea* oil) was used as a TLC marker for acetyl-TAG and lcTAG. The result is representative of 3 independent cultures.

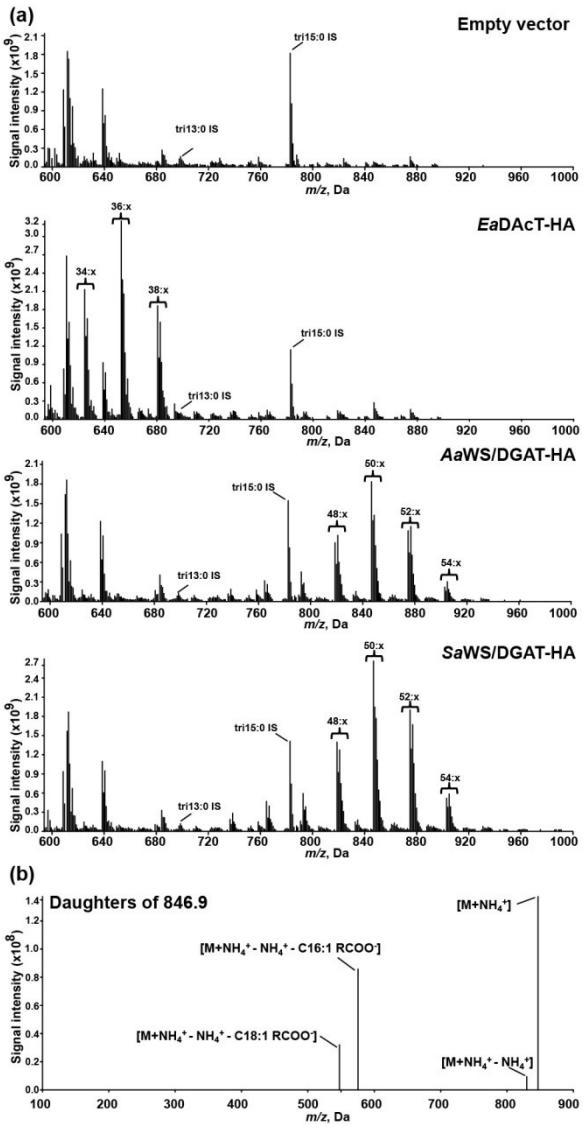

Figure 4.3 ESI-MS detection of long chain TAG in lipids of yeast expressing bifunctional proteins.

(a) Positive-ion electrospray ionization mass spectra of neutral lipid extracts from yeast expressing empty vector, bifunctional proteins or jojoba wax synthase (*Sc*WS). Tripentadecanoate (tri15:0) internal standard (IS) was added to the samples during lipid extraction and tritridecanoate (tri13:0) IS was added prior to mass spectrometry analysis. Peaks correspond to the *m/z* values of the $[M+NH_4]^+$ adducts of the intact TAG molecules. The number of acyl carbons in each cluster of peaks is indicated; for the clarity, the number of double bonds is not defined (x). (b) Representative ESI-MS^2 product ion scan of a long chain TAG from yeast expressing *Sa*WS/DGAT. Shown are fragment peaks from the long chain TAG with $[M+NH_4]^+$ adduct of mass of 846.9.

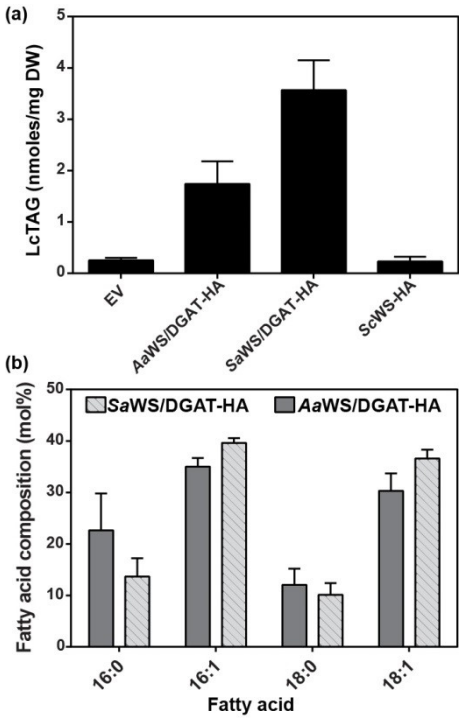

Figure 4.4 In vivo long chain TAG synthesis activity of novel WS/DGAT acyltransferases

(a) Quantification of long chain TAG produced by yeast expressing empty vector, novel WS/DGAT or jojoba wax synthase (*Sc*WS). Values represent the mean ± S.D. of long chain TAG derived from three replicate cultures. (b) Mean fatty acid composition of lcTAG from three independent yeast cultures expressing *Aa*WS/DGAT-HA or *Sa*WS/DGAT-HA. Values of the EV and *Sc*WS-HA were negligible. Error bars represent ± S.D of values derived from three replicate cultures.

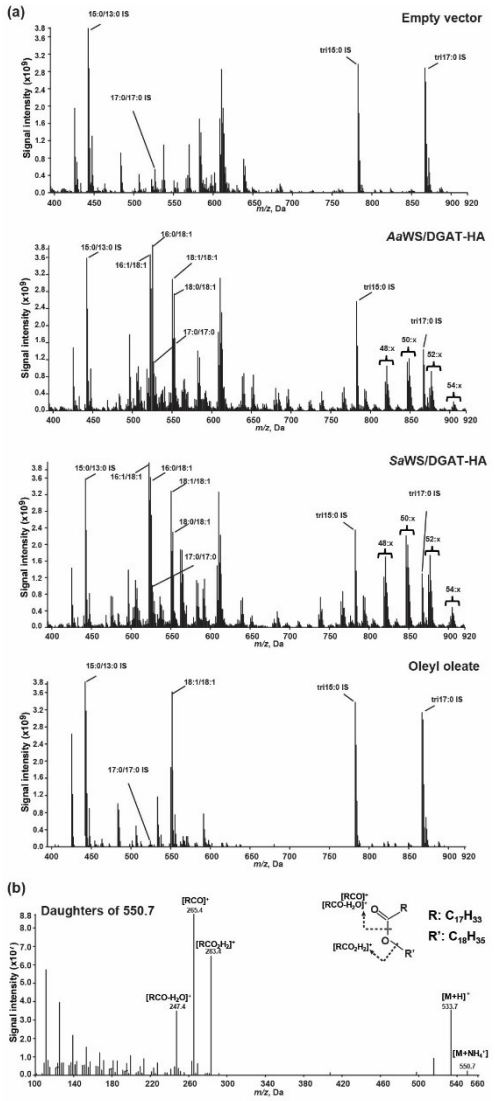

Figure 4.5 ESI-MS detection of WE in lipids of yeast expressing bifunctional proteins

(a) Positive-ion electrospray ionization mass spectra of neutral lipid extracts from yeast expressing empty vector, bifunctional acyltransferases and *Sc*WS cultured in the presence of oleyl alcohol. Heptadecanoyl heptadecanoate wax ester (17:0/17:0) and triheptadeacanoin (tri17:0) internal standards (IS) were added to the samples prior to lipid extraction. Pentadecanoyl tridecanoate wax ester (15:0/13:0) and tripentadecanoin (tri15:0) IS were added prior to mass spec analysis. A sample that contains oleyl oleate (18:1/18:1) and internal standards was run as an external positive control. (b) ESI-MS^2 product ion scan of wax ester from yeast expressing *Sa*WS/DGAT in the presence of oleyl alcohol. Shown are fragment peaks of wax ester with $[M+NH_4]^+$ adduct with mass of 550.7.

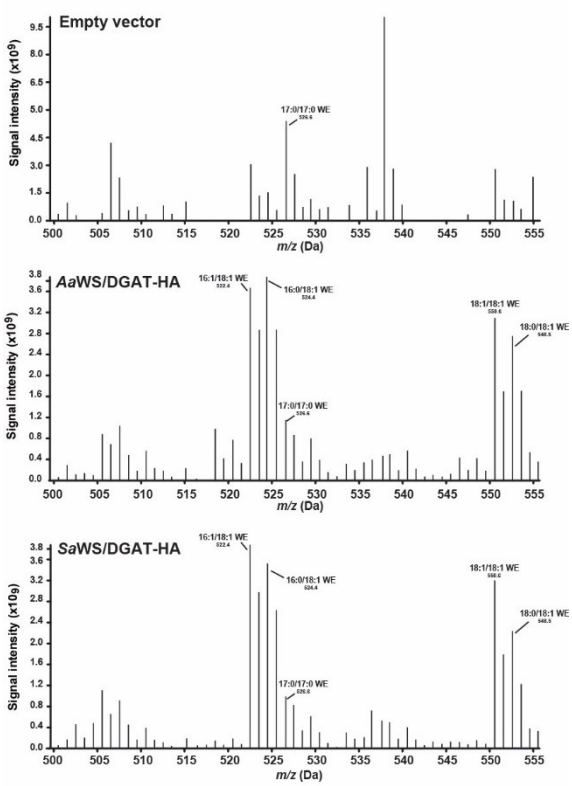

Figure 4.6 ESI-MS spectra from 500 to 555 *m/z* of lipid extracted from yeast expressing different proteins in the presence of oleyl alcohol

Positive-ion electrospray ionization mass spectra from 500 to 555 *m/z* of neutral lipid extracted from yeast expressing empty vector, bifunctional acyltransferases and *Sc*WS cultured in the presence of oleyl alcohol. Heptadecanoyl heptadecanoate wax ester (17:0/17:0) and triheptadeacanoin (tri17:0) internal standards (IS) were added to the samples prior to lipid extraction. Pentadecanoyl tridecanoate wax ester (15:0/13:0) and tripentadecanoin (tri15:0) IS were added during mass spec analysis.

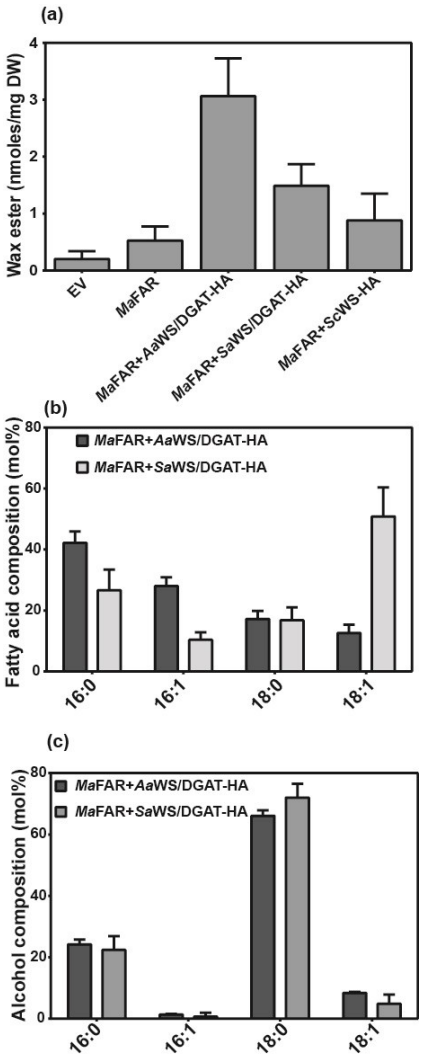

Figure 4.7 In vivo wax ester synthesis activity of novel WS/DGAT acyltransferases

(a) Quantification of wax ester produced by yeast expressing empty vector (EV), novel WS/DGAT and jojoba wax synthase (*Sc*WS) with or without *Ma*FAR. Values represent the mean ± S.D. of wax esters derived from three replicate cultures. Mean fatty acid composition (b) and mean alcohol composition (c) of wax ester from three independent yeast cultures expressing *Aa*WS/DGAT or *Sa*WS/DGAT with *Ma*FAR. Error bars represent ± S.D. of data derived from three replicate cultures.

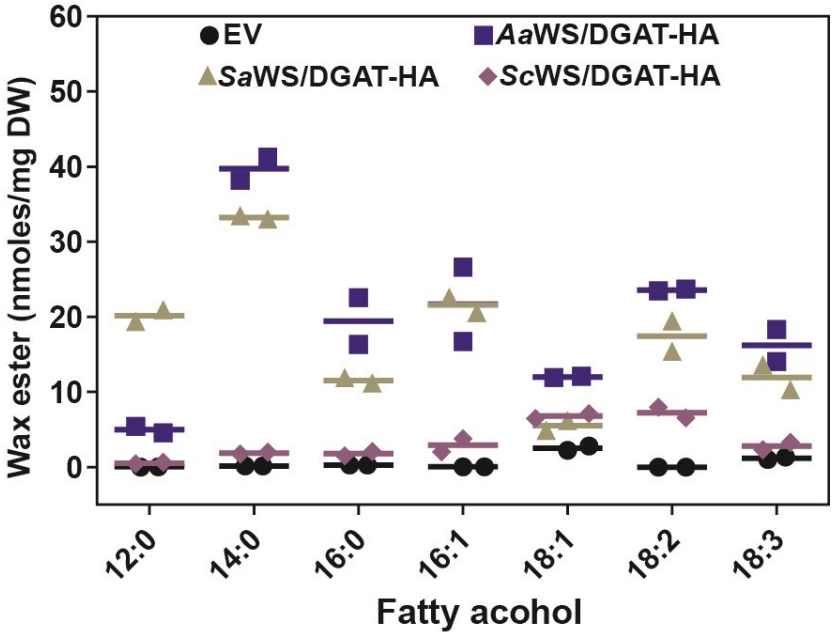

Figure 4.8 Fatty alcohol substrate selection of yeast expressing bifunctional acyltransferases

Scatter plot of WE accumulated in yeast expressing empty vector, *Aa*WS/DGAT-HA, *Sa*WS/DGAT-HA and *Sc*WS-HA in the presence of different fatty alcohols. Values derived from 2 independent cultures are shown. Horizontal lines represent mean values.

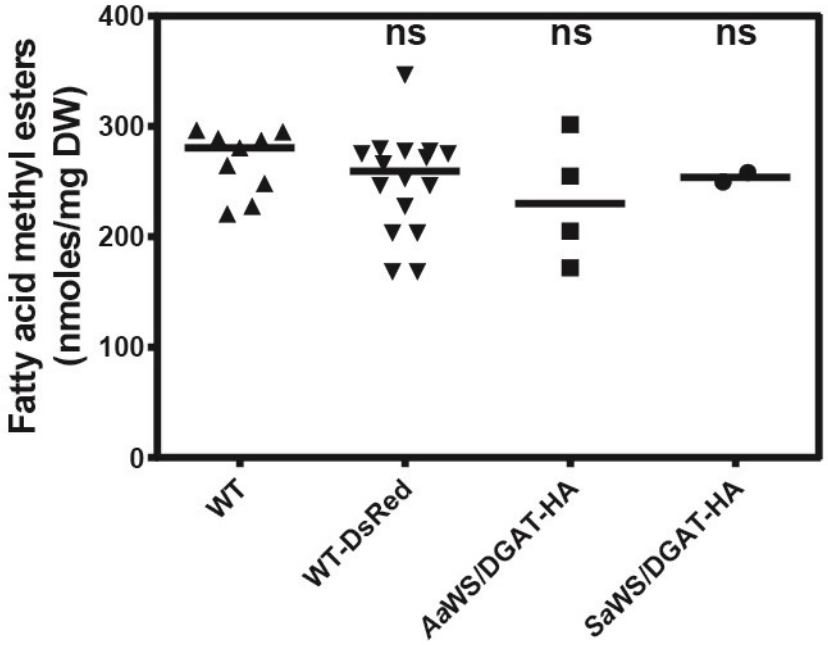

Figure 4.9 *Camelina sativa* transformed with *Sa*WS/DGAT or *Aa*WS/DGAT do not show increase in seed TAG

Scatter plot of the distribution of lcTAG in T_2 seeds from independent transgenic Camelina lines transformed with *Sa*WS/DGAT-HA, *Aa*WS/DGAT-HA or empty vector (EV). Null segregants were removed based on the absence of the visual DsRed transformation marker. Horizontal lines represent the mean value for each sample group. ns, not significant compared to the wild-type (WT) plants ($p>0.01$, Mann-Whitney *U*-test).

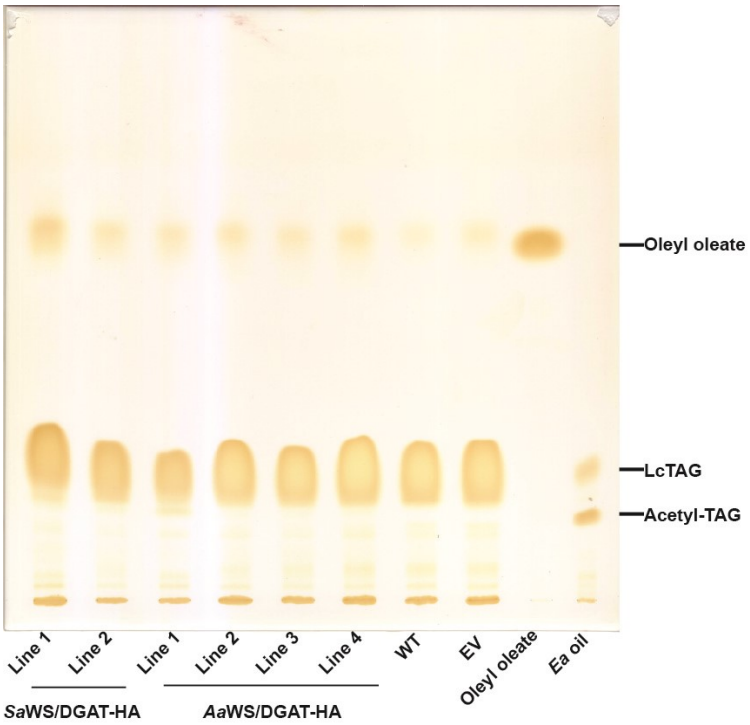

Figure 4.10 *Camelina sativa* transformed with *Sa*WS/DGAT or *Aa*WS/DGAT do not show increase in seed WE

TLC separation of lipid extracts from T_2 seeds of representative independent transgenic Camelina lines transformed with *Sa*WS/DGAT, *Aa*WS/DGAT or empty vector (EV). Null segregants were removed based on the absence of the visual DsRed transformation marker. Oil extracted from *Euonymus alatus* seeds (*Ea* oil) was used as a TLC marker for acetyl-TAG and lcTAG.

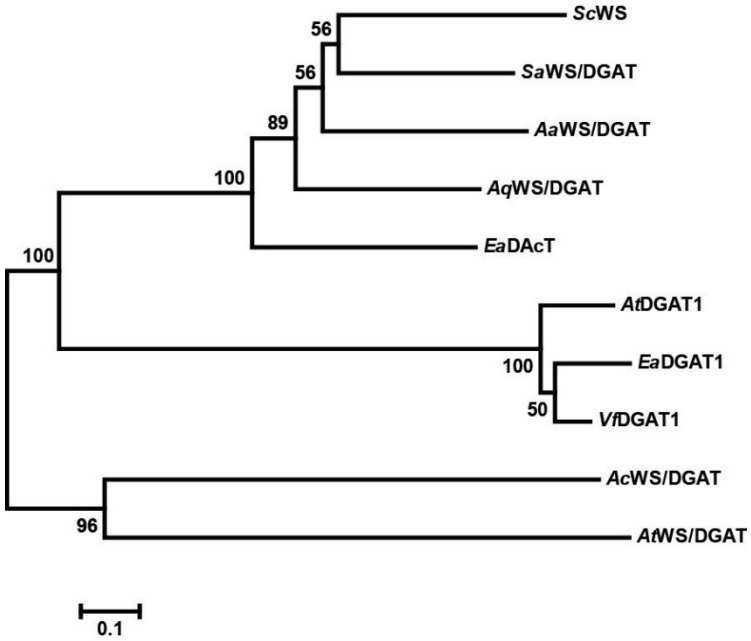

Figure 4.11 Phylogenetic comparison of MBOATs that acylate fatty alcohol or diacylglycerol and soluble WS/DGAT enzymes

Evolution analyses were conducted in MEGA6. The evolutionary history was inferred using the neighbor-joining method and the evolutionary distances were computed using the Poisson correction method. Bootstrap values are shown in percentage at nodes. The tree is drawn to scale, with branch lengths proportional to evolutionary distance. The 0.1 scale represents 10% change. Protein sequences used were DAcT, WS, DGAT1, and WS/DGAT. Abbreviations for each polypeptide indicating the organism of origin are: *Adonis aestivalis* (*Aa*)*; Acinobacter calcoaceticus (Ac); Akebia quinata* (*Aq*); *Arabidopsis thaliana (At), Euonymus alatus* (*Ea*)*; Sorbus aucuparia* (*Sa*); *Simmondsia chinensis* (*Sc*); *Vernicia fordii* (*Vf*).

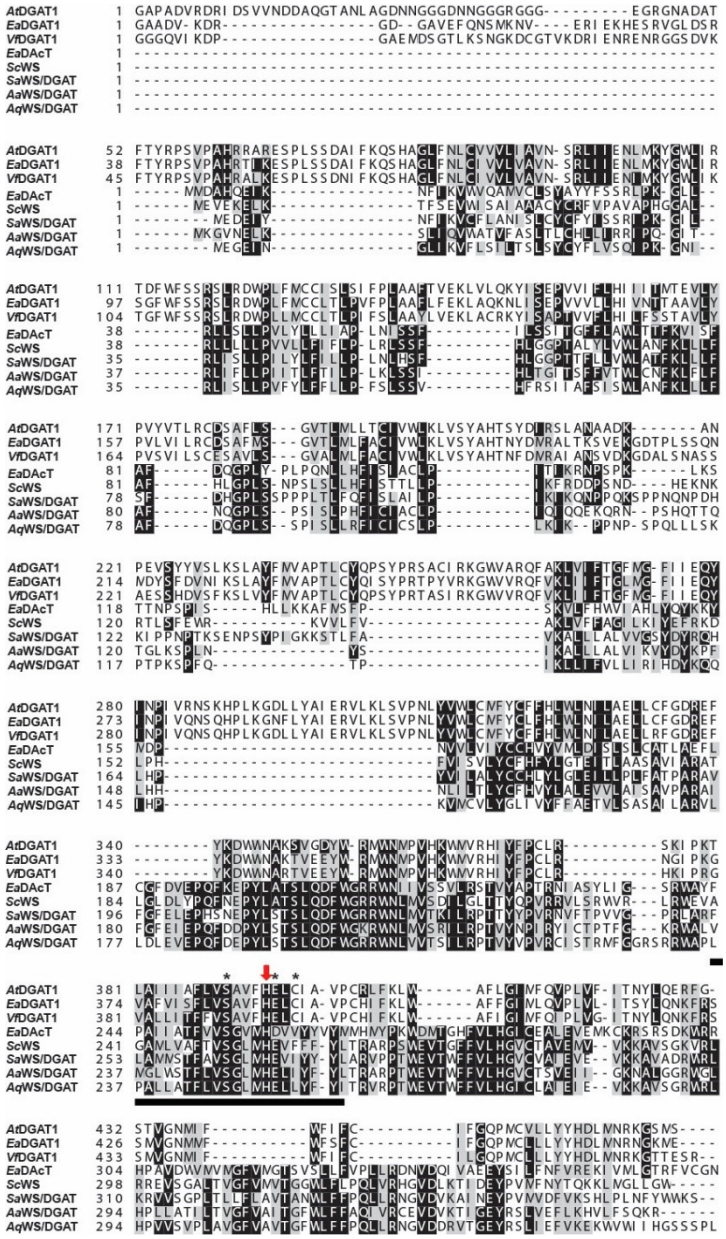

Figure 4.12 Multiple sequence alignment of the novel WS/DGAT bifunctional proteins with MBOATs that acylate fatty alcohol and diacylglycerol

Protein sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE). Residues similar to consensus are shaded in grey and residues identical to consensus are indicated by white letters on black background. Red arrow indicates the putative histidine active site of MBOATs. The MBOAT signature region is underlined. The asterisks denote amino acid residues aligned with residues essential for acetyltransferase activity of *Ea*DAcT. Proteins aligned were DAcT, WS, DGAT1 sequences from *Adonis aestivalis* (*Aa*)*; Akebia quinata* (*Aq*); *Arabidopsis thaliana (At), Euonymus alatus* (*Ea*)*; Sorbus aucuparia* (*Sa*); *Simmondsia chinensis* (*Sc*); *Vernicia fordii* (*Vf*).

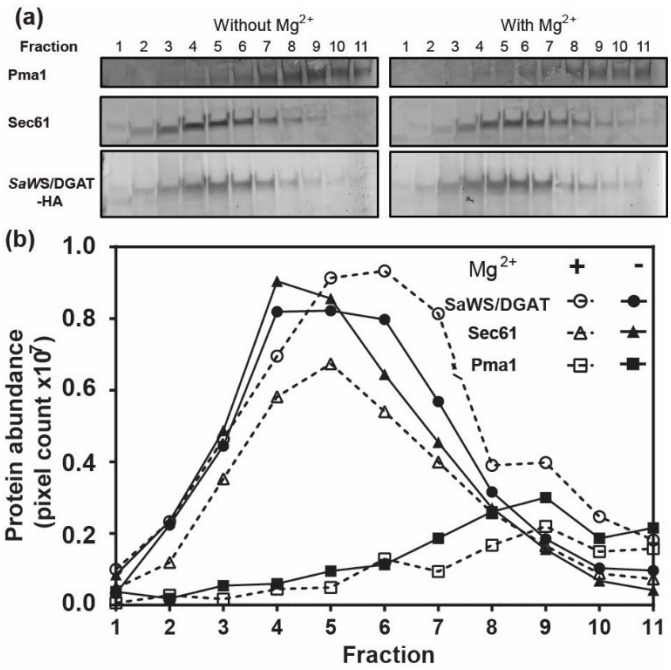

Figure 4.13 *Sa*WS/DGAT associates with the endoplasmic reticulum

Lysates from yeast cells expressing *Sa*WS/DGAT-HA were fractionated by centrifugation in a sucrose step gradient in the absence or presence of Mg^{2+}. (a) Immunoblots to detect the plasma membrane marker Pma1, the ER marker Sec61 and *Sa*WS/DGAT-HA in different fractions. (b) Quantification of immunoblot signal for *Sa*WS/DGAT-HA, Sec61 and Pma1 in different fractions.

Chapter 5 - Conclusions and future directions

*Ea*DAcT is an unusual MBOAT in that it catalyzes the transfer of acetate, the shortest acyl group, to the *sn-3* position of diacylglycerol. This study describes the mapping of the membrane topology of *Ea*DAcT which provides the first experimentally determined topology model for a plant MBOAT. Two different yet complementary approaches were employed to obtain this topology model. Protease protection assays were used to localize the orientation of the N- and C- termini of *Ea*DAcT relative to the membrane bilayer (Chapter 2, Figure 2.3). The orientation of the internal regions of the protein was determined by thiol- specific modification of cysteine residues (Chapter 2, Figure 2.7, Figure 2.8). In the proposed model, *Ea*DAcT contains four transmembrane domains with both the N- and C- termini orientated toward the lumen of the endoplasmic reticulum (Chapter 2, Figure 2.9). Similar to HHAT, GOAT and ACAT (Guo et al., 2005; Taylor et al., 2013; Matevossian and Resh, 2015), the MBOAT signature region of *Ea*DAcT is embedded in the membrane. This region of *Ea*DAcT may be close to the interface between the membrane and the cytoplasm. Topology models of HHAT and GOAT were proposed to possess at least one re-entrant loop (Taylor et al., 2013; Matevossian and Resh, 2015), a similar loop has not been found for *Ea*DAcT. Instead, *Ea*DAcT possesses a uniquely large cytoplasmic loop between the first and the second TMD (Chapter 2, Figure 2.9). While additional scanning of this region to investigate possible re-entrant loops is required to further refine the membrane topology of *Ea*DAcT, it will be interesting to test if the region is involved in substrate binding of the protein. This can be achieved by purifying a sufficient amount of the region between the first and the second TMD of *Ea*DAcT followed by substrate binding assays. The membrane topology of *Ea*DAcT proposed in this study provides a basis for further structure-function studies on the enzyme. Ultimately however, a high resolution structure of *Ea*DAcT is needed to elucidate the structural basis for the substrate specificity of the protein.

This study also describes the isolation of functional orthologs of *Ea*DAcT and the identification of amino acid residues important for enzymatic activity of *Ea*DAcT. Six proteins homologous to *Ea*DAcT were isolated from acetyl-TAG producing plant species and demonstrated to possess DAG acetyltransferase activity (Chapter 3, Figure 3.2, Figure 3.3). While *S. aucuparia, S. alnifolia* and *A. quinata* accumulate acetyl-TAG in their seeds, proteins of these species that share high sequence identity and similarity to *Ea*DAcT do not possess DAG

acetyltransferase activity. Instead, they catalyze the transfer of a long chain acyl-CoA to DAG and fatty alcohols (Chapter 4).This implies that the proteins responsible for acetyl-TAG synthesis in *S. aucuparia, S. alnifolia* and *A. quinata* may belong to a different group of enzymes capable of acetylating DAG. The BAHD enzyme family is attractive due its ability to acetylate a variety of alcohols (St-Pierre and De Luca, 2000). The cloning and characterization of BAHDs from *S. aucuparia, S. alnifolia* and *A. quinata* could allow the identification of putative alcohol acetyltransferases. Further biochemical characterization of these putative proteins is required to determine their ability to synthesize acetyl-TAG.

The sequence diversity of orthologs of *Ea*DAcT allowed us to identify amino acid residues conserved throughout MBOATs or within DAcT enzymes. The roles of these residues in *Ea*DAcT were studied by site-directed mutagenesis in conjunction with in vitro and in vivo acetyl-TAG synthesis assays. Our mutagenesis data demonstrated that S253, H257and D258 are essential for the enzyme activity of *Ea*DAcT (Chapter 3, Figure 3.8) suggesting that these residues may form a catalytic triad similar to that proposed for ACAT1 (Das et al., 2008). The presence of a negatively charged residue, aspartate or glutamate, at the position adjacent to the putative histidine active site is unique to MBOAT enzymes that acylate lipid substrates (Chapter 3, Figure 3.4). In HHAT, PORC and GOAT that acylate proteins and peptides, a glycine residue occupies that position instead (Caricasole et al., 2002; Buglino and Resh, 2008; Yang et al., 2008). In the case of *Ea*DAcT, the negative charge group adjacent to H257 was shown to be essential for enzyme activity of the protein (Chapter 3, Figure 3.8). Together, these findings suggest that the residue adjacent to the putative histine active site may be involved in substrate selection of MBOATs. However, further mutagenesis at this position in other MBOATs is required to validate this assumption. The prosposed catalytic triad of *Ea*DAcT resides in the third TMD (Chapte 2, Figure 2.9) and is thus accessible to cytosolic acetyl-CoA and membrane embedded DAG substrates. However, the substrate binding domains of *Ea*DAcT remains to be identified.

The additional of V263 present in all DAcT is essential for enzyme activity and important for stability of *Ea*DAcT, as deletion of the residue eliminated enzyme activity and reduced protein stability significantly (Chapter 3, Figure 3.8, Figure 3.9). Futhermore, mutation of individual residues unique to DAcT enzymes, including Q150, S222, D258 and V263, failed to

alter the substrate specificity of *Ea*DAcT (Chapter 3, Figure 3.7, Figure 3.10) indicating the involvement of multiple residues or amino acid regions in substrate specificity of the enzyme.

In future work, we will employ molecular evolution approaches that involve randomly swapping sequences between *Ea*DAcT and *Sc*WS to explore the sequence space important for substrate selection of the proteins (Stemmer, 1994; Coco et al., 2001). We will generate a library of *Ea*DAcT-*Sc*WS chimeras that have short DNA sequences of the genes randomly stitched together and take advantage of a powerful yeast-based method to select for changes in substrate specificity of the *Ea*DAcT and *Sc*WS. The selection method is based on our observation that expression of *Ea*DGAT1 or *Ea*DAcT but not *Sc*WS in the H1246 yeast strain devoid of TAG biosynthesis can complement the high oleic acid susceptibility phenotype of the yeast strain (Figure 5.1).

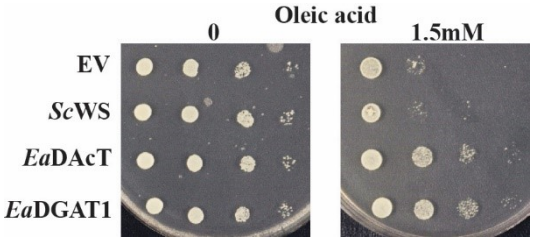

Figure 5.1 Expression of *Ea*DGAT1 or *Ea*DAcT complements the high oleic susceptibility phenotype of H1246 yeast train.

A dilution series of H1246 yeast expressing empty vector (EV) or different MBOATs was plated on defined media containing galactose (for protein induction) and with or without oleic acid.

We speculate that oleic acid is incorporated in DAG which is then acetylated by *Ea*DAcT or used by *Ea*DGAT1 to form neutral storage lipids. *Sc*WS lacks a suitable acyl acceptor pool, and thus cannot complement the H1246 phenotype. We will transform *Ea*DAcT-*Sc*WS chimeras to the H1246 yeast strain which will be subjected to selection on high oleic acid medium. Lipid extracts from survival yeast clones will be analyzed by ESI-MS and/or GC-MS. Chimeric proteins that exhibit TAG synthesis activity will be selected and their corresponding genes will be sequenced to identify amino acid residues/regions responsible for altering substrate specificity.

Additionally, this study presents the discovery and biochemical characterization of 3 novel MBOAT WS/DGAT bifunctional proteins from *S. aucuparia, A. aestivalis* and *A. quinata*. The enzymes co-localized with an ER membrane marker and are capable of synthesizing lcTAG and WE in vitro and in vivo. They are closely related to *Sc*WS, *Ea*DAcT, *At*ASAT1 and distantly

related to DGAT1 enzymes (Chapter 4, Figure 4.11). Additionally, the novel bifunctional enzymes are distinct from the catalytically similar enzymes from *A. thaliana* and bacteria (Chapter 4, Figure 4.11). The *Sa*WS/DGAT-HA and *Aa*WS/DGAT-HA exhibited stronger wax ester synthesis activity than *Sc*WS-HA (Chapter 4, Figure 4.1, Figure 4.8). The smaller size of the bifunctional proteins compared to DGAT1 proteins and their dual activity imply they can be used as models for studying the structual features that control the selection of fatty alcohol and diacylglycerol substrates of an acyl-CoA dependent acyltransferase. As the novel bifunctional WS/DGAT enzymes are closely related to *Sc*WS, they are also appropriate for the aforementioned yeast selection study. Expressing *Sa*WS/DGAT-HA or *Aa*WS/DGAT-HA in Camelina did not result in increased oil content or accumulation of WE in T_2 transgenic seeds (Chapter 4, Figure 4.9, Figure 4.10). We suspect that this may be dued to the insufficient expression of the proteins in the plants. We therefore will monitor the expression level of the proteins in the T_2 transgenic plants in conjunction with generating T_3 homozygote transgenic lines for further seed lipid analysis.

Reference List

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.,* 25, 3389-3402.

Bagby, M.O., and Smith, C.R.,Jr. (1967) Asymmetric triglycerides from *Impatiens edgeworthii* seed oil. *Biochim. Biophys. Acta,* 137, 475-477.

Bansal, S., and Durrett, T.P. (2016a) Defining the extreme substrate specificity of *Euonymus alatus* diacylglycerol acetyltransferase, an unusual membrane bound O-acyltransferase. *Biosci. Rep.,* 36, e00406.

Bansal, S., and Durrett, T.P. (2016b) Rapid Quantification of Low-Viscosity Acetyl-Triacylglycerols Using Electrospray Ionization Mass Spectrometry. *Lipids,* 51, 1093-1102.

Benghezal, M., Wasteneys, G.O., and Jones, D.A. (2000) The C-terminal dilysine motif confers endoplasmic reticulum localization to type I membrane proteins in plants. *Plant Cell,* 12, 1179-1201.

Berg, M., Tymoczko, L., and Stryer, L. (2002) Triacylglycerols Are Highly Concentrated Energy Stores. *In Biochemistry.* New York: W. H. Freeman.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G., and Elofsson, A. (2008) Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci. U. S. A.,* 105, 7177-7181.

Bibi, E., and Beja, O. (1994) Membrane topology of multidrug resistance protein expressed in *Escherichia coli*. N-terminal domain. *J. Biol. Chem.,* 269, 19910-19915.

Bogdanov, M., Zhang, W., Xie, J., and Dowhan, W. (2005) Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis. *Methods,* 36, 148-171.

Broome-Smith, J.K., Tadayyon, M., and Zhang, Y. (1990) Beta-lactamase as a probe of membrane protein assembly and protein export. *Mol. Microbiol.,* 4, 1637-1644.

Brumlik, M.J., and Buckley, J.T. (1996) Identification of the catalytic triad of the lipase/acyltransferase from Aeromonas hydrophila. *J. Bacteriol.,* 178, 2060-2064.

Buglino, J.A., and Resh, M.D. (2010) Identification of conserved regions and residues within Hedgehog acyltransferase critical for palmitoylation of Sonic Hedgehog. *PLoS One,* 5, e11195.

Buglino, J.A., and Resh, M.D. (2008) Hhat is a palmitoylacyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J. Biol. Chem.,* 283, 22076-22088.

Caldo, K.M., Greer, M.S., Chen, G., Lemieux, M.J., and Weselake, R.J. (2015) Purification and properties of recombinant *Brasisica napus* diacylglycerol acyltransferase 1. *FEBS Lett.,* 589, 773-778.

Cao, G., Goldstein, J.L., and Brown, M.S. (1996) Complementation of mutation in acyl-CoA:cholesterol acyltransferase (ACAT) fails to restore sterol regulation in ACAT-defective sterol-resistant hamster cells. *J. Biol. Chem.,* 271, 14642-14648.

Cao, H., Chapital, D.C., Shockey, J.M., and Klasson, K.T. (2011) Expression of tung tree diacylglycerol acyltransferase 1 in *E. coli*. *BMC Biotechnol.,* 11. doi: 10.1186/1472-6750-11-73.

Cao, J., Cheng, L., and Shi, Y. (2007) Catalytic properties of MGAT3, a putative triacylgycerol synthase. *J. Lipid Res.,* 48, 583-591.

Caricasole, A., Ferraro, T., Rimland, J.M., and Terstappen, G.C. (2002) Molecular cloning and initial characterization of the MG61/PORC gene, the human homologue of the Drosophila segment polarity gene Porcupine. *Gene,* 288, 147-157.

Cases, S., Smith, S.J., Zheng, Y.W., Myers, H.M., Lear, S.R., Sande, E., Novak, S., Collins, C., Welch, C.B., Lusis, A.J., Erickson, S.K., and Farese, R.V.,Jr. (1998) Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proc. Natl. Acad. Sci. U. S. A.,* 95, 13018-13023.

Chang, C.C., Huh, H.Y., Cadigan, K.M., and Chang, T.Y. (1993a) Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary cells. *J. Biol. Chem.,* 268, 20747-20755.

Chang, S., Puryear, J., and Cairney, J. (1993b) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter,* 11, 113-116.

Chang, T.Y., Li, B.L., Chang, C.C., and Urano, Y. (2009) Acyl-coenzyme A:cholesterol acyltransferases. *Am. J. Physiol. Endocrinol. Metab.,* 297, E1-9.

Chen, Q., Steinhauer, L., Hammerlindl, J., Keller, W., and Zou, J. (2007) Biosynthesis of phytosterol esters: identification of a sterol o-acyltransferase in Arabidopsis. *Plant Physiol.,* 145, 974-984.

Cheng, J.B., and Russell, D.W. (2004) Mammalian wax biosynthesis. II. Expression cloning of wax synthase cDNAs encoding a member of the acyltransferase enzyme family. *J. Biol. Chem.,* 279, 37798-37807.

Chevreux, B., Wetter, T., and Suhai, S. (1999) Genome Sequence Assembly Using Trace Signals and Additional Sequence Information.99, 46.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., and Monticello, D.J. (2001) DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.,* 19, 354-359.

Daniel, J., Deb, C., Dubey, V.S., Sirakova, T.D., Abomoelak, B., Morbidoni, H.R., and Kolattukudy, P.E. (2004) Induction of a novel class of diacylglycerol acyltransferases and triacylglycerol accumulation in *Mycobacterium tuberculosis* as it goes into a dormancy-like state in culture. *J. Bacteriol.,* 186, 5017-5030.

Das, A., Davis, M.A., and Rudel, L.L. (2008) Identification of putative active site residues of ACAT enzymes. *J. Lipid Res.,* 49, 1770-1781.

Deber, C.M., Wang, C., Liu, L.P., Prior, A.S., Agrawal, S., Muskat, B.L., and Cuticchia, A.J. (2001) TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci.,* 10, 212-219.

Ding, B.J., Lager, I., Bansal, S., Durrett, T.P., Stymne, S., and Lofstedt, C. (2016) The Yeast ATF1 Acetyltransferase Efficiently Acetylates Insect Pheromone Alcohols: Implications for the Biological Production of Moth Pheromones. *Lipids,* 51, 469-475.

Durrett, T.P., Benning, C., and Ohlrogge, J. (2008) Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J.,* 54, 593-607.

Durrett, T.P., McClosky, D.D., Tumaney, A.W., Elzinga, D.A., Ohlrogge, J., and Pollard, M. (2010) A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in Euonymus and transgenic seeds. *Proc. Natl. Acad. Sci. U. S. A.,* 107, 9464-9469.

Dyer, J.M., Stymne, S., Green, A.G., and Carlsson, A.S. (2008) High-value oils from plants. *Plant J*., *54,* 640-655.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5, 113.

Gaupp, R., and Adams, W. (2004) Acid esters of mono- and diglycerides. In *Emulsifiers in Food Technology.* (Whitehurst RJ ed., ed). Oxford, UK: Blackwell Publishing, pp. 59.

Geller, D., Taglicht, D., Edgar, R., Tam, A., Pines, O., Michaelis, S., and Bibi, E. (1996) Comparative topology studies in Saccharomyces cerevisiae and in *Escherichia coli*. The N-terminal half of the yeast ABC protein Ste6. *J. Biol. Chem.*, 271, 13746-13753.

Guo, Z.Y., Lin, S., Heinen, J.A., Chang, C.C., and Chang, T.Y. (2005) The active site His-460 of human acyl-coenzyme A:cholesterol acyltransferase 1 resides in a hitherto undisclosed transmembrane domain. *J. Biol. Chem.,* 280, 37814-37826.

Gutierrez, J.A., Solenberg, P.J., Perkins, D.R., Willency, J.A., Knierman, M.D., Jin, Z., Witcher, D.R., Luo, S., Onyia, J.E., and Hale, J.E. (2008) Ghrelin octanoylation mediated by an orphan lipid transferase. *Proc. Natl. Acad. Sci. U. S. A.,* 105, 6320-6325.

Hennerdal, A., and Elofsson, A. (2011) Rapid membrane protein topology prediction. *Bioinformatics,* 27, 1322-1323.

Hofmann, K. (2000) A superfamily of membrane-bound O-acyltransferases with implications for wnt signaling. *Trends Biochem. Sci.,* 25, 111-112.

Hofvander, P., Doan, T.T., and Hamberg, M. (2011) A prokaryotic acyl-CoA reductase performing reduction of fatty acyl-CoA to fatty alcohol. *FEBS Lett.,* 585, 3538-3543.

Huang, A. (1992) Oil bodies and oleosins in seeds. *Annu. Rev. Plant Physiol. Plant Mol. Biol*., 43, 177-200.

Iven, T., Herrfurth, C., Hornung, E., Heilmann, M., Hofvander, P., Stymne, S., Zhu, L.H., and Feussner, I. (2013) Wax ester profiling of seed oil by nano-electrospray ionization tandem mass spectrometry. *Plant. Methods,* 9, 24-4811-9-24.

Jander, G., Cronan, J.E.,Jr, and Beckwith, J. (1996) Biotinylation in vivo as a sensitive indicator of protein secretion and membrane protein insertion. *J. Bacteriol.,* 178, 3049-3058.

Jones, D.T. (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics,* 23, 538-544.

Kalscheuer, R., and Steinbuchel, A. (2003) A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in Acinetobacter calcoaceticus ADP1. *J. Biol. Chem.,* 278, 8075-8082.

Kast, C., Canfield, V., Levenson, R., and Gros, P. (1996) Transmembrane organization of mouse P-glycoprotein determined by epitope insertion and immunofluorescence. *J. Biol. Chem.,* 271, 9240-9248.

Kim, H., Park, J.H., and Kim, D.J. (2016) Functional analysis of *diacylglycerol acyltransferase1* genes from *Camelina sativa* and effects of *CsDGAT1B* overexpression on seed mass and storage oil content in *C. sativa*.10, 141-153.

Kleiman, R., Miller, R.W., Earle, F.R., and Wolff, I.A. (1967) (S)-1,2-diacyl-3-acetins: Optically active triglycerides from *Euonymus verrucosus* seed oil. *Lipids,* 2, 473-478.

Konitsiotis, A.D., Jovanovic, B., Ciepla, P., Spitaler, M., Lanyon-Hogg, T., Tate, E.W., and Magee, A.I. (2015) Topological analysis of Hedgehog acyltransferase, a multipalmitoylated transmembrane protein. *J. Biol. Chem.,* 290, 3293-3307.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.,* 305, 567-580.

Kroon, J.T., Wei, W., Simon, W.J., and Slabas, A.R. (2006) Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. *Phytochemistry,* 67, 2541-2549.

Lardizabal, K.D., Metz, J.G., Sakamoto, T., Hutton, W.C., Pollard, M.R., and Lassner, M.W. (2000) Purification of a jojoba embryo wax synthase, cloning of its cDNA, and production of high levels of wax in seeds of transgenic arabidopsis. *Plant Physiol.,* 122, 645-655.

Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L. (2008) Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis. *Plant Physiol.,* 148, 97-107.

Li, R., Yu, K., Hatanaka, T., and Hildebrand, D.F. (2010) Vernonia DGATs increase accumulation of epoxy fatty acids in oil. *Plant. Biotechnol. J.,* 8, 184-195.

Li, Y., Beisson, F., Pollard, M., and Ohlrogge, J. (2006) Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. *Phytochemistry,* 67, 904-915.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., Debono, A., Durrett, T.P.*, et al.* (2013) Acyl-lipid metabolism. In *Arabidopsis Book,* 11, e0161.

Lin, S., Cheng, D., Liu, M.S., Chen, J., and Chang, T.Y. (1999) Human acyl-CoA:cholesterol acyltransferase-1 in the endoplasmic reticulum contains seven transmembrane domains. *J. Biol. Chem.,* 274, 23276-23285.

Liu, J., Rice, A., McGlew, K., Shaw, V., Park, H., Clemente, T., Pollard, M., Ohlrogge, J., and Durrett, T.P. (2015) Metabolic engineering of oilseed crops to produce high levels of novel acetyl glyceride oils with reduced viscosity, freezing point and calorific value. *Plant. Biotechnol. J.,* 13, 858-865.

Liu, Q., Siloto, R.M., Lehner, R., Stone, S.J., and Weselake, R.J. (2012) Acyl-CoA:diacylglycerol acyltransferase: molecular biology, biochemistry and biotechnology. *Prog. Lipid Res.,* 51, 350-377.

Liu, Q., Siloto, R.M., Snyder, C.L., and Weselake, R.J. (2011) Functional and topological analysis of yeast acyl-CoA:diacylglycerol acyltransferase 2, an endoplasmic reticulum enzyme essential for triacylglycerol biosynthesis. *J. Biol. Chem.,* 286, 13115-13126.

Logan, C.Y., and Nusse, R. (2004) The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.,* 20, 781-810.

Long, J.C., Wang, S., and Vik, S.B. (1998) Membrane topology of subunit a of the F1F0 ATP synthase as determined by labeling of unique cysteine residues. *J. Biol. Chem.,* 273, 16235-16240.
Lu, C., and Kang, J. (2008) Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by Agrobacterium-mediated transformation. *Plant Cell Rep.,* 27, 273-278.

Lu, X., Lin, S., Chang, C.C., and Chang, T.Y. (2002) Mutant acyl-coenzyme A:cholesterol acyltransferase 1 devoid of cysteine residues remains catalytically active. *J. Biol. Chem.,* 277, 711-718.

Marshall, K.E., Thomas, R.H., Roxin, A., Chen, E.K., Brown, J.C., Gillies, E.R., and Sinclair, B.J. (2014) Seasonal accumulation of acetylated triacylglycerols by a freeze-tolerant insect. *J. Exp. Biol.,* 217, 1580-1587.

Masumoto, N., Lanyon-Hogg, T., Rodgers, U.R., Konitsiotis, A.D., Magee, A.I., and Tate, E.W. (2015) Membrane bound O-acyltransferases and their inhibitors. *Biochem. Soc. Trans.,* 43, 246-252.

Matevossian, A., and Resh, M.D. (2015) Membrane topology of hedgehog acyltransferase. *J. Biol. Chem.,* 290, 2235-2243.

Mayer, K.M., and Shanklin, J. (2007) Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC Plant. Biol.,* 7, 1. doi: 10.1186/1471-2229-7-1.

McCartney, A.W., Dyer, J.M., Dhanoa, P.K., Kim, P.K., Andrews, D.W., McNew, J.A., and Mullen, R.T. (2004) Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *Plant J.,* 37, 156-173.

McFie, P.J., Stone, S.L., Banman, S.L., and Stone, S.J. (2010) Topological orientation of acyl-CoA:diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the N terminus in dimer/tetramer formation. *J. Biol. Chem.,* 285, 37377-37387.

Milcamps, A., Tumaney, A.W., Paddock, T., Pan, D.A., Ohlrogge, J., and Pollard, M. (2005) Isolation of a gene encoding a 1,2-diacylglycerol-sn-acetyl-CoA acetyltransferase from developing seeds of *Euonymus alatus*. *J. Biol. Chem.,* 280, 5370-5377.

Moss, C.W., Shinoda, T., and Samuels, J.W. (1982) Determination of cellular fatty acid compositions of various yeasts by gas-liquid chromatography. *J. Clin. Microbiol.,* 16, 1073-1079.

Nugent, T., and Jones, D.T. (2012) Detecting pore-lining regions in transmembrane protein sequences. *BMC Bioinformatics,* 13, 169-2105-13-169.

Oelkers, P., Behari, A., Cromley, D., Billheimer, J.T., and Sturley, S.L. (1998) Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes. *J. Biol. Chem.,* 273, 26765-26771.
Ozawa, A., Speaker, R.B.,3rd, and Lindberg, I. (2009) Enzymatic characterization of a human acyltransferase activity. *PLoS One,* 4, e5426.

Panikashvili, D., Shi, J.X., Schreiber, L., and Aharoni, A. (2009) The Arabidopsis DCR encoding a soluble BAHD acyltransferase is required for cutin polyester formation and seed hydration properties. *Plant Physiol.,* 151, 1773-1789.

Paul, S., Gable, K., and Dunn, T.M. (2007) A six-membrane-spanning topology for yeast and Arabidopsis Tsc13p, the enoyl reductases of the microsomal fatty acid elongating system. *J. Biol. Chem.,* 282, 19237-19246.

Pedersen, A.G., and Nielsen, H. (1997) Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. *Proc. Int. Conf. Intell. Syst. Mol. Biol.,* 5, 226-233.

Peelman, F., Vinaimont, N., Verhee, A., Vanloo, B., Verschelde, J.L., Labeur, C., Seguret-Mace, S., Duverger, N., Hutchinson, G., Vandekerckhove, J., Tavernier, J., and Rosseneu, M. (1998) A proposed architecture for lecithin cholesterol acyl transferase (LCAT): identification of the catalytic triad and molecular modeling. *Protein Sci.,* 7, 587-599.

Peters, C., Tsirigos, K.D., Shu, N., and Elofsson, A. (2016) Improved topology prediction using the terminal hydrophobic helices rule. *Bioinformatics,* 32, 1158-1162.

Rios-Esteves, J., Haugen, B., and Resh, M.D. (2014) Identification of key residues and regions important for porcupine-mediated Wnt acylation. *J. Biol. Chem.,* 289, 17009-17019.

Roesler, K., Shen, B., Bermudez, E., Li, C., Hunt, J., Damude, H.G., Ripp, K.G., Everard, J.D., Booth, J.R., Castaneda, L., Feng, L., and Meyer, K. (2016) An Improved Variant of Soybean Type 1 Diacylglycerol Acyltransferase Increases the Oil Content and Decreases the Soluble Carbohydrate Content of Soybeans. *Plant Physiol.,* 171, 878-893.

Romano, J.D., and Michaelis, S. (2001) Topological and mutational analysis of *Saccharomyces cerevisiae* Ste14p, founding member of the isoprenylcysteine carboxyl methyltransferase family. *Mol. Biol. Cell,* 12, 1957-1971.

Saha, S., Enugutti, B., Rajakumari, S., and Rajasekharan, R. (2006) Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol.,* 141, 1533-1543.

Sahin-Toth, M., Dunten, R.L., and Kaback, H.R. (1995) Design of a membrane protein for site-specific proteolysis: properties of engineered factor Xa protease sites in the lactose permease of Escherichia coli. *Biochemistry,* 34, 1107-1112.

Sandager, L., Gustavsson, M.H., Stahl, U., Dahlqvist, A., Wiberg, E., Banas, A., Lenman, M., Ronne, H., and Stymne, S. (2002) Storage lipid synthesis is non-essential in yeast. *J. Biol. Chem.,* 277, 6478-6482.
Schmieder, R., and Edwards, R. (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics,* 27, 863-864.

Schrag, J.D., Li, Y.G., Wu, S., and Cygler, M. (1991) Ser-His-Glu triad forms the catalytic site of the lipase from Geotrichum candidum. *Nature,* 351, 761-764.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics,* 28, 1086-1092.

Shen, H., and Chou, J.J. (2008) MemBrain: improving the accuracy of predicting transmembrane helices. *PLoS One,* 3, e2399.

Shindou, H., and Shimizu, T. (2009) Acyl-CoA:lysophospholipid acyltransferases. *J. Biol. Chem.,* 284, 1-5.

Shockey, J.M., Gidda, S.K., Chapital, D.C., Kuan, J.C., Dhanoa, P.K., Bland, J.M., Rothstein, S.J., Mullen, R.T., and Dyer, J.M. (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *Plant Cell,* 18, 2294-2313.

Sidorov, R.A., Trusov, N.A., and Zhukov, A.V. (2013) Accumulation of neutral acylglycerols during the formation of morphologo-anatomical structure of euonymus Fruits. *Russian Journal of Plant Physiology,* 60, 800-811.

Silhavy, T.J., Shuman, H.A., Beckwith, J., and Schwartz, M. (1977) Use of gene fusions to study outer membrane protein localization in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.,* 74, 5411-5415.

Smith, S.J., Cases, S., Jensen, D.R., Chen, H.C., Sande, E., Tow, B., Sanan, D.A., Raber, J., Eckel, R.H., and Farese, R.V.,Jr. (2000) Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nat. Genet.,* 25, 87-90.

Stemmer, W.P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.,* 91, 10747-10751.

Stone, S.J., Levin, M.C., Zhou, P., Han, J., Walther, T.C., and Farese, R.V.,Jr. (2009) The endoplasmic reticulum enzyme DGAT2 is found in mitochondria-associated membranes and has a mitochondrial targeting signal that promotes its association with mitochondria. *J. Biol. Chem.,* 284, 5352-5361.

Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., and Steinbuchel, A. (2005) The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from Acinetobacter sp. strain ADP1: characterization of a novel type of acyltransferase. *J. Bacteriol.,* 187, 1369-1376.

St-Pierre, B., and De Luca, V. (2000) Evolution of acyltransferase genes: Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism. *Recent Advances in Phytochemistry,* 34, 285-315.

Takada, R., Satomi, Y., Kurata, T., Ueno, N., Norioka, S., Kondoh, H., Takao, T., and Takada, S. (2006) Monounsaturated fatty acid modification of Wnt protein: its role in Wnt secretion. *Dev. Cell.,* 11, 791-801.

Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A., and Kumar, S. (2012) Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci. U. S. A.,* 109, 19333-19338.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.,* 28, 2731-2739.

Taylor, M.S., Ruch, T.R., Hsiao, P.Y., Hwang, Y., Zhang, P., Dai, L., Huang, C.R., Berndsen, C.E., Kim, M.S., Pandey, A.*, et al.* (2013) Architectural organization of the metabolic regulatory enzyme ghrelin O-acyltransferase. *J. Biol. Chem.,* 288, 32211-32228.

Tsirigos, K.D., Peters, C., Shu, N., Kall, L., and Elofsson, A. (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.,* 43, W401-7.

Ulmschneider, M.B., and Sansom, M.S. (2001) Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta,* 1512, 1-14.

van Geest, M., and Lolkema, J.S. (2000) Membrane topology of the Na(+)/citrate transporter CitS of Klebsiella pneumoniae by insertion mutagenesis. *Biochim. Biophys. Acta,* 1466, 328-338.

Viklund, H., and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics,* 24, 1662-1668.

von Heijne, G. (2006) Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.,* 7, 909-918.

von Heijne, G. (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.,* 225, 487-494.

Wang, Y., Toei, M., and Forgac, M. (2008) Analysis of the membrane topology of transmembrane segments in the C-terminal hydrophobic domain of the yeast vacuolar ATPase subunit a (Vph1p) by chemical modification. *J. Biol. Chem.,* 283, 20696-20702.

Weselake, R.J., Madhavji, M., Szarka, S.J., Patterson, N.A., Wiehler, W.B., Nykiforuk, C.L., Burton, T.L., Boora, P.S., Mosimann, S.C., Foroud, N.A.*, et al.* (2006) Acyl-CoA-binding and

self-associating properties of a recombinant 13.3 kDa N-terminal fragment of diacylglycerol acyltransferase-1 from oilseed rape. *BMC Biochem.,* 7, 24. doi: 10.1186/1471-2091-7-24.

Wessels, H.P., and Spiess, M. (1988) Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. *Cell,* 55, 61-70.

White, S.H., and Wimley, W.C. (1999) Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.,* 28, 319-365.

Winkler, F.K., D'Arcy, A., and Hunziker, W. (1990) Structure of human pancreatic lipase. *Nature,* 343, 771-774.

Wurie, H.R., Buckett, L., and Zammit, V.A. (2011) Evidence that diacylglycerol acyltransferase 1 (DGAT1) has dual membrane topology in the endoplasmic reticulum of HepG2 cells. *J. Biol. Chem.,* 286, 36238-36247.

Yang, H., Bard, M., Bruner, D.A., Gleeson, A., Deckelbaum, R.J., Aljinovic, G., Pohl, T.M., Rothstein, R., and Sturley, S.L. (1996) Sterol esterification in yeast: a two-gene process. *Science,* 272, 1353-1356.

Yang, H.O., Park, J.S., Cho, S.H., Yoon, J.Y., Kim, M.G., Jhon, G.J., Han, S.Y., and Kim, S.H. (2004) Stimulatory effects of monoacetyldiglycerides on hematopoiesis. *Biol. Pharm. Bull.,* 27, 1121-1125.

Yang, J., Brown, M.S., Liang, G., Grishin, N.V., and Goldstein, J.L. (2008) Identification of the acyltransferase that octanoylates ghrelin, an appetite-stimulating peptide hormone. *Cell,* 132, 387-396.

Yen, C.L., and Farese, R.V.,Jr. (2003) MGAT2, a monoacylglycerol acyltransferase expressed in the small intestine. *J. Biol. Chem.,* 278, 18532-18537.

Yen, C.L., Stone, S.J., Cases, S., Zhou, P., and Farese, R.V.,Jr. (2002) Identification of a gene encoding MGAT1, a monoacylglycerol acyltransferase. *Proc. Natl. Acad. Sci. U. S. A.,* 99, 8512-8517.

Yu, C., Kennedy, N.J., Chang, C.C., and Rothblatt, J.A. (1996) Molecular cloning and characterization of two isoforms of *Saccharomyces cerevisiae* acyl-CoA:sterol acyltransferase. *J. Biol. Chem.,* 271, 24157-24163.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.,* 18, 821-829.

Zhu, L.H., Krens, F., Smith, M.A., Li, X., Qi, W., van Loo, E.N., Iven, T., Feussner, I., Nazarenus, T.J., Huai, D.*, et al.* (2016) Dedicated Industrial Oilseed Crops as Metabolic Engineering Platforms for Sustainable Industrial Feedstock Production. *Sci. Rep.,* 6, 22181.