Changing the substrate specificity of arogenate dehydratases (ADTs) from Arabidopsis thaliana.

Megan ES Smith-Uffen
The University of Western Ontario

Supervisor
Dr. Susanne Kohalmi
The University of Western Ontario

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Megan ES Smith-Uffen 2014

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/2551

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca.
Changing the substrate specificity of arogenate dehydratases (ADTs) from

*Arabidopsis thaliana*.

(Spine title: Changing the substrate specificity of ADTs)

(Thesis format: Monograph)

by

Megan Smith-Uffen

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario Canada

© Megan Smith-Uffen 2014
Abstract

Phenylalanine (Phe), an essential aromatic amino acid, serves as a precursor for protein synthesis and a variety of secondary metabolites in plants. Two pathways are known for Phe biosynthesis. In the first, prephenate dehydratases (PDTs) convert prephenate to phenylpyruvate, which is transaminated to Phe. In the second, prephenate is transaminated to arogenate, which is converted to Phe by arogenate dehydratases (ADTs). ADTs and PDTs are structurally very similar, as are their substrates. Six ADTs (ADT1-ADT6) have been identified in Arabidopsis thaliana. ADT1 and ADT2 can recognize both prephenate and arogenate as substrates whereas ADT3 – ADT6 are solely arogenate-accepting. Twenty ADT domain-swapping chimeras were generated through overlap extension PCR and were tested for PDT function in a pha2 complementation assay. Through targeted mutagenesis it was identified that a Phe341Leu substitution in ADT5 was sufficient to introduce PDT function to this previously solely-arogenate accepting ADT. This research represents the first identification of any amino acid that discriminates an arogenate-only ADT from an ADT that is able to accept prephenate.

Keywords: Arabidopsis thaliana, phenylalanine biosynthesis, arogenate dehydratase, prephenate dehydratase, overlap extension PCR, pha2 complementation assay, in silico analysis, ACT regulatory domain.
Acknowledgments

First, I would like to thank my supervisor Dr. Susanne Kohalmi for all of the support that she has provided me throughout my research. Her guidance, both technical and personal, made this degree not just possible but also enjoyable. I would also like to thank my advisors Dr. Sangeeta Dhaubhadel and Dr. Kathleen Hill for offering advice and new perspectives on my project.

I would like to thank my lab mates, Sara Rad and Travis Howes for being a constant source of positivity and support. As well, I would like to thank the many undergraduate students that contributed to our lab: Tehmina Ahmad, Andra Jejerjen, William Pirjamali, Brenna Lindsay, and Ornella Kljakic.

In addition, I would like to thank Dr. Jim Karagiannis for generously allowing me to use his nanodrop and 30°C shaker/incubator almost every week. I would also like to thank Dr. Rima Menassa for allowing me to complete my Western blots in her lab, under the guidance of her students Sean Miletic and Angelo Kaldis.

Lastly, I would like to thank my friends and family for supporting me throughout this research. In particular, I would like to thank Jason Speir for his patience and encouragement during the most challenging parts of this degree.
Abstract.................................................................................................................................................. II
Acknowledgments...................................................................................................................................... III
List of Tables ........................................................................................................................................ VI
List of Figures .......................................................................................................................................... VII
List of Appendices .................................................................................................................................. IX
List of Abbreviations ............................................................................................................................... X

1 Introduction .............................................................................................................................................. 1
   1.1 The importance of phenylalanine ................................................................................................. 1
   1.2 Phenylalanine biosynthesis ............................................................................................................ 3
   1.3 Sequence comparison of PDTs and ADTs .................................................................................. 6
   1.4 Substrate specificity ....................................................................................................................... 11
      1.4.1 Substrate specificity of ADTs in vitro ................................................................................... 11
      1.4.2 Substrate specificity of ADTs in vivo ................................................................................... 13
   1.5 ADT and PDT domain organization ......................................................................................... 13
   1.6 ADT and PDT 3D structure ......................................................................................................... 14
   1.7 The importance of few amino acids ......................................................................................... 25
   1.8 Thesis hypothesis and objectives ............................................................................................. 25

2 Materials and Methods ....................................................................................................................... 26
   2.1 Media, buffers and gels ............................................................................................................... 26
      2.1.1 Media .................................................................................................................................... 26
      2.1.2 Buffers ................................................................................................................................... 27
      2.1.3 Gels ....................................................................................................................................... 28
   2.2 Strains and plasmids ................................................................................................................... 29
      2.2.1 Strains .................................................................................................................................... 29
      2.2.2 Plasmids ............................................................................................................................... 29
      2.3 Cloning procedure .................................................................................................................... 32
   2.4 PCR amplification ......................................................................................................................... 32
      2.4.1 Primer Design: ...................................................................................................................... 32
      2.4.2 Overlap extension PCR ......................................................................................................... 32
   2.5 Chimera name annotation ........................................................................................................... 46
List of Tables

Table 1. List of bacterial PDT sequences used in PDT sequence alignment. ...... 9
Table 2. Ability of ADTs to use prephenate versus arogenate as a substrate.... 12
Table 3. Primer sequences for domain swapping........................................... 35
Table 4. Primer sequences for targeted mutagenesis. .................................. 40
Table 5. Components for fragment amplification and overlap extension PCR. .. 44
Table 6. Thermocycler conditions for fragment amplification and overlap
extension PCR. .................................................................................................. 45
Table 7. Composition and expected lengths of ADT2/ADT4 chimeras.......... 54
Table 8. Composition and expected lengths of ADT2/ADT5 chimeras.......... 56
Table 9. Composition and expected lengths of ACT domain chimeras........... 58
List of Figures

Figure 1. The last two steps of Phenylalanine biosynthesis. ........................................ 4
Figure 2. Phylogenetic tree of plant ADTs and bacterial PDTs. ................................. 7
Figure 3. Schematic of the domain structure of plant ADTs and bacterial PDTs. ................................ ................................ ................................................................. 15
Figure 4. Ct-PDT monomer and dimer 3D conformation............................................ 18
Figure 5. Schematic of open (R) and closed (T) states of the PDT enzyme. ....... 21
Figure 6. 3D homology modeling of ADT5 and Ct-PDT. ..................................... 23
Figure 7. Gateway® compatible donor vector and destination vector. .................. 30
Figure 8. Schematic of Gateway® cloning strategy................................................. 33
Figure 9. Schematic of overlap extension PCR....................................................... 42
Figure 10. Schematics for ADT2/ADT4 chimeras................................................ 47
Figure 11. Schematics for ADT2/ADT5 chimeras................................................ 49
Figure 12. Schematics for N-terminal ACT domain constructs............................ 52
Figure 13. Schematic of complementation assay to test for PDT function........... 63
Figure 14. Protein sequence alignment of bacterial PDTs.................................... 67
Figure 15. Protein sequence alignment of Arabidopsis ADTs............................... 73
Figure 16. ADT2/ADT4 chimeras.......................................................................... 78
Figure 17. Complementation of the pha2 phenotype for ADT2/ADT4 chimeras. 81
Figure 18. Construct design for the N-terminal ACT domain............................... 84
Figure 19. ADT2/ADT5 chimeras.......................................................................... 87
Figure 20. Complementation of the pha2 phenotype for ADT2/ADT5 chimeras. 89
Figure 21. N-terminal ACT domain chimeras..................................................... 93
Figure 22. Complementation of the pha2 phenotype for N-terminal ACT domain chimeras.................................................................................................................. 95
Figure 23. Expression of ADT constructs in pha2 yeast...................................... 98
Figure 24. Leu317 and the 3D homology model for ADT2.................................... 102
Figure 25. Phe341 in the 3D homology modeled ADT5...................................... 104
Figure 26. Analysis of Leu317 from *Arabidopsis* and *Petunia* ADT sequences.
List of Appendices

Appendix A ........................................................................................................... 122
List of Abbreviations
NOTE: SI units are not listed

ACT aspartokinase, chorismate mutase, Tyr A domain
ADT arogenate dehydratase
APS ammonium persulfate
Asn Asparagine
BCA bicinchoninic acid
bp base pair
BME β-mercaptoethanol
CAT catalytic domain
CDT cyclohexadienyl dehydratase
CFP cyano fluorescent protein
CM chorismate mutase
DNA deoxyribonucleic acid
ECL enhanced chemiluminescence
EDTA ethylenediaminetetraacetic acid
His histidine
Kb kilobase
LB lysogeny broth
Leu leucine
Met methionine
PAGE polyacrylamide gel electrophoresis
PBS phosphate-buffered saline
PCR polymerase chain reaction
PDB Protein Data Bank
PDT prephenate dehydratase
Phe phenylalanine
PMSF phenylmethanesulfonylfluoride
RE restriction endonuclease or restriction enzyme
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPM</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SD</td>
<td>synthetic dextrose</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>SOC</td>
<td>super optimal catabolite-repression</td>
</tr>
<tr>
<td>TEMED</td>
<td>tetramethylethlenediamine</td>
</tr>
<tr>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>TP</td>
<td>transit peptide</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>TSP</td>
<td>total soluble protein</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
</tbody>
</table>
1 Introduction

The research presented in this thesis is concerned with the final step of the phenylalanine (Phe) biosynthesis pathway in *Arabidopsis thaliana*. This step is catalyzed by arogenate dehydratases (ADTs). A family of six ADTs (ADT1-ADT6) have been identified in *Arabidopsis*.

1.1 The importance of phenylalanine

Phe is an aromatic amino acid that is synthesized through the shikimate pathway in plants, bacteria and fungi (Herrmann and Weaver 1999). Phe is required for protein synthesis in all organisms, and therefore is an important dietary component for organisms that cannot synthesize the amino acid themselves, such as humans. In the United States alone, approximately 13,000 tons of Phe is artificially synthesized annually for use as a dietary supplement. This is valued at 198 million USD (Demain and Fang 2000).

In plants, Phe is also a precursor to a diverse family of organic compounds. These compounds, such as flavonoids, alkaloids, lignin and suberin have a wide variety of important functions within the plant.

Flavonoids are aromatic molecules that are created when Phe is converted to 4-coumaroyl-CoA (Falcone Ferreyra et al. 2012). These molecules represent the major red, blue and purple pigments in plants. These pigments recruit pollinators and seed dispersers (Winkel-Shirley 2001), which allow plants to improve their reproductive success. In addition, flavonoid pigments make flowering plants aesthetically pleasing and aesthetics such as these drive the success of the fast-growing flower industry. In 2012, flower and flower-related item sales were responsible for 27.8 billion dollars of revenue in the United States alone (U.S. Bureau of Economic Analysis, July 31, 2013).

Alkaloids are a class of nitrogen-containing organic compounds that are derived from Phe and tyrosine (Tyr; Facchini 2001). These molecules are key mediators of the plant’s resistance towards pests. For example, loline alkaloids, which are insecticidal compounds produced in grasses that are
infected with endophytic fungi, increase resistance of the infected grasses to aphid predation (Wilkinson et al. 2000).

Lignin and suberin are Phe-derived biopolymers. Lignin is a broad term for a large group of complex cross-linking polymers of aromatic alcohols. This cross-linking framework is found in plant secondary cell walls and confers rigidity to the cell and stability to the plant as a whole (Vanholme et al. 2010). Lignin is also quite hydrophobic, which makes it very important for water conduction in vascular plants (Vogt 2010). Lignin crosslinking in the secondary plant cell wall acts as an obstacle for water absorption into the cell, and therefore lignin is a crucial component of cells involved in water transport, such as xylem tracheids (Peter and Neale 2004). In fact, lignin is one of the most abundant biopolymers on earth. Between 20 and 30% of the total carbon that is fixed through photosynthesis is channeled through the Phe biosynthesis pathway to the production of lignin, although this number differs greatly depending on whether the plant is woody or herbaceous (Bonawitz and Chapple 2010). Lignin has been a major focus of the biofuel industry due to its role in the inhibition of microbial fermentation of the plant cell wall. This inhibition decreases the efficiency of the energy conversion process in biofuel production, resulting in significant revenue losses (Frei 2013).

Suberin is a cell wall biopolymer that is made up of both polyphenolic and polyaliphatic domains (Bernards 2002). Similar to lignin, suberin functions as a physical barrier to water movement within the plant. However, suberin also plays a significant role in the wound healing process of plants. Wounded plants focus their metabolic energy on the formation of this biopolymer, which can seal off the wound and therefore prevent infection (Yang and Bernards 2007). For example, Soybean plants with increased suberin levels display partial resistance to Phytophthora sojae, a soil-born pathogen which causes root rot (Thomas et al. 2007). Root rot is responsible for millions of dollars of potential Soybean revenue losses every year (Kaufmann and Gerdemann 1958).

Phe is produced in plants and microorganisms but not humans. Therefore, enzymes in the Phe biosynthesis pathway can serve as drug targets against microbial pathogens (Husain et al. 2001), or herbicides against
invading plants (Tohge et al. 2013). In plants, the shikimate pathway has been a target for herbicidal inhibition in the past (Tohge et al. 2013).

1.2 Phenylalanine biosynthesis

Phe biosynthesis begins with the shikimate pathway in plants, fungi, and bacteria. The shikimate pathway is comprised of seven enzymatic reactions which lead to the production of chorismate (Herrmann 1995). Chorismate is the last common precursor of the three aromatic amino acids tryptophan (Trp), tyrosine (Tyr) and Phe. In Phe biosynthesis, chorismate is converted to prephenate by chorismate mutase (CM) (Herrmann and Weaver 1999; Tzin and Galili 2010). Two steps are then necessary to convert prephenate to Phe: a decarboxylation/dehydration step and an aminotransferase step.

The order of these two steps, as well as the enzymes that catalyze them, differ and therefore Phe biosynthesis can be divided into two pathways. These pathways are called the prephenate pathway and the arogenate pathway (Figure 1).

In the prephenate pathway, prephenate is converted to phenylpyruvate via a decarboxylation/dehydration reaction. This reaction is catalyzed by a prephenate dehydratase (PDT; EC4.2.1.51). Phenylpyruvate is then converted to Phe by a phenylpyruvate aminotransferase (Cotton and Gibson 1965; Fazel et al. 1980; Bentley 1990). In the arogenate pathway, prephenate is first transaminated to arogenate and arogenate is then converted to Phe in a decarboxylation/dehydration reaction that is catalyzed by an arogenate dehydratase (ADT; EC4.2.1.911). PDTs and ADTs both catalyze a decarboxylation/dehydration reaction, and recognize substrates (prephenate and arogenate, respectfully) that differ by only a single amino group (Figure 1).

Microorganisms predominantly synthesize Phe via the prephenate pathway (Im and Pittard 1971). Plants, however, predominantly synthesize Phe
Figure 1. The last two steps of Phenylalanine biosynthesis.

Chorismate, the end-product of the shikimate pathway, is converted to prephenate. In the prephenate pathway, prephenate is decarboxylated/dehydrated by a PDT to phenylpyruvate. Phenylpyruvate is then converted to phenylalanine by a PPAT. In the arogenate pathway, prephenate is first converted to arogenate by a PAT, and is then converted via a decarboxylation/dehydration reaction to phenylalanine by an ADT.
via the arogenate pathway (Byng et al. 1981), with only one exception identified in the literature (Yoo et al. 2013) in which an aminotransferase was identified in *Arabidopsis* that could convert phenylpyruvate to phenylalanine. In addition, some microorganisms can use both the arogenate and the prephenate pathways to synthesize Phe (Zhao et al. 1992). These organisms employ enzymes called cyclohexadienyl dehydratases (CDTs), which can accept both arogenate and prephenate as substrates. However, CDTs are quite different from ADTs and PDTs at the sequence level.

Some microorganisms employ a PDT-CM fusion protein, known as a P-protein (Zhang et al. 1998). These proteins can both convert chorismate to prephenate and convert prephenate to phenylpyruvate.

### 1.3 Sequence comparison of PDTs and ADTs

While searching for putative PDT-domain encoding proteins in *Arabidopsis thaliana*, Cho et al. (2007) and Ehlting et al. (2005) identified a family of six PDT-like enzymes. Through biochemical and yeast complementation analyses, these enzymes were shown to produce Phe predominantly via the arogenate pathway (Cho et al. 2007; Bross et al. 2011), and therefore were named ADT1-ADT6 (*ADT1*, At1g11790; *ADT2*, At3g07630; *ADT3*, At2g27820; *ADT4*, At3g44720; *ADT5*, At5g22630; and *ADT6*, At1g08250). ADT1 and ADT2 contain introns, whereas ADT3-ADT6 do not.

Phylogenetic analysis suggests that the bacterial and plant enzymes are evolutionarily distinct (Figure 2). The plant sequences used in this tree include the *Arabidopsis* ADTs listed above, as well as three *Petunia hybrida* ADTs (PhADT1: ACY79502.1; PhADT2: ACY79503.1; PhADT3: ACY79504.1) and five *Oriza sativa* ADTs (OsADT1: CBC52502.1; OsADT2: CBC52495.1; OsADT3: CBC52493.1; OsADT4: CBC87919.1; OsADT5: CAW84916.1). The bacterial sequences are listed in Table 1. When comparing the bacterial PDTs, the sequences branch into two groups: one containing Gram-positive bacteria and one containing Gram-negative bacteria, with the exception of *Deinococcus radiodurans* and *Magnetococcus marinus*, which are Gram-positive but cluster
Figure 2. Phylogenetic tree of plant ADTs and bacterial PDTs.

Rooted phylogenetic tree generated with DNAMAN using a bootstrap of 1000. Branch point numbers represent the boot-strapping values and horizontal scale indicates sequence divergence. All sequences that are included in this tree contain both a catalytic domain and an ACT regulatory domain, the transit peptides of the plant sequences have been removed. Dashed lines separate the clustering of the plant sequences into subgroups and bacterial sequences into Gram-positive and Gram-negative groups (with the exception of Deinococcus radiodurans and Magnetococcus marinus, which are Gram-positive but cluster with the Gram-negative ADTs). The plant ADT sequences are highlighted in green and the enterobacterial PDT sequences are highlighted in yellow. A CDT sequence from Pseudomonas aeruginosa is included as a representative of these enzymes.
Table 1. List of bacterial PDT sequences used in PDT sequence alignment.

<table>
<thead>
<tr>
<th>Bacteria Species</th>
<th>Accession Number</th>
<th>Amino Acids Removed(^a)</th>
<th>Gram Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agrobacterium tumefaciens</em></td>
<td>EGL66829.1</td>
<td>6</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Bacillus anthracis</em></td>
<td>WP_000621720.1</td>
<td>2</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>AEP91783.1</td>
<td>-</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Bradyrhizobium japonicum</em></td>
<td>WP_024340564.1</td>
<td>4</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Brucella suis</em></td>
<td>AIB30268.1</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Chlorobium tepidum</em></td>
<td>NP_662549.1</td>
<td>4</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Deinococcus radiodurans</em></td>
<td>WP_010887790.1</td>
<td>-</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Enterococcus faecalis</em></td>
<td>WP_024797082.1</td>
<td>-</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Erwinia herbicola</em></td>
<td>CAH09395.1</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>WP_001324213.1</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>AHL61944.1</td>
<td>-</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em></td>
<td>WP_011835672.1</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Listeria innocua</em></td>
<td>WP_003767020.1</td>
<td>1</td>
<td>Positive</td>
</tr>
</tbody>
</table>
Table 1. (continued). List of bacterial PDT sequences used in PDT sequence alignment.

<table>
<thead>
<tr>
<th>Bacteria Species</th>
<th>Accession Number</th>
<th>Amino Acids Removed&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Gram Stain</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Listeria monocytogenes</em></td>
<td>AHN32779.1</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Magnetococcus marinus</em></td>
<td>ABK44389.1</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Magnetospirillum magnetotacticum</em></td>
<td>WP_018726497.1</td>
<td>9</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Mesorhizobium loti</em></td>
<td>WP_019863035.1</td>
<td>6</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>WP_024914965.1</td>
<td>1</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Pseudomonas stutzeri</em></td>
<td>AHL75120.1</td>
<td>1</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Rhizobium meliloti</em></td>
<td>NP_384330</td>
<td>6</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Rhodopseudomonas palustris</em></td>
<td>WP_013501421.1</td>
<td>1</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Salmonella typhimurium</em></td>
<td>AGK68175.1</td>
<td>1</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>AIA28436.1</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Streptococcus pneumoniae</em></td>
<td>KDE93426.1</td>
<td>1</td>
<td>Positive</td>
</tr>
<tr>
<td><em>Thermoanaerobacter tengcongensis</em></td>
<td>NP_622663</td>
<td>7</td>
<td>Negative</td>
</tr>
<tr>
<td><em>Vibrio cholerae</em></td>
<td>NP_230354</td>
<td>1</td>
<td>Negative</td>
</tr>
</tbody>
</table>

<sup>a</sup>Number of amino acids removed from the N-terminus.
with the Gram-negative ADTs. In general, Gram-positive bacteria express a monofunctional enzyme while Gram-negative bacteria express a bifunctional P-Protein.

There is a high level of sequence diversity among the bacterial PDTs. For example, the bacterial sequences used in the phylogenetic alignment in Figure 2 have an overall identity of only 45.8%. The PDTs from *Chlorobidium tepidum* and *Staphylococcus aureus*, which will be discussed further (section 1.6; Boavida et al. 2005) have a sequence identity of only 27.3% even though they are functionally very similar (Tan et al. 2008). Due to this diversity, it is currently impossible to identify the exact amino acids that can distinguish between an ADT and a PDT based on the protein sequences alone.

When aligned with ADT sequences from several other plant species, the six *Arabidopsis* ADTs branch into three subgroups (Figure 2). ADT1 and ADT2 branch into Sub Groups I and II, respectively, and ADT3-ADT6 group together in Sub Group III. Of the *Arabidopsis* ADTs, ADT4 and ADT5 show the highest degree of sequence identity (82% identity).

### 1.4 Substrate specificity

#### 1.4.1 Substrate specificity of ADTs in vitro

*In vitro* biochemical analysis of the six *Arabidopsis* ADTs (Cho et al. 2007) indicates that all six enzymes preferentially accept arogenate as a substrate (Table 2). However, when tested for the ability to accept prephenate as a substrate, three of these enzymes (ADT1, ADT2 and ADT6) also displayed a low level of PDT function. The catalytic efficiency values ($k_{cat}/K_m$) of ADT1, ADT2 and ADT6 were 28, 32, and 98 times greater for arogenate recognition than prephenate, respectively. This dual substrate recognition can be found in ADT enzyme families in other plant species as well. For example, enzymes with the capacity to accept both arogenate and prephenate have also been identified in *Petunia hybrida*, in which two out of a family of three ADTs are also able to
Table 2. Ability of ADTs to use prephenate versus arogenate as a substrate.

<table>
<thead>
<tr>
<th>ADT</th>
<th>Substrate</th>
<th>V&lt;sub&gt;max&lt;/sub&gt; [pkat µg&lt;sup&gt;-1&lt;/sup&gt;]</th>
<th>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; [M&lt;sup&gt;-1&lt;/sup&gt; S&lt;sup&gt;-1&lt;/sup&gt;]</th>
<th>V&lt;sub&gt;max&lt;/sub&gt; [pkat µg&lt;sup&gt;-1&lt;/sup&gt;]</th>
<th>k&lt;sub&gt;cat&lt;/sub&gt;/K&lt;sub&gt;m&lt;/sub&gt; [M&lt;sup&gt;-1&lt;/sup&gt; S&lt;sup&gt;-1&lt;/sup&gt;]</th>
<th>pha2&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT1</td>
<td>Arogenate</td>
<td>31.00</td>
<td>1050</td>
<td>0.28</td>
<td>38</td>
<td>Yes</td>
</tr>
<tr>
<td>ADT2</td>
<td>Arogenate</td>
<td>60.60</td>
<td>7650</td>
<td>1.6</td>
<td>240</td>
<td>Yes</td>
</tr>
<tr>
<td>ADT3</td>
<td>Arogenate</td>
<td>5.17</td>
<td>1140</td>
<td>-</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>ADT4</td>
<td>Arogenate</td>
<td>52.32</td>
<td>490</td>
<td>-</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>ADT5</td>
<td>Arogenate</td>
<td>11.26</td>
<td>620</td>
<td>-</td>
<td>-</td>
<td>No</td>
</tr>
<tr>
<td>ADT6</td>
<td>Arogenate</td>
<td>42.61</td>
<td>1560</td>
<td>0.4</td>
<td>16</td>
<td>No</td>
</tr>
</tbody>
</table>

<sup>a</sup>: data from Cho et al., 2007.

<sup>b</sup>: data from Bross et al., 2011.
accept prephenate as a substrate (Maeda et al. 2010). The catalytic efficiency values for arogenate recognition of these enzymes were approximately 100 and 200 times greater than those for prephenate recognition. Before the results presented in this study, the specific amino acid sequence(s) that can distinguish an arogenate-only ADT from an arogenate- and prephenate-accepting ADT had not been identified.

1.4.2 Substrate specificity of ADTs in vivo

ADT1-ADT6 were subsequently tested for their ability to recognize prephenate as a substrate in a yeast complementation assay in vivo (Bross et al. 2011). Saccharomyces cerevisiae, also called “bakers yeast”, utilizes only the prephenate pathway to produce Phe. The unicellular eukaryote contains one PDT gene, PHA2, which encodes for the only PDT protein in this organism. When PHA2 is knocked out, the yeast cannot survive without either an exogenous supply of Phe or a transformed gene that expresses a protein with PDT activity.

The six ADT enzymes were expressed in a pha2 knockout S. cerevisiae strain. ADT1 and ADT2 were able to complement the PDT knockout phenotype, while ADT3-ADT6 were not. Although ADT6 displayed PDT function in the in vitro biochemical assays, it had the lowest prephenate catalytic efficiency (Table 2), and therefore it was understandable that it could not support growth in the in vivo assays. Thus, it was concluded that ADT1 and ADT2 are preferentially arogenate-accepting enzymes with low levels of prephenate recognition, and ADT3-ADT6 are solely arogenate-accepting enzymes (Bross et al. 2011).

1.5 ADT and PDT domain organization

Domain structures have been defined for both the plant ADTs and the bacterial PDTs (Cho et al. 2007)Figure 3). Both enzymes contain a catalytic domain and a C-terminal ACT domain, which is named after three of the proteins in its domain family (Aspartokinase, Chorismate mutase, TyrA).
In bacterial PDTs, the catalytic domain binds the prephenate substrate (Zhang et al. 2000; Hsu et al. 2004) and therefore it is hypothesized to be involved in substrate-binding in plant ADTs as well. Typically, ACT domain dimers bind specific amino acid ligands at their interface, resulting in the either the activation of inhibition of enzymatic function (Aravind and Koonin 1999; Liberles et al. 2005). ACT domain homologues are found in a wide range of enzymes that are regulated by amino acids, such as valine-inhibited acetalactate synthase, lysine-inhibited aspartokinase, and Phe-inhibited phenylalanine-4-hydroxylase (Chipman and Shaanan 2001). In addition to the catalytic and ACT domain, plant ADTs contain an N-terminal transit peptide. The transit peptide was defined based on ~115 amino acids that don't align with the bacterial PDTs and has been shown to target ADTs to the chloroplasts (Cho et al. 2007; Bross 2011).

When comparing PDTs across bacteria species, the ACT domain is more highly conserved than the PDT domain (Tan et al. 2008). When comparing the six Arabidopsis ADTs, the catalytic and ACT domains are the most highly conserved domains (62.0-97.8% and 61.5-91.7% similarity, respectively), whereas the transit peptide sequences are far more diverse (Crawley 2004)

1.6 ADT and PDT 3D structure

Three crystal structures exist in the literature for bacterial PDTs (Tan et al. 2008; Vivan et al. 2008). These PDTs come from the bacterial species Mycobacterium tuberculosis, S. aureus, and C. tepidum. Although these enzymes have low protein sequence identity (32.22% overall) they have highly similar overall 3D structures. The structures support the domain organization just described (section 1.5). The catalytic domain is further divided into two subdomains: the PDTa and PDTb subdomains, which are each comprised of a central β-sheet in between three α-helices (Tan et al. 2008). The PDTb subdomain is inserted within the PDTa via two linker regions, and in between these two subdomains lies the catalytic cleft. This solvent accessible cleft is the site of prephenate binding in the bacterial PDT. The ACT domain is connected to the PDTa subdomain and interacts with this subdomain extensively.
Figure 3. Schematic of the domain structure of plant ADTs and bacterial PDTs.

Both the plant and bacterial enzymes contain a catalytic domain (blue) and a C-terminal ACT (Aspartokinase, Chorismate mutase, Tyr A) regulatory domain (green). In addition, the plant enzymes contain an N-terminal transit peptide (orange) for localization to the chloroplast.
Transit peptide | ADT/PDT catalytic domain | ACT regulatory domain

PLANT

BACTERIA
Figure 4. Ct-PDT monomer and dimer 3D conformation.

(A) Top: Ribbon drawing of the *Chlorobium tepidum* PDT monomer. The catalytic domain is divided into two subdomains: the PDTa and PDTb subdomains. The catalytic cleft is located in between the PDTa and PDTb subdomains and the ACT domain is attached to the PDTa subdomain. Bottom: schematic of the PDT monomer in the closed (T) state (top view).

(B) Top: Ribbon drawing of the *Chlorobium tepidum* PDT homodimer, consisting of two monomers (blue and red). The dimer is shown in the closed (T) state and two phenylalanine molecules are shown bound to two pockets at the ACT dimer interface (beige). The catalytic clefts are indicated in the aligned PDTa subdomains. Bottom: schematic of the PDT dimer in the closed (T) state (top view). Original PDB file (2QMX) is from Tan *et al.*, 2008 and images are obtained from UCSF Chimera.
Two PDT monomers form a homodimer, the basic catalytic unit of the PDT enzyme (Tan et al. 2008; Vivan et al. 2008), with extensive interactions across both PDT and ACT domains (Figure 4B, top). The two catalytic clefts, one from each PDT monomer, align to form an extended active site across the interface of the PDT domain dimer (Tan et al. 2008). These clefts are in close proximity, indicating that there may be some communication between the two active sites during the decarboxylation/dehydration reaction. The ACT domains of the two monomers also align, creating two hydrophobic pockets at the ACT dimer interface. Two Phe molecules can simultaneously bind these hydrophobic pockets. When Phe is absent, the dimer exists in an open (“Relaxed”) state in which the prephenate substrate has access to the extended catalytic cleft formed between the two monomers. However, the binding of the two Phe molecules to the ACT dimer interface induces an overall conformational shift from this open (R) state to a closed (“Tense”) state, in which access to the cleft is greatly reduced (Tan et al. 2008; Figure 4, bottom; Figure 5). During this shift, the two ACT domains move closer to each other. This conformational change is propagated along the ACT-PDT interface, resulting in a shift in the PDT domains as well. The two PDTb subdomains push towards each other and the two PDTa subdomains move away from each other. This movement offsets the alignment of the extended catalytic cleft, reducing the access of prephenate to and the release of phenylpyruvate from the catalytic site and results in the allosteric inhibition of enzymatic function. Interestingly, PDTs are also able to form tetramers in solution, although the significance of these structures and their involvement in the allosteric conformational change is not currently known (Tan et al. 2008).

Although there are no crystal structures available for the plant ADTs, the basic 3D structure of these enzymes can be estimated by homology modeling against the known PDT protein structures. For example, the ADT5 monomer can be modeled against the Ct-PDT monomer protein (Figure 6) using a 3D homology modeling program called UCSF Chimera. This program predicts that
Figure 5. Schematic of open (R) and closed (T) states of the PDT enzyme.

A schematic of the allosteric regulation mechanism that is triggered when phenylalanine binds to the ACT domain interface of the bacterial PDT tetramer. Shown is the PDT homodimer, made up of two monomers (blue and red; bottom view). Binding of two phenylalanine molecules at the ACT interface between the two dimers results in an overall conformation shift that affects substrate binding in the catalytic sites of the complex. Adapted from Tan et al., 2008.
Figure 6. 3D homology modeling of ADT5 and Ct-PDT.

(A) Ribbon diagram of the overlay of the PDT monomer from *Chlorobium tepidum* (blue) and homology modeled ADT5 from *Arabidopsis thaliana* (pink).
(B) Ribbon diagram of homology modeled ADT5 (Model #1.2) from *Arabidopsis thaliana*. Original Ct-PDT PDB file is from Tan et al., 2008; PDB: 2QMX, images are obtained from UCSF Chimera and the homology modeled ribbon diagram was created using UCSF Chimera's “Modeller” program.
although these two proteins share only 41.5% sequence identity, they have a very similar overall monomer structure.

1.7 The importance of few amino acids

The diversity in the bacterial PDT sequences makes it difficult to determine the amino acids that are important for prephenate substrate recognition. As a consequence of this, the exact amino acids responsible for prephenate recognition in plant ADTs are currently unknown. In *Arabidopsis*, two of the six members of the ADT family display both ADT and PDT function. Therefore, there must be some differences on the sequence level that can be used to identify this.

Evidence in the literature suggests that enzymatic function can be converted between two similar proteins by introducing only a few amino acid substitutions to the protein sequence (Jäckel *et al.* 2008). This raises the question of where to look within the *Arabidopsis* ADTs for the amino acids that are responsible for PDT activity in ADT1 and ADT2. It seems intuitive that these amino acids would be located within the catalytic cleft, the part of the protein that is responsible for binding the substrate. However, it has been shown in other studies that mutations do not need to be within a proposed catalytic site to affect substrate specificity (Jeffery *et al.* 2000; Mendonça and Marana 2011). Therefore, in order to be exhaustive, the entire ADT protein sequence must be analyzed to find the amino acid(s) that can distinguish a prephenate-accepting ADT from a solely arogenate-accepting ADT.

1.8 Thesis hypothesis and objectives

ADTs and PDTs have many similarities. They have similar domain organization, they both catalyze a decarboxylation/dehydration reaction, and they recognize substrates that differ by only the presence of a single amino group. Due to these similarities, I hypothesize that PDT function can be introduced to an
aroxygenate-only *Arabidopsis* ADT (ADT3-ADT6) through the alteration of only a few amino acids. This research will identify the amino acid(s) that can distinguish between a prephenate-accepting ADT like ADT1 and ADT2 and a solely arogenate-accepting ADT like ADT3 – ADT6.

My first objective is to narrow down the domain (transit peptide, catalytic, or ACT) that contains the sequences for prephenate recognition in the *Arabidopsis* ADTs. This will be accomplished by generating ADT chimeras containing domains from both a prephenate-accepting ADT (ADT1/ADT2) and an arogenate-specific ADT (ADT3-ADT6). Since it has been demonstrated that biochemical analysis may not reflect an *in vivo* system (Yao et al. 2007; Bross et al. 2011), I will test for PDT function of these ADT chimeras in a pha2 yeast complementation assay. Only chimeras with the domain that contains sequences that can confer PDT function will be able to sustain growth of this PDT knockout yeast strain.

Once the domain(s) have been identified, my second objective is to revisit the sequence analysis and, through targeted amino acid substitutions and another yeast complementation assay, identify the exact amino acid(s) that can distinguish between a prephenate-accepting and a solely arogenate-accepting ADT. I have chosen ADT2 (both arogenate- and prephenate-accepting) and ADT4 and ADT5 (solely arogenate-accepting) for this research.

### 2 Materials and Methods

#### 2.1 Media, buffers and gels

##### 2.1.1 Media

For the creation of all solid media, 20 g of agar was added to 1 L of liquid media.

*Synthetic dextrose (SD) medium*
Per 1 L: 20 g glucose, 6.7 g yeast nitrogen base without amino acids, 1.5 g dropout powder (Amberg et al. 2005). Glucose was substituted with equal weights of galactose or raffinose as required.

**Yeast peptone dextrose (YPD) medium**

Per 1 L: 20 g glucose, 10 g yeast extract, 20 g bacto-peptone (Amberg et al. 2005).

**Amino acid dropout powder**

Per 25 L: 1 g adenine sulphate, 0.5 g L-arginine-HCL, 2.5 g L-aspartic acid, 2.5 g L-glutamic acid, 0.75 g L-isoleucine, 1.5 g L-leucine, 0.75 g L-lysine-HCL, 0.5 g L-methionine, 1.25 g phenylalanine, 9.35 g L-serine, 5 g L-threonine, 1 g L-tryptophan, 0.75 g L-tyrosine, 0.5 g uracil, 3.725 g L-valine (Sherman et al., 1983), 2 mL histidine (His) liquid stock.

Drop-out powder containing the appropriate amino acids for selection and functional testing were made from the above amino acids. 1.5 g of drop-out powder was used per 1 L of SD medium before autoclaving.

**Lysogeny broth (LB)**

Per 1 L: 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl, 200 µL 5 N NaOH (Sambrook and Russell 2001).

**Super Optimal Catabolite-Repression (SOC) medium**

Per 1 L: 20 g bacto-tryptone, 5 g yeast extract, 2 mL 5 M NaCl, 2.5 mL 5 M KCl, 10 mL 1 M MgCl₂, 10 mL 1 M MgSO₄, 20 mL 1 M glucose.

**Antibiotics**

Stock solutions of ampicillin (100 mg/mL), gentamycin (50 mg/mL) and kanamycin (60 mg/mL) were prepared by dissolving each in double distilled water. Dissolved antibiotics were filter sterilized (Sambrook and Russell 2001), aliquoted and stored at -20°C. Stocks were added to the media after autoclaving to a final concentration of 100 µg/mL, 15 µg/mL and 60 µg/mL for ampicillin, gentamycin and kanamycin, respectively.

2.1.2 Buffers

5x Sodium dodecyl sulfate protein sample buffer
Per 10 mL: 1.2 mL of 0.5 M Tris-HCl (pH 6.8), 2.5 mL 100% glycerol, 2 mL 10% sodium dodecyl sulfate (SDS), 0.25 mL β-mercaptoethanol (BME), 0.5 mL 2% (w/v) bromophenol blue, 3.55 mL double distilled water (Sambrook and Russell 2001).

**50x TAE buffer**

Per 1 L: 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5 M ethylenediaminetetraacetic acid (EDTA) (pH 8.0), 1 L double distilled water. 50X TAE buffer was diluted to 1X with sterile double distilled water before use.

**Phosphate-buffered saline (PBS)/Tween-20**

Per 1 L: 8 g NaCl, 0.2 g KCl, 1.44 g Na$_2$HPO$_4$, 0.24 g KH$_2$PO$_4$, 0.5 mL Tween-20 (Sambrook and Russell 2001). pH was adjusted to 7.4 with HCl.

**Yeast breaking buffer**

Per 1 L: 500 mL of 0.1 M sodium phosphate (pH 7.4), 2 mL of 0.5 EDTA (pH 8.0), 250 mL of 20% glycerol, 10 mL of 100 mM phenylmethanesulfonylfluoride (PMSF), 238 mL double distilled water (pYEST2/NT vector manual, Invitrogen Cat. No. V8252-20).

### 2.1.3 Gels

**Agarose separating gel**

1% (w/v) agarose gels were used to separate and visualize PCR products and fragments from restriction enzyme digests.

Per 30 mL: 30 mL 1x TAE buffer, 0.3 g of electrophoresis-grade agarose.

**Stacking gel**

10% (w/v) acrylamide separating gels were used to resolve proteins from *S. cerevisiae* (Sambrook and Russell 2001).

Per 20 mL: 7.9 mL double distilled water, 6.7 mL 30% (w/v) acrylamide mix, 5 mL 1.5 M Tris (pH 8.8), 0.2 mL 10% SDS, 0.2 mL 10% ammonium persulfate (APS), 8 µL tetramethylethylenediamine (TEMED).

A 5% (w/v) stacking gel was also cast over the top of the 10% (w/v) separating gel. This allowed the loaded sample proteins to be concentrated into a tight band during the first few minutes of electrophoresis before entering the separating portion of the gel.
Per 10 mL: 6.8 mL water, 1.7 mL 30% acrylamide mix, 1.25 mL 1.0 M Tris (pH 6.8), 0.1 mL 10% SDS, 0.1 mL 10% APS, 10 µL TEMED (Sambrook and Russell 2001).

2.2 Strains and plasmids

2.2.1 Strains

_E. coli_ DH5α and DH10β strains (Invitrogen Cat. No. 11319019 and 18290015, respectively) were used for maintenance and cloning of plasmid DNA. Liquid cultures were grown at 250 RPM at 37°C in LB. Cells harboring plasmids were grown in media supplemented with appropriate antibiotics for selection.

The “pha2” haploid PDT knockout _S. cerevisiae_ strain (pha2: Matα, _his3Δ1_, _leu2Δ0_, _lys2Δ0_, _ura3Δ0_, YNL316c::kanMX; Saccharomyces Genome Database; Invitrogen Cat. No. YSC1021-549308) was used for complementation assays. Yeast were grown at 30°C and liquid yeast cultures were grown at 220 RPM.

2.2.2 Plasmids

The full-length _ADT_ genes were cloned into a modified pEZT-NL vector (D. Ehrhardt, Carnegie Institution of Washington), where the 3’ GFP sequence was replaced with a CFP sequence by a previous student in the lab to allow the expression of a C-terminal CFP fusion protein (Bross _et al._ 2011). The inserts contain a start codon and lack a stop codon, to allow for the continuous expression of the CFP fusion protein.

The pDONR™221 (Figure 7A; Invitrogen Cat. No. 12536-017) vector was used as a donor vector. This vector contains an kanamycin marker for selection in _E. coli_.

The pAG423GAL-ccdB-CFP yeast-compatible vector (Figure 7B; Alberti _et al._ 2007) was used as a destination vector. This vector contains an ampicillin resistance marker for selection in _E. coli_ and a His marker for selection in _S. cerevisiae_.

Figure 7. Gateway® compatible donor vector and destination vector.

(A) Diagram of pDONR™221 donor vector which contains attP sites for recombination with the attB sites that flank the PCR product, to generate the entry vector. Upon recombination, the ccdB gene is displaced by the gene insert. The KanR resistance gene is used for selection of the recombined entry vector in E. coli. (B) Diagram of the pAG423GAL-ccdB-ECFP destination vector. This vector contains attR sites for recombination with the attL sites in the entry vector to generate the expression vector. Upon recombination the ccdB gene is displaced by the gene insert. The AmpR resistance gene is used for selection of the recombined expression vector in E. coli and the HIS3 gene is used for selection of yeast the recombined expression vector.
A

pDONR\textsuperscript{TM}221

$Kan^R$

attP2

ccdB

attP1

B

pAG423GAL-ccdB-ECFP

$HIS3$

ECFP

attR2

ccdB

attR1

$GAL1$

$Amp^R$
2.3 Cloning procedure

Gateway® Cloning recombination technology was used to recombine the ADT inserts into a yeast compatible destination vector following the manufacturer's manual (Invitrogen MAN0000282; Figure 8). Gateway® BP Clonase® II Enzyme mix (Invitrogen Cat. No. 11789020) was used for the recombination of the full-length ADTs and ADT chimeras into the pDONR™221 vector to generate the entry vector. Gateway® LR Clonase® II Enzyme mix (Invitrogen Cat. No. 11791100) was used for the recombination of the inserts into the pAG423GAL-ccdb-ECFP yeast-compatible destination vector.

2.4 PCR amplification

2.4.1 Primer Design:
All primers were tested for self-complementarity, primer-pair complementarity, balanced GC content and similar melting temperatures within primer-pairs using DNAMAN (Lynnon BioSoft, Version 6; Table 3, Table 4).

2.4.2 Overlap extension PCR
Chimeras containing domains from both ADT2 and ADT4, or ADT2 and ADT5 were generated through overlap extension PCR (Figure 9). All PCR conditions were established using Tfi polymerase (Invitrogen Cat. No. 30342024) prior to amplification with Platinum Taq High Fidelity Polymerase (Invitrogen Cat. No. 11304011) (Table 5). PCR products were amplified using previously cloned ADT full-length gene sequences as a template (Bross 2011), as previously stated (section 2.2.2). Reactions were performed using a Touchgene Gradient Thermocycler (Techne TC-512; Cat. No. 353.55144.E1).

Annealing temperatures varied based on the primers that were used in the reaction (Table 3; Table 4), and therefore are denoted as X in Table 6.
Figure 8. Schematic of Gateway® cloning strategy.

(A) Each construct is recombined into the pDONRTM221 vector (blue). This is facilitated by the outer att recombination sites (orange) in a BP reaction, generates an entry vector. (B) The insert is then recombined from the entry vector into the pAG423GAL-ccdB-ECFP yeast compatible destination vector (green) using the att recombination sites (orange) in an LR reaction. This reaction generates an expression vector. Each expression vector is then transformed into pha2 yeast for the complementation assay.
A

Chimeras

BP reaction

pDONRTM221

B

pAG423GAL-ccdB-ECFP

Entry Vector

LR reaction

Expression Vector
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' to 3')</th>
<th>Direction</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>attB1-ADT2FL*</td>
<td>GGGGACAAAGTTTGTAACAAAAAAGCAGGCTTTATGGCAATGCACACTGTT</td>
<td>FWD</td>
<td>65</td>
</tr>
<tr>
<td>attB2-ADT2FL*</td>
<td>GGGGACCACTTTGTACAAAGAAAGCTGGGTAGAGCAATTAGTTG</td>
<td>REV</td>
<td>65</td>
</tr>
<tr>
<td>attB1-ADT4FL*</td>
<td>GGGGACAAAGTTTGTAACAAAAAAGCAGGCTAGATGCAACGCTCAGCAGCAGGCAGCCG</td>
<td>FWD</td>
<td>65</td>
</tr>
<tr>
<td>attB2-ADT4FL*</td>
<td>GGGGACCACTTTGTACAAAGAAAGCTGGGTAGAGCAATTAGTTG</td>
<td>REV</td>
<td>65</td>
</tr>
<tr>
<td>attB1-ADT5FL*</td>
<td>GGGGACAAAGTTTGTAACAAAAAAGCAGGCTATGCAAACCATTTCGCG</td>
<td>FWD</td>
<td>67</td>
</tr>
<tr>
<td>attB2-ADT5FL*</td>
<td>GGGGACCACTTTGTACAAAGAAAGCTGGGTAGAGCAATTAGTTG</td>
<td>REV</td>
<td>67</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence (5’ to 3’)[b]</td>
<td>Direction[c]</td>
<td>Annealing Temperature (°C)[d]</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------</td>
<td>--------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>ADT2/4CATACTF</td>
<td>GCACCGTTGCATGGTTCTAGTCTACGTGTTCGCATCAGGG AGTACGAGGT</td>
<td>FWD</td>
<td>59</td>
</tr>
<tr>
<td>ADT4/2TPR</td>
<td>TCGTACTCCCTGATACGCAACACGTAGACTAGACCATGCA ACGGTGC</td>
<td>REV</td>
<td>59</td>
</tr>
<tr>
<td>ADT2/4TPCATR</td>
<td>GGCAGCAAAGACGATGCTAGTCTTTGAAGAGTCTGTTAGTTG CAGGGATAATGGG</td>
<td>REV</td>
<td>59</td>
</tr>
<tr>
<td>ADT4/2ACTF</td>
<td>CCCATTATCCCTGGAACCTAACAAGACTCTTTCAAGACTAGCAT CGTCTTGGCTGCC</td>
<td>FWD</td>
<td>59</td>
</tr>
<tr>
<td>ADT2/4ACTF</td>
<td>CCTATAATCCACCGACCGACCGGCCATTAAAGACAAGTAT AGTTTTCTCGTTA</td>
<td>FWD</td>
<td>59</td>
</tr>
</tbody>
</table>
Table 3. (continued) Primer sequences for domain swapping.

<table>
<thead>
<tr>
<th>Name^a</th>
<th>Sequence (5' to 3')^b</th>
<th>Direction^c</th>
<th>Annealing Temperature (°C)^d</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT2/4TPR</td>
<td>GGGACGCCTTGGTAAGCTACACGAACGGCGGAGCCATT AGAGACAGA</td>
<td>REV</td>
<td>59</td>
</tr>
<tr>
<td>ADT4/2CATACTF</td>
<td>TCTGTCTCTAATGGCTCCCGCTTTCTGTAGCTACAAGG</td>
<td>FWD</td>
<td>59</td>
</tr>
<tr>
<td>ADT5/2MACTR</td>
<td>CCTCCAGAGGCTCAGGAGGATCTTCTATCCTGCGCGGC</td>
<td>REV</td>
<td>67</td>
</tr>
<tr>
<td>ADT2/5MACTF</td>
<td>TTGACGAAAATCGAATCAGGCGGCTAGGAAACATCCTC TCCGACATCTGGAGG</td>
<td>FWD</td>
<td>62</td>
</tr>
<tr>
<td>ADT2/5MACTR</td>
<td>CGCCTACGACTCTAAGGGGACAATTCTCTGAAAGGCGGC</td>
<td>REV</td>
<td>62</td>
</tr>
<tr>
<td>ADT5/2MACTF</td>
<td>CTCACAAAGATTGGAAAGCCGCCCTTTACAGAATTGTCCC</td>
<td>FWD</td>
<td>65</td>
</tr>
</tbody>
</table>
Table 3. (continued). Primer sequences for domain swapping.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Direction</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT5/2MCATR</td>
<td>GGCTAAGCACCCTTCTCAAATCTTTCAATACGAACTCCGGAAGAGCTAAGAGACAG</td>
<td>REV</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>CTCCGGGAAGAGCTAAGAGACAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADT2/5MCATF</td>
<td>CTGTCTCTTAGCTCTCCAAGTGCTATTGAAGATTGGAGAAGGGTGCTTAGCC</td>
<td>FWD</td>
<td>62</td>
</tr>
<tr>
<td>ADT2/5TPR</td>
<td>AACTCCTTGATACGCGACACG</td>
<td>REV</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>AACGCGGGAGCCATTAGAGACAGGGTGCTTAGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADT5/2CATACTF</td>
<td>TGTCTCTATGCGTCCCGCCTTGCTGTGCTGTATCAAGGAGTGTATGTT</td>
<td>FWD</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>AGCGGTCGTTTAGCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADT2/5TPCATR</td>
<td>GAAAAACTATACTTTGCTTTAAAGAGTCTGGTAGTCCCCAGGAT</td>
<td>REV</td>
<td>62</td>
</tr>
<tr>
<td>ADT5/2ACTF</td>
<td>ATTATCCCTGGAACATACAGACTCTTTAAGACAAGTATGTTTTTCG</td>
<td>FWD</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>AGCGGTCGTTTAGCC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3. (continued). Primer sequences for domain swapping.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)</th>
<th>Direction</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT2/5CATACTF</td>
<td>CGCCGTCTCATGGATCTACTCTTTCCGTTG TGCCTATCAGGAGTACG</td>
<td>FWD</td>
<td>65</td>
</tr>
<tr>
<td>ADT5/2TPR</td>
<td>TACTCCCTGATACGCAACACGAAGAGTA GATCCATGAGACGG</td>
<td>REV</td>
<td>65</td>
</tr>
<tr>
<td>ADT2/5ACTF</td>
<td>ATCATTCCCCGTACCGATCGTCCCGTTTA AGACAAAGTATAGTTTTCTCG</td>
<td>FWD</td>
<td>62</td>
</tr>
<tr>
<td>ADT5/2TPCATR</td>
<td>AGAAAAACTTACTTGTCTTTAAACGGACGA TCGGTCGGGGGAATG</td>
<td>REV</td>
<td>62</td>
</tr>
</tbody>
</table>

*a*: outer primer; *b* double underline: 5’ GGGG for docking of Gateway® enzymes; *bold*: attB1/attB2 sequences; *bold and underlined*: added to maintain frame; *italics*: complementary to ADT2; *underlined*: complementary to ADT5; non-annotated: complementary to ADT4; *FWD*: forward amplification; *REV*: reverse amplification. *Annealing temperatures listed for outer primers were used for full-length ADT and chimera amplification, and those listed for the inner primers were used for overlap extension PCR.*
Table 4. Primer sequences for targeted mutagenesis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' to 3')</th>
<th>Direction</th>
<th>Annealing Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-V397LF</td>
<td>CGCGTTAGCGGAGTTGCAAGAGTACACGTCATTCC</td>
<td>FWD</td>
<td>60</td>
</tr>
<tr>
<td>5-V397LR</td>
<td>GGAATGACGTGTACTCTTTGCAACTCCGCTAACGCG</td>
<td>REV</td>
<td>63</td>
</tr>
<tr>
<td>4-V396LF</td>
<td>CGCGCTTTCGGAGCTTCAAGAGTACACGTCGTTCC</td>
<td>FWD</td>
<td>60</td>
</tr>
<tr>
<td>4-V396LR</td>
<td>GGAACGACGTGTACTCTTTGGAACTCCGCAACGCGG</td>
<td>REV</td>
<td>60</td>
</tr>
<tr>
<td>5-F341LF</td>
<td>GTGCTTTCCGCGTTTGCCTTCAAAACATTAAGGGTTGACG</td>
<td>FWD</td>
<td>62</td>
</tr>
<tr>
<td>5-F341LR</td>
<td>CGTCAAACATGTTTGCAGCGAAGCGGAAACGCGAAGCAC</td>
<td>REV</td>
<td>62</td>
</tr>
<tr>
<td>5-S345NF</td>
<td>CCGTTTGCCTTCGAAACATTAAGGGTTGACGAAAATCGAATCG</td>
<td>FWD</td>
<td>64</td>
</tr>
</tbody>
</table>

*Note: The bold sequence indicates the mutation targeted.*
Table 4. (continued). Primer sequences for targeted mutagenesis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’ to 3’)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Direction&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Annealing Temperature (°C)&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-S345NR</td>
<td>CGGCGCGATCGATTCTCGTCAAATTAATGGTTTCCGAAACGCA</td>
<td>REV</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>AACGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-DM-F</td>
<td>GTGGCTGTCGCTTACGCGTCGAAACATTAATTGACGAAA</td>
<td>FWD</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>TCGAATCGCGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-DM-R</td>
<td>CCGCGATCTGATTCTCGTCAAATTAATGGTTTCCGAAAGCGCAAAC</td>
<td>REV</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>GCGGAAAGCAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Bold and double-underlined: nucleotide substitutions; Underlined: complementary to ADT5; non-underlined: complementary to ADT4; <sup>b</sup>FWD: forward amplification; REV: amplification. These primers were used in combination with the outer primers listed in Table 2. <sup>c</sup>Annealing temperature for overlap extension PCR.
Figure 9. Schematic of overlap extension PCR.

(A) Fragments containing 15-24 bp of overlapping sequences are generated in a first set of PCRs using outer primers that are specific to the template (straight arrows) and inner primers that contain a combination of sequences from both of the ADT templates (curved arrows). (B) In a third PCR reaction, both fragments that were generated in (A) are used as templates. The overlapping sequences facilitate the annealing of the fragments, and the DNA polymerase extends from the fragments’ 3’ ends, generating the full-length sequence. Only the outer primers are used in this reaction, ensuring that the full-length chimera is amplified.
A

**ADT2**

```
TP  CAT  ACT
```

```
5'     3'
```

```
5'     3'
```

```
5'     3'
```

**ADT4**

```
TP  CAT  ACT
```

```
5'     3'
```

```
5'     3'
```

B

**Chimera**

```
5'     3'
```

```
5'     3'
```

```
5'     3'
```

```
5'     3'
```
Table 5. Components for fragment amplification and overlap extension PCR.

<table>
<thead>
<tr>
<th>Component</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 High Fidelity PCR Buffer</td>
<td>1X</td>
</tr>
<tr>
<td>10 mM MgSO₄</td>
<td>2.0 mM</td>
</tr>
<tr>
<td>10 mM dNTP Mix</td>
<td>0.2 mM each</td>
</tr>
<tr>
<td>10 µM Forward Primer</td>
<td>0.2 µM</td>
</tr>
<tr>
<td>10 µM Reverse Primer</td>
<td>0.2 µM</td>
</tr>
<tr>
<td>Template DNA*</td>
<td>&lt;500 ng</td>
</tr>
<tr>
<td>Platinum® Taq DNA Polymerase</td>
<td>1 U/rxn</td>
</tr>
<tr>
<td>High Fidelity (5 U/µL)</td>
<td></td>
</tr>
<tr>
<td>Autoclaved, distilled water</td>
<td>-</td>
</tr>
</tbody>
</table>

*For overlap extension PCR reactions: the two templates to be recombined were both added to this final concentration.
Table 6. Thermocycler conditions for fragment amplification and overlap extension PCR.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Temperature (°C)</th>
<th>Time</th>
<th>Number of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94</td>
<td>10 min</td>
<td>1</td>
</tr>
<tr>
<td>Denature</td>
<td>94</td>
<td>15 secs</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>55</td>
<td>30 secs</td>
<td>5*</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>Denature</td>
<td>94</td>
<td>15 secs</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>X</td>
<td>30 secs</td>
<td>30</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72</td>
<td>7 min</td>
<td>1</td>
</tr>
</tbody>
</table>

*For 5-SLEEG and 5-F341L overlap extension PCR: the number of cycles for these PCR reactions was increased to 10 and primers were not added until after these 10 cycles were complete.
The 5-SLEEG and 5-F341L full-length constructs (see section 2.5 for more information on these chimeras) could not be successfully amplified through overlap extension PCR with the standard cycle conditions, and so a modified protocol was used. In this modified protocol, the number of cycles at a 55°C annealing temperature was increased from five to ten, and primers were only added to the reaction after these cycles were complete (Table 6). This was done to ensure that the overlapping sequences would anneal first and the DNA polymerase would extend from the 3’ ends of the fragments, generating the full-length product. Then, once the primers were added, this full-length product was used as the template.

M13 primers were used for sequencing all of the ADT inserts in the entry vectors.

2.5 Chimera name annotation

The chimeras are named based on the protein sequences. All descriptions and positions correspond to the amino acid level.

The six chimeras that contained domain sequences from ADT2 and ADT4 were the first six chimeras to be generated, and therefore these constructs are named Chimera 1 – Chimera 6 (Figure 10). The next seven chimeras that were generated contained domain sequences from ADT2 and ADT5 and were named Chimera 7 – Chimera 13 (Figure 11).

The chimeras that swap the nucleotides that correspond to the SLEEG motif in the ACT domain of ADT2 and ADT5 were named 2-SLEEG and 5-SLEEG, respectively. 2-SLEEG has the nucleotide sequences that correspond to the entire ADT2 transit peptide and catalytic domain, and the N-terminal ACT domain sequence until G299, ADT5 sequences from the corresponding ADT5 amino acid (S331) make up the C-terminal end of the ACT domain in this chimera. 5-SLEEG has the nucleotide sequences that correspond to the entire ADT5 transit peptide and catalytic domain, and the N-terminal ACT domain
Figure 10. Schematics for ADT2/ADT4 chimeras.

Full-length ADT2 (white) and ADT4 (purple) were assembled through overlap extension PCR to generate Chimera 1 – Chimera 6, which have every possible combination of the three ADT domains.

TP: transit peptide; CAT: catalytic domain; ACT: ACT regulatory domain.
Figure 11. Schematics for ADT2/ADT5 chimeras.

Full-length ADT2 (white) and ADT5 (green) were assembled through overlap extension PCR to generate Chimera 7 – Chimera 13 which have every possible combination of the three ADT domains.

TP: transit peptide; CAT: catalytic domain; ACT: ACT regulatory domain.
sequences until S331 and ADT2 sequences make up G299 to the C-terminal ACT domain (Figure 12).

The chimeras that contain the nucleotide mutations that result in a single amino acid substitution are named as follows: *(the ADT template) – (native amino acid) (position) (substituted amino acid).* For example, the chimera named 5-F341L contains the nucleotide sequences that express the transit peptide, catalytic and ACT domains from ADT5 with the exception that a Phe has been substituted for a Leu at amino acid position 341 (Figure 12).

For the overlap extension PCR, the fragments that contain overlapping sequences for the amplification of the ADT construct are named “*Fragment Xa*” or “*Fragment Xb*”, where X represents the chimera number or name. For example, the overlapping fragments that were used as templates to generate Chimera 1 through overlap extension PCR are named *Fragment 1a* and *Fragment 1b*.

### 2.6 Gel purification and quantification of DNA

DNA fragments were size separated by gel electrophoresis on a 1% (w/v) agarose gel in TBA buffer. The DNA was stained with RedSafe™ Nucleic Acid Staining Solution (FroggaBio Cat. No. 21141) and visualized under UV light. Fragments of the correct size (Table 7; Table 8; Table 9) were excised from the gel and were purified using the PureLink Quick Gel Extraction Kit (Invitrogen Cat. No. K210012). Samples were resuspended in 50 µL of Elution Buffer (“E5” provided in the kit; 10mM Tris-HCL, pH 8.5) and the DNA concentration and the 260/280 values were measured using a Nanodrop™ 1000 Spectrophotometer (Thermo Scientific Cat. No. ND-1000). DNA fragments were considered acceptable for use in PCR or cloning if the concentration was between 25 and 75 ng/µL and the 260/280 values were between 1.7 and 1.9.

If sample concentration was between 10 ng/µL and 25 ng/µL, samples were evaporated using a Scientific™ Savant™ DNA SpeedVac™ Concentrator
Figure 12. Schematics for N-terminal ACT domain constructs.

Overlap extension PCR was used to generate 2-SLEEG, 5-SLEEG, 5-F341L, 5-S345N, 5-DM, 5-V397L, and 4-V396L constructs using ADT2, ADT4, and ADT5 templates.

TP: transit peptide; CAT: catalytic domain; ACT: ACT regulatory domain.
Table 7. Composition and expected lengths of \( ADT2/ADT4 \) chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT4 Domains</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nt</td>
</tr>
<tr>
<td><strong>Fragment 1a</strong></td>
<td>TP</td>
<td>CAT, ACT</td>
<td>315</td>
</tr>
<tr>
<td><strong>Fragment 1b</strong></td>
<td></td>
<td>CAT, ACT</td>
<td>918</td>
</tr>
<tr>
<td><strong>Fragment 2a</strong></td>
<td>CAT, ACT</td>
<td>TP</td>
<td>867</td>
</tr>
<tr>
<td><strong>Fragment 2b</strong></td>
<td></td>
<td>TP</td>
<td>393</td>
</tr>
<tr>
<td><strong>Fragment 3a</strong></td>
<td>TP, CAT</td>
<td></td>
<td>813</td>
</tr>
<tr>
<td><strong>Fragment 3b</strong></td>
<td></td>
<td>ACT</td>
<td>354</td>
</tr>
<tr>
<td><strong>Fragment 4a</strong></td>
<td>ACT</td>
<td>TP, CAT</td>
<td>329</td>
</tr>
<tr>
<td><strong>Fragment 4b</strong></td>
<td></td>
<td>ACT</td>
<td>957</td>
</tr>
<tr>
<td><strong>Fragment 5a</strong></td>
<td>CAT</td>
<td>TP</td>
<td>1020</td>
</tr>
<tr>
<td><strong>Fragment 5b</strong></td>
<td></td>
<td>ACT</td>
<td>296</td>
</tr>
<tr>
<td><strong>Fragment 6a</strong></td>
<td>TP</td>
<td>CAT</td>
<td>341</td>
</tr>
<tr>
<td><strong>Fragment 6b</strong></td>
<td>ACT</td>
<td></td>
<td>1011</td>
</tr>
</tbody>
</table>
Table 7. (continued). Composition and expected lengths of ADT2/ADT4 chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT4 Domains</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nt</td>
</tr>
<tr>
<td>FL ADT2</td>
<td>TP, CAT, ACT</td>
<td></td>
<td>1146</td>
</tr>
<tr>
<td>FL ADT4</td>
<td></td>
<td>TP, CAT, ACT</td>
<td>1275</td>
</tr>
<tr>
<td>Chimera 1</td>
<td>TP</td>
<td>CAT, ACT</td>
<td>1197</td>
</tr>
<tr>
<td>Chimera 2</td>
<td>CAT, ACT</td>
<td>TP</td>
<td>1224</td>
</tr>
<tr>
<td>Chimera 3</td>
<td>TP, CAT</td>
<td>ACT</td>
<td>1291</td>
</tr>
<tr>
<td>Chimera 4</td>
<td>ACT</td>
<td>TP, CAT</td>
<td>1210</td>
</tr>
<tr>
<td>Chimera 5</td>
<td>CAT</td>
<td>TP, ACT</td>
<td>1300</td>
</tr>
<tr>
<td>Chimera 6</td>
<td>TP, ACT</td>
<td>CAT</td>
<td>1292</td>
</tr>
</tbody>
</table>

Amino acid length and molecular weight are only listed for chimeras used in the Western Blot. nt: nucleotide length; aa: amino acid length; MW\(^a\): molecular mass in kDa of the native protein. Add 27.7 kDa to this value for the molecular weight of the ADT-ECFP fusion protein.
Table 8. Composition and expected lengths of *ADT2/ADT5* chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT5 Domains</th>
<th>Length nt</th>
<th>Length aa</th>
<th>Length MW&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Fragment 7a</em></td>
<td>TP</td>
<td>CAT, ACT</td>
<td>297</td>
<td>795</td>
<td>900</td>
</tr>
<tr>
<td><em>Fragment 7b</em></td>
<td></td>
<td>CAT, ACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 8a</em></td>
<td>TP, CAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 8b</em></td>
<td></td>
<td>ACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 9a</em></td>
<td>CAT, ACT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 9b</em></td>
<td></td>
<td>TP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 10a</em></td>
<td>ACT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 10b</em></td>
<td></td>
<td>TP, CAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fragment 11a</em></td>
<td>TP, CAT, N-ACT</td>
<td>ACT-C</td>
<td></td>
<td></td>
<td>1071</td>
</tr>
<tr>
<td><em>Fragment 11b</em></td>
<td>ACT-C</td>
<td></td>
<td></td>
<td></td>
<td>237</td>
</tr>
<tr>
<td><em>Fragment 12a</em></td>
<td>TP, CAT, N-ACT</td>
<td>ACT-C</td>
<td></td>
<td></td>
<td>1039</td>
</tr>
<tr>
<td><em>Fragment 12b</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>264</td>
</tr>
<tr>
<td><em>Fragment 13a</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>660</td>
</tr>
<tr>
<td><em>Fragment 13b</em></td>
<td>CAT-C, ACT</td>
<td></td>
<td></td>
<td></td>
<td>600</td>
</tr>
</tbody>
</table>
Table 8. (continued). Composition and expected lengths of ADT2/ADT5 chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT5 Domains</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>nt</td>
</tr>
<tr>
<td>FL ADT2</td>
<td>TP, CAT, ACT</td>
<td></td>
<td>1146</td>
</tr>
<tr>
<td>FL ADT5</td>
<td></td>
<td>TP, CAT, ACT</td>
<td>1278</td>
</tr>
<tr>
<td>Chimera 7</td>
<td>TP</td>
<td>CAT, ACT</td>
<td>1217</td>
</tr>
<tr>
<td>Chimera 8</td>
<td>TP, CAT</td>
<td>ACT</td>
<td>1361</td>
</tr>
<tr>
<td>Chimera 9</td>
<td>CAT, ACT</td>
<td>TP</td>
<td>1267</td>
</tr>
<tr>
<td>Chimera 10</td>
<td>ACT</td>
<td>TP, CAT</td>
<td>1203</td>
</tr>
<tr>
<td>Chimera 11</td>
<td>ACT-C</td>
<td>TP, CAT, N-ACT</td>
<td>1328</td>
</tr>
<tr>
<td>Chimera 12</td>
<td>TP, CAT, N-ACT</td>
<td>ACT-C</td>
<td>1223</td>
</tr>
<tr>
<td>Chimera 13</td>
<td>CAT-C, N-ACT</td>
<td>TP, N-CAT</td>
<td>1210</td>
</tr>
</tbody>
</table>

Amino acid length and molecular weight are only listed for chimeras used in the Western Blot. nt: nucleotide length; aa: amino acid length; MW*: molecular weight in kDa of the native protein. Add 27.7 kDa to this value for the molecular weight of the ADT-ECFP fusion protein; N-: N-terminal sequences; -C: C-terminal sequences.
Table 9. Composition and expected lengths of ACT domain chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT4 Domains</th>
<th>ADT5 Domains</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT-C&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>5-F341La</td>
<td></td>
<td></td>
<td>ACT-C&lt;sup&gt;b&lt;/sup&gt;</td>
<td>306</td>
</tr>
<tr>
<td>5-F341Lb</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>994</td>
</tr>
<tr>
<td>5-S345Na</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>1084</td>
</tr>
<tr>
<td>5-S345Nb</td>
<td></td>
<td>ACT-C&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>226</td>
</tr>
<tr>
<td>5-DMa</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td>1092</td>
</tr>
<tr>
<td>5-DMb</td>
<td></td>
<td>ACT-C&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td>232</td>
</tr>
<tr>
<td>5-SLEEGa</td>
<td>ACT-C&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>1052</td>
</tr>
<tr>
<td>5-SLEEGb</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td>267</td>
</tr>
<tr>
<td>2-SLEEGa</td>
<td>TP, CAT, N-ACT&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>270</td>
</tr>
<tr>
<td>2-SLEEGb</td>
<td></td>
<td>ACT-C&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
<td>1117</td>
</tr>
<tr>
<td>4-V396La</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>1207</td>
</tr>
<tr>
<td>4-V396Lb</td>
<td></td>
<td>ACT-C&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>146</td>
</tr>
<tr>
<td>5-V397La</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>1216</td>
</tr>
</tbody>
</table>
Table 9. (continued). Composition and expected lengths of ACT domain chimeras.

<table>
<thead>
<tr>
<th>Fragments and Chimeras</th>
<th>ADT2 Domains</th>
<th>ADT4 Domains</th>
<th>ADT5 Domains</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>nt</td>
</tr>
<tr>
<td>5-V397L&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>ACT-C&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>168</td>
</tr>
<tr>
<td>5-F341L</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td>1278</td>
</tr>
<tr>
<td>5-3S45N</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td>1278</td>
</tr>
<tr>
<td>5-DM</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td>1278</td>
</tr>
<tr>
<td>5-SLEEG</td>
<td>ACT-C&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td>TP, CAT, N-ACT&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1269</td>
</tr>
<tr>
<td>2-SLEEG</td>
<td>TP, CAT, N-ACT&lt;sup&gt;g&lt;/sup&gt;</td>
<td>ACT-C&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
<td>1236</td>
</tr>
<tr>
<td>4-V396L</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td>1275</td>
</tr>
<tr>
<td>5-V397L</td>
<td></td>
<td>TP, CAT, ACT&lt;sup&gt;j&lt;/sup&gt;</td>
<td></td>
<td>1278</td>
</tr>
</tbody>
</table>

Amino acid length and molecular weight are only listed for chimeras used in the Western Blot. nt: nucleotide length; aa: amino acid length; MW<sup>a</sup> molecular weight in kDa of the native protein. Add 27.7 kDa to this value for the molecular weight of the ADT-ECFP fusion protein; N-: N-terminal sequences; -C: C-terminal sequences; <sup>b</sup>: sequences with 1021T>C substitution; <sup>c</sup>: sequences with 1034G>A substitution; <sup>d</sup>: sequences with both 1021T>C and 1034G>A substitutions; <sup>e</sup>: sequences from 898G to C-terminus; <sup>f</sup>: sequences from N-terminus to 1012G; <sup>g</sup>: sequences from N-terminus to 898G; <sup>h</sup>: sequences from 1012G to C-terminus; <sup>i</sup>: sequences with 1186G>T substitution; <sup>j</sup>: sequences with 1189G>C substitution.
(Thermo Scientific Cat. No. 20-548-132) and resuspended in sterile double distilled water to increase the concentration to within the acceptable range.

2.7 Plasmid DNA isolation, DNA sequencing and analysis

Plasmid DNA was isolated from DH5α E. coli using an alkaline lysis with SDS miniprep protocol (Sambrook and Russell 2001). Plasmid DNA for sequencing was isolated using the QIAprep Spin Miniprep Kit (QIAGEN Cat. No. 12123). Isolated DNA concentration and quality was analyzed using Nanodrop™ 1000 Spectrophotometer (Thermo Scientific Cat. No. ND-1000).

Entry vectors were sent to the London Regional Genomics Centre in the Robarts Research Institute for sequencing and only inserts lacking mutations were used for functional analysis (The University of Western Ontario).

2.8 Transformations

2.8.1 E. coli transformations

DH5α E. coli cells were made chemically competent through a Rubidium Chloride Competent Cell procedure (Renzette 2011) and plasmid DNA was transformed into the competent E. coli cells using the One Shot® TOP10 Chemically Competent E. coli transformation procedure (Invitrogen Cat. No. C4040-03 Manual). All liquid cultures of E. coli were grown at 37°C and 225 RPM unless otherwise stated. Alterations to the transformation procedure were as follows: incubation time in a 37°C shaker was increased to 1.5 hours immediately following the heat shock, and the amount of cells plated on solid selective LB was increased to 250 µL.

Transformed E. coli cells were plated on solid selective LB plates and incubated at 37°C overnight. Colonies were picked the next day and grown overnight in liquid selective LB. Bacterial stocks were stored in 25% sterile glycerol at -80°C.
2.8.2 pha2 yeast transformations

*S. cerevisiae* pha2 cells were made competent and transformed using the lithium acetate/single stranded carrier DNA/polyethylene glycol method (Gietz and Woods 2002). Yeast was transformed with one of the following: an ADT chimera, full-length ADT2, ADT4 or ADT5, or the empty pAG423GAL-ccdB-ECFP destination vector. Cells were plated on appropriately supplemented solid selective SD media and incubated at 30°C for 3 to 7 days. Yeast colonies were then re-streaked on the same type of media, incubated at 30°C for 3 more days, and were stored in 15% (w/v) sterile glycerol at -80°C. For all yeast experiments, cell density was determined by counting yeast cells using a Bright-Line™ Haemacytometer (Sigma-Aldrich Cat. No. Z359629).

2.9 Protein induction, isolation and detection in *S. cerevisiae*

A Western blot assay was used to determine if the pha2 yeast was able to express the ADT proteins. Samples for this assay included yeast that contained the ADT chimeras, ADT2, ADT4 or ADT5 full-length proteins, or the empty destination vector, untransformed yeast, and non-induced yeast which contained Chimera 8 but what grown in non-inducing medium. All yeast cultures were grown in liquid media at 30°C and 225 RPM unless otherwise stated.

First, pre-cultures of the yeast samples were grown in appropriately supplemented non-inducing liquid media (raffinose; West Jr et al. 1984) for 24 hours. The cells were harvested in mid-late logarithmic growth phase (~7 x 10⁶ -7 x 10⁷) by centrifuging at 1500 x g for 5 minutes. They were then transferred to appropriately supplemented inducing liquid media (galactose; West Jr et al. 1984), with the exception of the non-induced sample which was transferred to appropriately supplemented non-inducing liquid media (glucose; West Jr et al. 1984). They were grown for approximately 24 hours and cells were collected at a cell density of 5 x 10⁶ cells/mL. Cell lysates were prepared according to the manufacturer’s instruction manual for the pYES2/NT vector (Invitrogen Cat. No. V8252-20). The total soluble protein (TSP) was quantified using a Pierce™ BCA
Protein Assay Kit (Pierce Cat. No. 23227; Smith et al. 1985) and samples were diluted to 750 µg/mL of protein and stored at -80°C in 5X denaturing SDS buffer with BME until the Western blot assay.

Fifteen µg of TSP from each sample was size separated on a 10% (w/v) SDS-PAGE. A Precision Plus Protein™ Dual Color Standards ladder (BioRad Cat. No. 161-0374) and a GFP-hydrophobin standard (1:1000 dilution; Gutiérrez et al. 2013) were also loaded. The proteins were transferred to a nitrocellulose membrane (Invitrogen Cat. No. LC2000) using a semi dry transfer method (Towbin et al. 1979) and probed with primary anti-GFP antibody (1:5000 antibody dilution of A.v. monoclonal antibody (JL-8), Clontech Cat. No. 632380). The anti-GFP antibody was used because CFP and GFP are similar enough that both can be detected with this antibody (Bross et al. 2011). Subsequently, the blot was probed with a goat anti-mouse IgG (H+L)-horseradish peroxidase conjugated antibody (1:3000 dilution, Bio-Rad Cat. No. 170-6516). An enhanced chemiluminescence (ECL) detection system (GE Healthcare Cat. No. RPN2109) was used for visualization. Western blots were exposed for 1 minute and imaged using X-ray film.

2.10 Yeast complementation assays

PDT activity was tested for in a pha2 yeast complementation assay (Figure 13). The yeast samples described in the previous section (section 2.7) were grown in appropriately supplemented liquid raffinose media for 24 hours and were harvested at a cell density of at least 5 x 10^6 cells/mL. All liquid yeast cultures were grown at 30°C and 225 RPM unless otherwise stated. Aliquots of each were then centrifuged at 13000 x g for 30 seconds and resuspended in sterile double distilled water to a final concentration of 1 x 10^5 cells/mL.

All samples and controls were tested on plates of solid SD media. These plates contained either glucose (which represses the GAL1 promoter in the destination vector), raffinose (which neither represses or induces GAL1) or galactose (which induces GAL1) as a carbon source. Plates within each carbon
Figure 13. Schematic of complementation assay to test for PDT function.

(A) ADT chimeras are recombined into the yeast compatible destination vector. \textit{pha2} yeast are transformed with the destination vector carrying an ADT insert. (B) Yeast transformants are spotted on selective solid SD media containing galactose as a carbon source to induce transcription (right-facing arrow) from the GAL1 promoter (light green). The yeast transformants were spotted on solid SD media containing Phe as a positive control (green checkmark). Transformants were spotted on solid selective SD media lacking Phe to test for PDT function (purple question mark). Note: yeast transformants were also spotted on solid SD media containing glucose or raffinose as a carbon source.
A

B

+ Galactose + Phenylalanine
+ Galactose - Phenylalanine
source were divided into three treatments: one that was fully-supplemented, one which selected for transformed yeast (lacking His), and one which tested for PDT function (lacking His and Phe). pha2 transformant liquid cultures of 10 µL were spotted on each plate and growth was documented for 30 days.

2.11 3D Homology Modeling

UCSF Chimera was used for all molecular graphics and analyses presented in this study. The UCSF Chimera program was developed by the Resource for Biocomputing, Visualization and Informatics and the University of California, San Francisco and is supported by NIGMS (P41-GM103311).

To predict the structure and the position of the amino acids for distinguishing between a prephenate –accepting ADT and a solely arogenate-accepting ADT, in ADT2 and ADT5, the ADT2 and ADT5 protein sequences were homology modeled against a protein template with a known structure. Chimera’s “BLAST Protein” tool was used to identify the protein structure that has the highest similarity to ADT2 and ADT5. The ADTs were modeled to Chain A of the PDT protein from C. tepidum (PDB file 2QMX_A; Tan et al. 2008) using the “Modeller” tool. The program used two scores, a GA341 and a zDOPE score, to measure the quality of the models that were generated with this tool. The GA341 model score is derived from statistical potentials and a value greater than 0.7 generally indicates a reliable model (>95% probability of having the correct fold (Pettersen 2004). The zDOPE score, also known as the Discrete Optimized Protein Energy (DOPE) score, is a sample of native structures that do not depend on adjustable parameters and is based on an improved reference state that accounts for the spherical shape of the native structures. This score is used to derive an atomic distance-dependent statistical potential from a sample of native structures and a more negative zDOPE score indicates a better model (Pettersen 2004).

Amino acid substitutions were made using the Chimera’s “Rotamers” tool and the Dunbrack Rotamer library (Dunbrack Jr 2002). Interatomic “clashes” (unfavourable interactions in which atoms are too close together) and “contacts”
(both polar and nonpolar interactions) were found in atoms with a Van der Waal radius overlap of greater than or equal to 0.6 Ångstroms using the “Clashes & Contacts” tool. The “FindHBond” tool was used to identify potential hydrogen bonds between the amino acids of interest and their surrounding amino acids.

3 Results

3.1 *in silico* analysis of ADTs and PDTs

An *in silico* approach was used to identify amino acids that can potentially differentiate between the prephenate-accepting and the solely arogenate-accepting ADTs. To do this, both ADT and PDT protein sequences were analyzed and compared.

First, PDT protein sequences were selected from 26 different bacterial species (Table 1). These 26 bacteria were chosen so that they represent a diverse set of species. For example, sequences from both Gram-negative and Gram-positive bacteria were included. In contrast to *Arabidopsis* and other plant species that have ADT families, bacteria contain only one PDT. The goal of this alignment was to identify amino acids that were common to the PDTs. It was assumed that if an amino acid was important for PDT function it would be conserved among the bacterial PDTs in this alignment. However, there is a high level of sequence diversity between the bacterial sequences. For example, some sequences chosen for the alignment were only 24.6% identical, and there was an average of 45.8% identity within the entire group. Amino acids that were identical in 100% of the sequences were in highly conserved motifs that are known to be important for overall enzyme function (data not shown), and therefore unlikely to distinguish between an ADT and a PDT. Therefore, 100% identity was considered to be too stringent. Instead, amino acids that were similar in more than 75% of the sequences were identified. Forty amino acids were identified and are highlighted in blue in Figure 14.
Figure 14. Protein sequence alignment of bacterial PDTs.

Alignment of 27 PDT protein sequences from bacteria generated with DNAMAN. Dots: spacing introduced by the DNAMAN alignment program; blue highlight: 40 amino acid locations that were identified as similar in more than 75% of the bacterial sequences; yellow highlight: the 16 of 18 amino acid sites identified in the Arabidopsis ADT alignment that could be located; star: the 10 amino acid sites identified in the Arabidopsis ADT alignment that contained both the amino acid from ADT1/ADT2 and the amino acid from ADT3 – ADT6; arrow: the 6 amino acid sites identified in the Arabidopsis ADT alignment that did not contain the amino acid found in ADT3 – ADT6; green arrow: the amino acid position that corresponds to the Phe341Leu substitution in ADT5 (discussed in section 3.5).
Next, the six *Arabidopsis* ADT protein sequences were aligned. These sequences were aligned separately from the bacterial PDT sequences because the diverse PDT sequences disrupted the ADT alignment and made analysis difficult. The goal of the ADT alignment was to identify amino acid sites that may be responsible for the distinction between an ADT that can accept both arogenate and prephenate (ADT1 and ADT2) and a solely arogenate-accepting ADT (ADT3 – ADT6). To investigate this, an amino acid site was highlighted if it fulfilled three criteria:

1. There was a conserved amino acid between ADT1 and ADT2.
2. There was a conserved amino acid between ADT3-ADT6.
3. These two amino acids were different.

The 18 amino acid sites that met these requirements are highlighted in yellow in Figure 15.

Then, the results from the PDT alignment and the ADT alignment were compared. First, the 40 amino acids conserved in the bacterial PDTs were located in the ADT alignment. The goal of this comparison was to identify amino acids that were common between the bacterial PDTs and ADT1 and ADT2 but different from ADT3 – ADT6. However, all 40 of these amino acids that were conserved among the bacterial PDTs are present in both the ADT1/ADT2 and ADT3 – ADT6 groups. In most cases, these amino acids were present in all six of the *Arabidopsis* ADTs. These sites are highlighted in blue in Figure 15. Since they are found in both of the ADT groups, it is unlikely that these amino acids can distinguish between a prephenate-accepting and a solely arogenate-accepting ADT.

In an alternative approach, the 18 amino acid sites that were identified in the ADT alignment were located in the PDT alignment. Again, the goal of this comparison was to identify amino acid sites that were common between the bacterial PDT sequences and ADT1 and ADT2, but different from the amino acid in ADT3 – ADT6 at the same site. Due to the high level of diversity in the bacterial PDT sequences, only 16 of the 18 amino acid sites could be identified
Figure 15. Protein sequence alignment of Arabidopsis ADTs.

Partial alignment of the 6 ADT protein sequences from Arabidopsis thaliana generated with DNAMAN. The transit peptide sequences have been removed. Blue highlight: 40 amino acid locations that were identified in the bacterial PDTs as similar in more than 75% of the PDT sequences; yellow highlight: the 18 amino acid sites identified in the ADTs that differed between ADT1/ADT2 and ADT3 – ADT6; star: the 10 amino acid sites identified in the ADT alignment that contained both the amino acid from ADT1/ADT2 and the amino acid from ADT3 – ADT6 in the PDT sequences; question mark: the two amino acid sites identified in the ADTs that differed between ADT1/ADT2 and ADT3 – ADT6 that could not be found in the PDT sequences; arrow: the 6 amino acid sites identified in the ADT alignment that did not contain the amino acid found in ADT3 – ADT6 in the PDT sequences; green arrow: the amino acid position that corresponds to the Phe341Leu substitution in ADT5 (discussed in section 3.5).
with certainty. The two sites that could not be identified are labeled with a question mark in Figure 15 and the 16 amino acid sites that could be identified are highlighted in yellow in Figure 14. Of these 16 sites, 10 had the amino acid in ADT1/ADT2 as well as the amino acid found in ADT3 – ADT6 (stars in Figure 14). Therefore, these 10 amino acids were unlikely to be important for differentiating between a prephenate-accepting ADT and a solely arogenate-accepting ADT. The remaining 6 sites, however, did not contain the amino acid found in ADT3 – ADT6, which indicated that these amino acids might be important for ADT vs. PDT distinction (arrows in Figure 14).

As a third approach, an alignment of CDT protein sequences was analyzed. This was done because these enzymes can accept both arogenate and prephenate as substrates, and the comparison of their sequences with the ADTs and PDTs might reveal which amino acids are important for prephenate recognition. The CDT sequences were obtained from *Burkholderia pseudomallei* (YP_109985.1), *Acidovorax citrulli* (ABM33045.1), *Ilyobacter polytropus* (YP_003967695.1), *Bradyrhizobium diazoefficiens* (NP_771201.1), and *Pseudomonas aeruginosa* (CDH69854.1). However, the CDTs were found to be too distinct from both the plant ADT and bacterial PDT protein sequences to provide any valuable information (data not shown).

Therefore, the number of amino acid sites that might be important for distinguishing between a prephenate-accepting ADT and a solely arogenate-accepting ADT could not be narrowed down any further. Six is a large number of sites to target for testing protein function, particularly if a combination of several of these sites is required for prephenate recognition. Therefore, it was concluded that it was not practical to determine the amino acid(s) that are responsible for this distinction through specifically targeting these 6 amino acids.

### 3.2 Random mutagenesis

The results of the *in silico* analysis indicated that it was not practical to use targeted codon mutagenesis to determine which amino acids are required for
distinguishing between a prephenate-accepting ADT and a solely arogenate-accepting ADT. Therefore, it was decided to use another approach.

First, a random mutagenesis method was tried. One advantage of random mutagenesis is that it is an unbiased method that does not require any knowledge regarding putative sites of interest in the sequence to be mutated. A second advantage of this method is that it has the capacity to generate a large library of mutants that can be screened for changes in function caused by the mutation(s). The goal of this method was to randomly mutate an arogenate-specific ADT through error-prone PCR to introduce PDT function to the enzyme. Therefore, full-length ADT4 and ADT6 cDNA sequences (both solely arogenate-accepting; Cho et al. 2007) were selected for random mutagenesis. The ADT4 and ADT6 cDNA libraries were generated and transformed into pha2, the PDT-knockout yeast strain, to test for mutants that had acquired PDT function. However, it was difficult to generate a sufficient number of transformants. Only 2000 transformants were successfully generated. A subset of each library was sequenced, and approximately 15% of the sequences were identified as wild type. That means that of the 2000 transformants only approximately 1700 (85%) would be useful for the PDT screen. Many more transformants would need to be generated in order to exhaustively screen the libraries (Appendix A) and therefore it was concluded that the random mutation method would not be pursued any further.

3.3 Domain swapping

In a second approach it was decided to identify the domain (TP, catalytic, or ACT) that can distinguish between a prephenate-accepting ADT and a solely arogenate-accepting ADT. For this, a domain from a prephenate-accepting ADT was swapped with a domain from a solely arogenate-accepting ADT and the resulting “chimera” was tested for PDT function in a pha2 yeast complementation assay. If the chimera is able to complement the PDT knockout pha2 yeast strain, this indicates that the domain from the prephenate-accepting ADT contains sequences for prephenate recognition. ADT2 and ADT4 were initially chosen for this research. ADT2 was chosen for domain swapping because it displayed the
highest level of prephenate recognition of the six ADTs in biochemical analyses (240 M⁻¹s⁻¹; Cho et al. 2007). ADT4 was chosen because it had been previously studied by other members of the lab (White 2011; Karademir 2012), and has been identified as solely arogenate-accepting both *in vitro* and *in vivo* (Cho et al. 2007; Bross et al. 2011).

### 3.3.1 Generating ADT2/ADT4 chimeras

Overlap extension PCR was used to assemble the ADTs with swapped domains (Figure 9). In this method, two fragments are independently amplified from two different templates (Figure 9A). The 5' fragment is amplified with a 5' gene specific primer and a 3' primer that incorporates 15-24 bp of the other template. Conversely, the 3' fragment is amplified with a 3' gene specific primer and a 5' primer that incorporates 15-24 bp of the first template. Then, the two fragments are both used as templates in a third PCR reaction (Figure 9B) in which the overlapping sequences facilitate the annealing of the two fragments and can act as an extension starting point to generate the full-length chimera. Only the outer primers are used for this third PCR reaction.

A schematic for the six full-length ADT2/ADT4 chimeras that were generated through overlap extension PCR is outlined in Figure 10. These chimeras were designed to contain every possible combination of the three domains from ADT2 and ADT4. All initial PCR fragments generated are listed in Table 7 together with their expected length. PCR fragments were analyzed by gel electrophoresis to ensure that they were of appropriate length (Figure 16A).

The fragments were then used in combination as templates for the overlap extension PCR to generate the full-length chimeras. Expected sizes are given in Table 7 and the corresponding electrophoresis results are shown in Figure 16B.

After PCR amplification, the ADT chimera sequences were recombined into the donor vector. This recombination was facilitated by the att recombination sites that were integrated onto the 5' and 3' ends of the constructs by the outer Gateway® Recombination primers (Table 3; Figure 8). All inserts were sequenced and only inserts that contained the correct combination of domains and lacked PCR amplification errors were chosen. The chosen inserts were then
Figure 16. *ADT2/ADT4* chimeras.

(A) Agarose gels of *ADT2* and *ADT4* fragments containing the 15-24 bp of overlapping sequences. These fragments were used as templates in the overlap extension PCR to produce *Chimera 1 – Chimera 6*. (B) Agarose gel of *Chimera 1 – Chimera 6*.

L: 1 Kb ladder. All gels are 1% agarose and all sizes are in Kb.
recombined into the yeast compatible pAG423GAL-ccdB-ECFP destination vector.

3.3.2 Complementation assay of ADT2/ADT4 chimeras

To determine if Chimera 1 – Chimera 6 have PDT activity \textit{in vivo}, they were transformed into the \textit{pha2} yeast strain for the complementation assay. Full-length ADT2 was used as a positive control and full-length ADT4 was used as a negative control. In addition, yeast transformed with an empty destination vector, and untransformed yeast cells were used as negative controls.

Yeast cells were grown for 24 hours in appropriately supplemented liquid SD media containing raffinose. Raffinose media was chosen as it neither inhibits nor induces expression from the \textit{GAL1} promoter in the destination vector (Weinhandl \textit{et al.} 2014). The cells were spotted onto solid SD media plates supplemented as described previously (section 2.8). To ensure the reproducibility of the complementation results, two samples of yeast cells that were independently transformed with the same chimera were each spotted twice. Every complementation assay was repeated three times. The results for all replicates and repeats were identical. Therefore, a single set of representative data is shown.

Differential growth in the yeast samples was observed by day 13. Therefore, only data from this day is shown in Figure 17. By this day, the yeast transformed with ADT constructs were able to grow on media supplemented with Phe, regardless of the carbon source used. The \textit{pha2} strain was unable to grow on any media as it is untransformed and contains no vector. Yeast transformed with an empty destination vector was able to grow on all media as long as it was supplemented with Phe. All yeast samples containing an ADT construct were unable to grow on glucose or raffinose media if it lacked Phe. As demonstrated previously (Bross 2011), ADT2 was able to grow on inducing selective media lacking Phe whereas ADT4 did not. These controls indicate that all yeast samples have the ability to grow on all carbon sources if fully supplemented, that the \textit{GAL1} promoter is only induced on media containing
Figure 17. Complementation of the *pha2* phenotype for ADT2/ADT4 chimeras.

Growth of *pha2* yeast samples for the ADT2/ADT4 complementation assay on day 13. On the left, diagrams for the full-length chimeric sequences are shown. ADT2 domains are shown in white and ADT4 domains are shown in purple. The samples were spotted at a cell density of $1 \times 10^5$ cells/mL on selective solid SD media containing glucose, raffinose or galactose. For each of these carbon sources, the samples were spotted on media containing Phe and media lacking Phe.

*pha2*: un-transformed *pha2* yeast; pDEST: yeast transformed with the empty destination vector; Glu: Glucose; Raf: raffinose; Gal: galactose; +Phe: media containing Phe; -Phe: media lacking Phe.
galactose, and only an ADT with PDT function is able to grow on inducing medium lacking Phe.

Yeast expressing Chimera 1, Chimera 3 and Chimera 5 were unable to grow on galactose media lacking Phe. However, growth was observed for yeast transformed with Chimera 2, Chimera 4 and Chimera 6 (Figure 17). The ability of these chimeras to complement the PDT knockout *pha2* phenotype indicates that they can act as PDTs. The only ADT domain common to these three chimeras was the ACT regulatory domain from ADT2. Therefore, it was concluded that the ACT regulatory domain contains the sequences that can confer PDT function to an arogenate-only ADT *in vivo*.

### 3.4 Functional analysis of ADT2 and ADT5

To confirm that the ACT domain is able to confer PDT activity to an arogenate-only ADT, seven constructs with swapped domains were generated using ADT2 and ADT5 as templates. Again, ADT2 was chosen for its ability to complement the PDT knockout yeast phenotype (Bross et al. 2011). ADT5 was chosen, like ADT4, because it is a solely arogenate-accepting ADT (Cho et al. 2007; Bross et al. 2011).

A schematic for the seven full-length ADT2/ADT5 chimeras that were generated through overlap extension PCR is shown in Figure 11. Chimera 7 – Chimera 10 are analogs to Chimera 1 – Chimera 4, respectively, to test if the ADT2 ACT domain confers PDT activity to the ADT5 sequence. Three additional chimeras (Chimera 11 – Chimera 13) were created to narrow down whether it was the N-terminal half, the C-terminal half, or the catalytic:ACT domain border of the ACT domain that contains the sequences that can confer PDT to an arogenate-only ADT. The construct borders for Chimera 11 – Chimera 13 are shown in the ADT protein sequence alignment in Figure 18.

Primers were designed as described previously (Table 3; section 3.3.1), fragments and chimeras (Table 8) were amplified by PCR and overlap extension
Figure 18. Construct design for the N-terminal ACT domain.

Protein sequence alignment of partial catalytic and ACT domains of the six Arabidopsis ADTs generated with DNAMAN. Dark blue: C-terminal catalytic domain sequences; light green: N-terminal ACT domain sequences; dark green: C-terminal ACT domain sequences; black lines: the borders of Chimera 11 – Chimera 13. Amino acid sites that were identified as different for ADT1 and ADT2 vs. ADT3 – ADT6 are indicated in red. The SLEEG group was swapped between ADT2 and ADT5. All amino acid locations listed correspond to the location in the ADT5 protein sequence, with the exception of 4-V396L which corresponds to this location in the ADT4 sequence.
PCR and the resulting PCR fragments were checked for the proper sizes by gel electrophoresis (Figure 19A and 19B).

Chimera 7 – Chimera 13 were transformed into *pha2* and were tested for PDT activity using the complementation assay previously described (section 3.3.3). The same positive and negative controls were used with the exception that full-length ADT5 replaced ADT4 as a negative control (Figure 20). Differential growth was observed by day 11 and therefore images of yeast growth on this day are presented in Figure 20. Chimera 7, Chimera 9 and Chimera 11 were unable to support yeast growth on inducing medium lacking Phe. However, Chimera 8 and Chimera 10, Chimera 12, and Chimera 13 were able to support yeast growth on this media. The only region that was common between these chimeras was the N-terminal half of the ACT domain from ADT2.

Therefore, it was concluded that this region contains the sequences for the PDT activity of ADT2 and can confer PDT function to ADT5 *in vivo*. This area is defined by 38 amino acids that span from Phe286 to Leu323 in ADT2 (Figure 18).

### 3.5 *in silico* analysis of the N-terminal ACT domain

To identify the exact amino acids in the N-terminal ACT domain that can confer PDT function to an arogenate-only ADT, the protein sequence alignments were revisited. The purpose of this *in silico* analysis was to identify amino acid sites that were common between the bacterial PDT sequences and ADT1 and ADT2, but different from the amino acid in ADT3 – ADT6. Two sites fulfilled these criteria, the positions of which are indicated in Figure 18. At the first site, which corresponds to Phe341 in ADT5, ADT1 and ADT2 have a leucine (Leu) and ADT3 – ADT6 have a Phe. This position was substituted from a Phe to a Leu in the ADT5 sequence (5-F341L; Figure 12). At the second site, which corresponds to Ser345 in ADT5, ADT1 and ADT2 contain an asparagine (Asn) and ADT3 – ADT6 contain a serine (Ser). This position was substituted from a Ser to an Asn in ADT5 (5-S345N; Figure 12). A double
Figure 19. **ADT2/ADT5 chimeras.**

(A) Agarose gels of **ADT2** and **ADT5** fragments containing the 15-24 bp of overlapping sequences. These fragments were used as templates in the overlap extension PCR to produce **Chimera 7** – **Chimera 13**. (B) Agarose gel of **Chimera 7** – **Chimera 13**.

L: 1 Kb ladder. All gels are 1% agarose and all sizes are in Kb.
Figure 20. Complementation of the pha2 phenotype for ADT2/ADT5 chimeras.

Growth of pha2 yeast samples for the ADT2/ADT5 complementation assay on day 11. On the left, diagrams for the full-length chimeric sequences are shown. ADT2 domains are shown in white and ADT5 domains are shown in green. The samples were spotted at a cell density of $1 \times 10^5$ cells/mL on selective solid SD media containing glucose, raffinose or galactose. For each of these carbon sources, the samples were spotted on media containing Phe and media lacking Phe.

pha2: un-transformed pha2 yeast; pDEST: yeast transformed with the empty destination vector; Glu: Glucose; Raf: raffinose; Gal: galactose; +Phe: media containing Phe; -Phe: media lacking Phe
<table>
<thead>
<tr>
<th>Construct</th>
<th>TP</th>
<th>CAT</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADT5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chimera 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pha2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pDEST</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>+Phe</th>
<th>Glu</th>
<th>Raf</th>
<th>Gal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-Phe</th>
<th>Glu</th>
<th>Raf</th>
<th>Gal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
mutant was also generated that substituted both of these amino acids in ADT5 (5-DM; Figure 12).

In addition, a group of amino acids upstream of these two sites differs between ADT1/ADT2 and ADT3 – ADT6 (Figure 18). The amino acids corresponding to this group in the bacterial PDTs more closely resembled the group in ADT1/ADT2 than ADT3 – ADT6. This group will be referred to as the SLEEG group, due to the N-SLEEG-C motif observed in this group in ADT1 and ADT2. This region was swapped between ADT2 and ADT5 and two constructs were designed: 5-SLEEG and 2-SLEEG. 5-SLEEG contained ADT5 sequences from the N-terminus of the transit peptide to the last amino acid of the SLEEG region (S331) and ADT2 sequences from the corresponding ADT2 site (G299) to the C-terminus of the ACT domain. 2-SLEEG contained the reciprocal of this.

For comparison, an amino acid site that differed between ADT1/ADT2 and ADT3 – ADT6, but was located in the C-terminal ACT domain, was chosen as a negative control. This site corresponds to L263 in ADT2, V396 in ADT4, and V397 in ADT5 (section 3.1; Figure 18) and was substituted for the ADT2 amino acid in ADT4 and ADT5.

3.6 Functional analysis of sequences within the N-terminal ACT domain

A schematic for the seven full-length ADT chimeras that were generated through overlap extension PCR is outlined in Figure 12.

Primer design, fragment amplification and overlap extension PCR were performed as previously described (section 3.3.1) to generate the new constructs. For the generation of the point mutations, the primers contained nucleotide changes causing amino acid substitutions (Table 4). Primers for the amplification of 5-F341L contain a T>C substitution at nucleotide 1021, primers for 5-S345N contain G>A1034, primers for 5-DM contain both of these mutations, primers for 5-V397L contain G>T1189 and primers for 4-V396L contain G>C1186.
Fragments were amplified as previously described (section 3.3.1) and analyzed with gel electrophoresis to check for the proper sizes (Figure 21A; Table 9). The fragments were used as templates to generate the full-length chimeras through overlap extension PCR as previously described (section 3.3.1) with the exception of 5-SLEEG and 5-F341L. These chimeras could not be successfully amplified using the standard thermocycler program and thus were generated using the modified program described in Table 6. In this modified program, the primers were added to the reaction after 10 cycles at an annealing temperature of 55°C. This was done to ensure that amplification would only occur once the fragments had annealed. Then, the polymerase would extend from the overlapping sequences, generating the full-length chimeras. The full-length products are then used as a template in the remaining cycles after the outer primers are added. All chimeras were analyzed by gel electrophoresis to ensure that they were of appropriate length (Figure 21B; Table 9).

The chimeras were tested then for PDT function in a complementation assay as previously described (section 3.3.3), with the exception that both full-length ADT4 and ADT5 were used as negative controls.

Differential growth of the chimeras was observed by day 13 and therefore images of yeast growth on this date are presented in Figure 22. 5-SLEEG was unable to grow on fully supplemented SD media. Therefore, it was concluded that this chimera was unsuccessfully transformed into the pha2 yeast strain and it was not considered for further study. 2-SLEEG, 5-S345N, 5-V397L, and 4-V396L were unable to support growth on inducing medium lacking Phe. However, 5-F341L and 5-DM were both able to complement the PDT knockout phenotype on this medium. The only substitution that these two chimeras have in common is the Phe341Leu substitution, in which a Phe was substituted for a Leu at amino acid position 341 in ADT5. Therefore, it was concluded that the substitution of a Leu for Phe at amino acid position 341 in ADT5 can result in the introduction of PDT function to this arogenate-only ADT.
Figure 21. N-terminal ACT domain chimeras.

(A) Agarose gels of ACT domain fragments containing the 15-24 bp of overlapping sequences. These fragments were used as templates in the overlap extension PCR to produce 5-F341L, 5-S345N, 5-DM, 5-SLEEG, 2-SLEEG, 4-V396L and 5-V397L. (B) Agarose gel of 5-F341L, 5-S345N, 5-DM, 5-SLEEG, 2-SLEEG, 4-V396L and 5-V397L chimeras.
L: 1 Kb ladder. All gels are 1% agarose and all sizes are in Kb.
Figure 22. Complementation of the *pha2* phenotype for N-terminal ACT domain chimeras.

Growth of *pha2* yeast samples for the N-terminal ACT domain complementation assay on day 13. On the left, diagrams for the full-length chimeric sequences are shown. ADT2 domains are shown in white, ADT4 domains are shown in purple, and ADT5 domains are shown in green. The samples were spotted at a cell density of $1 \times 10^5$ cells/mL on selective solid SD media containing glucose, raffinose or galactose. For each of these carbon sources, the samples were spotted on media containing Phe and media lacking Phe.

*pha2*: un-transformed *pha2* yeast; pDEST: yeast transformed with the empty destination vector; Glu: Glucose; Raf: raffinose; Gal: galactose; +Phe: media containing Phe; -Phe: media lacking Phe.
3.7 Protein induction and expression in *pha2* *Saccharomyces cerevisiae*

Many of the ADT chimeras were unable to support growth of the *pha2* yeast strain. It was important to ensure that the chimeras were expressed under inducing conditions, otherwise the lack of yeast growth could be due to an inability of the yeast to express the chimeras instead of a lack of PDT activity.

Total soluble protein (TSP) was extracted from *pha2* yeast cells that contained a single ADT protein and grown in liquid inducing media and a western blot was performed as previously described (section 2.7; Figure 23). Equal protein concentrations were size separated on an SDS PAGE gel, transferred onto a nitrocellulose membrane and used for the western blot.

As all of the ADT proteins are ECFP fusion proteins, they were probed with an anti-GFP antibody. Anti-GFP was used because ECFP and GFP are similar enough that this antibody can detect both proteins. A GFP standard was also loaded as a positive control to ensure that the anti-GFP antibody could bind to GFP. Yeast transformed with full-length ADT2, ADT4 and ADT5 were analyzed as positive controls. Yeast transformed with Chimera 11 and grown in non-inducing conditions (“non-induced yeast”), yeast transformed with an empty destination vector, and untransformed yeast were analyzed as negative controls.

ECFP was seen in all of the positive controls, but it was not detected in the negative controls. A band of approximately 27 kDa was detected in the yeast transformed with an empty destination vector (“pDEST” in Figure 23), which corresponds to the 27.7 kDa size of the ECFP fusion protein, confirming that ECFP can be detected with the GFP antibody.

ECFP was detected for all of the ADT-ECFP fusion proteins with the exception of 5-SLEEG (Figure 23), which was considered to be untransformed. For all samples, the top band corresponds to the expected size of the chimera protein (Table 9) with the fusion protein attached. In addition, a band that is approximately 10 kDa smaller was seen in yeast samples that were transformed with Chimeras 2, 3, 7, 5-V397L, and full-length ADT5. Ten kDa corresponds to
Figure 23. Expression of ADT constructs in \textit{pha2} yeast.

Western blots of all ADT constructs. Total soluble protein was isolated and quantified using a BCA assay, 15 µg of total soluble protein was size separated on 10% SDS-PAGE and Western blots were exposed for 1 min. (A) Western blot containing ADT2/ADT4, ADT2/ADT5, and C-terminal ACT domain amino acid substitution constructs and positive and negative controls. The concentration of the GFP standard (lane 1) in this blot is 150 ng/µL. Due to the high signal observed for this standard this concentration was decreased in (B). (B) Western blot containing the N-terminal ADT2/ADT5 chimeras and ADT5 mutants, and full-length ADT2 and ADT5. The GFP standard concentration in this blot in 75 ng/µL. GFP: GFP standard; \textit{pha2}: empty \textit{pha2} yeast; pDEST: yeast transformed with the empty pAG423GAL-ccdB-ECFP; N: yeast transformed with Chimera 11 but grown under non-inducing conditions (glucose). All sizes are in kDa.
the approximate size of the transit peptide. This indicates that the transit peptide is cleaved in these samples, and the smaller band represents the cleavage product that lacks the transit peptide. ADT-ECFP fusion proteins without the transit peptide can still be detected by the GFP antibody because the ECFP tag is at the C-terminus of the protein. In addition, a small band corresponding to the size of the transit peptide was also detected in yeast transformed with Chimeras 3, 5, 6, 7, 8, 12, 5-V397L and full-length ADT5. This indicates that the ECFP is cleaved in these samples.

In conclusion, all of the ADT-ECFP fusion proteins (with the exception of 5-SLEEG) can be detected and hence are translated by the *pha2* yeast. Therefore, the lack of growth that was seen in many of these samples in the complementation assay is due to a lack of PDT activity and not an inability of the yeast to express the protein.

### 3.8 3D Homology modeling of the ACT domain

The results of the *pha2* complementation assays indicate that a Phe to Leu substitution at amino acid position 341 in ADT5 was able to confer PDT activity to this enzyme. It is possible that the change in substrate specificity observed in this altered ADT is due to a conformational change that is caused by the Phe341Leu substitution. There is no structural information available for any of the plant ADTs, which makes it difficult to confirm this prediction. However, there are computer programs that can predict a protein’s structure by modeling its sequence against that of a related protein with a known structure, using it as a structural template. Based on the template’s structural information, this “3D homology modeling” can predict what the overall conformation of the uncharacterized protein will look like. For this study, a program named UCSF Chimera was used to homology model the structure of ADT2 and ADT5. The goal of this modeling was to predict whether the Phe341Leu substitution in ADT5 has an impact on its surrounding amino acids or the ADT5 overall conformation, therefore altering the substrate specificity of this enzyme. The amino acid position in ADT2 that corresponds to this substitution is Leu317. It was expected
that the Phe341Leu substitution’s interactions with its surrounding residues would more closely resemble those of the Leu317 in ADT2 than the Phe341 in ADT5. In addition, it was expected that this substitution would alter the ADT5 overall structure to make it more similar to ADT2, a prephenate-accepting ADT.

First, the structure of ADT2 was predicted. To do this, a protein template with a known crystal structure had to be chosen. A BLAST search was performed to determine the most similar amino acid sequence for which there was a crystal structure available in the Protein Data Bank (PDB). The PDT protein from *C. tepidum* (*Ct*-PDT; QMXW) had the highest similarity to the ADT2 sequence. *Ct*-PDT, as described in section 1.6, is a homodimer. However, only a single chain or subunit can be modeled at a time with the UCSF Chimera program (Figure 24A). Therefore, the ADT2 sequence was modeled after the *Ct*-PDT monomer. The location of Leu317 was identified in the homology modeled ADT2 structure. The location of Leu317 in ADT2 was predicted to be in an α-helix in the ACT domain (Figure 24B). This correlates with the ACT dimer interface in *Ct*-PDT, and the site of Phe ligand binding. Then, the “FindHbond” tool was used to predict potential hydrogen bonds between Leu317 and its surrounding amino acids. This tool predicted that Leu317 forms a hydrogen bond with Ala313 (Figure 24B).

As with ADT2, a BLAST search was performed with the ADT5 sequence and it was found that *Ct*-PDT had the highest similarity to the ADT5 sequence. ADT5 was modeled against *Ct*-PDT and the position of Phe341 was identified (Figure 25A). Similar to Leu317, Phe341 was predicted to be in the α-helix that lies at the ACT domain interface and Phe ligand binding site in the *Ct*-PDT dimer. A potential hydrogen bond between Phe341 and Ser337 was predicted with the “FindHBond” tool (Figure 25B). Interestingly, Ser337 in ADT5 corresponds to the same amino acid position as Ala313 in ADT2. So, both Leu317 in ADT2 and Phe341 in ADT5 are predicted to form hydrogen bonds with the amino acids 4 positions towards their N-terminus. Then, the “Rotamers” tool was used to substitute Phe341 to a Leu in the 3D homology modeled ADT5. This tool can substitute an amino acid and will also predict the possible rotations in space in
Figure 24. Leu317 and the 3D homology model for ADT2.

(A) 3D homology model for an ADT2 monomer. This model is based on the structural information the monomer of the PDT protein from *Chlorobium tepidum* (Ct-PDT) using the UCSF Chimera program. (B) Zoomed in view of the ADT2 ACT domain to highlight the position of Leu317 (green). The predicted hydrogen bond between Leu317 and Ala313 (blue) is shown in orange.
Figure 25. Phe341 in the 3D homology modeled ADT5.

(A) ADT5 homology modeled monomer. The ADT5 sequence was homology modeled against the monomer of the PDT homodimer from *Chlorobium tepidium* (*Ct*-PDT) using UCSF Chimera’s “Modeller” program. (B) Zoomed view of Phe341 (green), the amino acid required for PDT activity in ADT2. The predicted hydrogen bond between Phe341 and Ser337 is shown in orange. (C) Zoomed view of Leu341 (green), which was substituted for Phe341. The predicted hydrogen bond between Leu341 and Ser337 is shown in orange.
which the new amino acid might exist. The tool assigns a probability to each of these different “rotamers”. Nine rotamers were identified for ADT5 and all of these rotamers were predicted to form a hydrogen bond with the same Ser337. The rotamer that was predicted with the greatest probability (75.3%) is shown in Figure 25C. The “Clashes and Contacts” tool was used to predict if any of the rotamers could physically interact with their surrounding amino acids, possibly resulting in a displacement those surrounding amino acids or an overall conformational change in the protein. However, no clashes or contacts were identified for any of the nine rotamers (data not shown).

In summary, the predictions made by the UCSF Chimera program do not indicate that the Phe341Leu substitution in ADT5 has any substantial effects on interactions with amino acids in its nearby surroundings or the overall protein conformation.

4 Discussion

The main goal of this study was to identify the amino acids that can distinguish between ADTs and PDTs. To accomplish this, a family of six Arabidopsis ADTs, two of which have PDT function (ADT1 and ADT2) and four of which are only ADTs (ADT3 – ADT6) were studied. Three members of this family, ADT2, ADT4 and ADT5 were chosen for analysis through domain swapping and yeast complementation assays. Nineteen ADT chimeras were generated through overlap extension PCR to consist of Arabidopsis sequences from ADT2, which is a prephenate-accepting ADT, and ADT4 or ADT5, which are solely arogenate-accepting ADTs. The chimeras were tested for PDT activity in a pha2 PDT knockout yeast strain. I was able to identify that a Phe to Leu substitution at amino acid position 341 in ADT5 can confer PDT function to this arogenate-only ADT.
4.1 Complementation results

*S. cerevisiae* is a useful organism in which to study protein function because it can be grown as a haploid and therefore the allele it carries determines its phenotype. If one allele is knocked out, the gene function is eliminated in the yeast strain entirely. A mutant *S. cerevisiae* yeast strain named *pha2* was used for the complementation assay in this study. *S. cerevisiae* has a single *PDT* coding sequence (*PHA2*), making it an ideal candidate for this assay. When *PHA2* is knocked out, the strain is unable to synthesize Phe and can only survive if either an exogenous supply of Phe is provided or a gene with PDT activity is transformed into and expressed in the yeast (Bross *et al.* 2011). Hence, this provides a simple way to identify PDT activity in a strain that contains an altered protein. Yeast growth indicates that the protein can complement the *pha2* phenotype and therefore has PDT activity. Lack of yeast growth indicates that the protein cannot complement the *pha2* phenotype, and therefore does not have PDT activity. This *pha2* complementation assay has already been successfully used to test the six *Arabidopsis* ADTs for PDT function (Bross *et al.* 2011). In this study, ADT1 and ADT2 were able to complement the *pha2* phenotype and ADT3 – ADT6 were not. Hence, ADT1 and ADT2 have PDT function and ADT3 – ADT6 do not.

ADT domain swapping chimeras were tested for PDT activity using the *pha2* complementation assay. The region that can confer PDT function to an arogenate-only ADT was systematically narrowed down to a single amino acid position. This amino acid was a Phe to Leu substitution at amino acid position 341 in ADT5, which conferred PDT activity to this enzyme. This position was previously identified in an *in silico* approach (section 3.1) as a position which could potentially distinguish between ADT and PDT function. It was first identified in the *Arabidopsis* ADTs (green arrow in Figure 15), because it fulfilled three requirements: that there was a conserved amino acid between ADT1 and ADT2 (Leu), that there was a conserved amino acid between ADT3 – ADT6 (Phe), and that these two amino acids are different.
Then, this position was located in the bacterial PDT sequences (green arrow in Figure 14). Importantly, Leu was identified at this position in the PDTs, while Phe was not. However, although Leu is present at this position in some of the bacterial sequences, it is not highly conserved. In fact, it is only found in only 3 of the 26 sequences in the alignment in Figure 14. In addition, several other amino acids such as Trp, Thr, Asn Arg, and Lys are found at this position. Interestingly, these amino acids have a wide range of different characteristics: Leu and Trp are nonpolar and hydrophobic, Thr and Asn are polar and hydrophilic, and Arg and Lys are positively charged. This indicates that Leu is not the only amino acid at this position which can confer PDT function.

To check if the presence of a Leu at position 341 in ADT5 can be used to predict PDT function in plant ADTs, the *Arabidopsis* ADT sequences were compared to ADT sequences from *P. hybrida*. The *P. hybrida* ADT sequences are the only other plant ADTs which have been biochemically analyzed (Maeda *et al.* 2010). There are three enzymes in the *P. hybrida* ADT family, *Ph-ADT1*, *Ph-ADT2* and *Ph-ADT3* (ACY79502.1, ACY79503.1, ACY79504.1, respectively) all of which function predominantly as ADTs. However, *Ph-ADT2* and *Ph-ADT3* can also convert prephenate to phenylpyruvate (act as PDTs). An alignment of the *P. hybrida* and the *Arabidopsis* ADT sequences is shown in Figure 26. Interestingly, neither of the prephenate-accepting *Ph-ADTs* have a Leu at the amino acid position corresponding to the Phe341Leu substitution that resulted in PDT function in ADT5. In fact, *Ph-ADT2* has a Met while *Ph-ADT3* has an Ala. This does not indicate that a Leu at this position can be used to predict PDT function in plant ADTs. However, *Ph-ADT1*, which is solely arogenate-accepting, has Phe at this location. This is the amino acid also found in the arogenate-specific *Arabidopsis* ADT3 – ADT6 at this position. This could indicate that a Phe at this position might define ADT function, rather than a Leu at this position defines PDT function. If this is the case, it is possible that many amino acids at this position can allow for the acceptance of prephenate as a substrate but only a Phe at this location eliminates prephenate recognition. Consistent with this assumption, none of the bacterial PDT sequences that were
Figure 26. Analysis of Leu317 from *Arabidopsis* and *Petunia* ADT sequences.

Alignment of partial catalytic and ACT domain sequences from the six *Arabidopsis* ADTs (At-ADT1 – At-ADT6) and the three Ph-ADT sequences from *Petunia hybrida* (Ph-ADT1 – Ph-ADT3) generated with DNAMAN. At-ADT1, At-ADT2, Ph-ADT2, and Ph-ADT3 are all predominantly arogenate-accepting but also have a low level of prephenate recognition. All other enzymes in this alignment are solely arogenate-accepting. Sequences from the N-terminal catalytic domain are highlighted in blue and sequences from the C-terminal ACT domain are highlighted in green. The position of Phe341Leu, the amino acid substitution that conferred PDT function to ADT5, is highlighted in red. The amino acid number at this position differs between the ADT enzymes because the lengths of the transit peptides in these enzymes are highly variable. At this position: the ADTs with PDT function have a Met (Ph-ADT2), an Asn (Ph-ADT3), or a Leu (At-ADT1 and At-ADT2) whereas all of the solely arogenate-accepting ADTs contain a Phe (Ph-ADT1 and At-ADT3 – At-ADT6).
<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ph-ADT2</td>
<td>NVLAQDIDQQDSDNVTFRMLAREPIIPGTDKFKTSVFSLD..EGPGVLFKALAVFAMRNINLTIESRPLQKQALR</td>
<td>337</td>
</tr>
<tr>
<td>Ph-ADT3</td>
<td>KVLSAFARNSLTKIESRPRHNRPRLVDDANVGTKHFEY..MFYVDFASMDVSIFRNALAEQTSFLRLVS</td>
<td>420</td>
</tr>
<tr>
<td>At-ADT1</td>
<td>DILAEIQDDVNNVTFLILAREPPRPTDRPYKTSIVFSL..EGPGVLFKALAVFALRSINLSKIESRPQRRPLR</td>
<td>336</td>
</tr>
<tr>
<td>At-ADT2</td>
<td>NIVAKDIQDDCNVTFRMLAREPIIPGTNRFLKTSIVFSL..EGPGVLFKALAVFALRQINLTIESRPLKHPLR</td>
<td>329</td>
</tr>
<tr>
<td>At-ADT3</td>
<td>EILEGQDDSNVTRFVMAREPIIPRTDRPFKTSIVFAH..EKGTCVFKVLSAFARNSLTIKESRPNHVPR</td>
<td>353</td>
</tr>
<tr>
<td>At-ADT4</td>
<td>QILADGIQDDPGVTRLMAREPIIPRTDRPFKTSIVFAQEHHGTSLFKVLSAFARDISLTKIESRPHMNPLR</td>
<td>359</td>
</tr>
<tr>
<td>At-ADT5</td>
<td>QILADGIQDDAGNVTRFLMLAREPPRPTDRPFKTSIVFAQEHHGTSLFKVLSAFARNSLTKIESRPHQCPR</td>
<td>360</td>
</tr>
<tr>
<td>At-ADT6</td>
<td>EILEGQDDSVTRFVMAREPIIPRTDRPFKTSIVFAH..EKGTSVLFKVLSAFARDISLTKIESRPHMNPR</td>
<td>348</td>
</tr>
<tr>
<td>Ph-ADT1</td>
<td>GLDILEQGIQDDLSNVTFRMLAREPIIPRTDRPFKTSIVFA..HDKGTSVLFKVLSAFARNSLTKIESRPHMNRP</td>
<td>359</td>
</tr>
</tbody>
</table>
analyzed carried a Phe at the corresponding position in the bacterial alignment (green arrow in Figure 14).

So, distinguishing between an ADT and a PDT based on this amino acid position alone might be more complicated than originally anticipated. A Phe at this position clearly allows for ADT function in ADT5, but future studies will need to be completed to determine if this is a more general phenomena.

However, the results from the complementation analysis in this study clearly show that a Phe to Leu substitution at amino acid position 341 can confer PDT function to ADT5, an arogenate-specific ADT. Although this may be surprising, there are many examples in the literature in which a single amino acid change has altered the function of an entire protein. For example, a single amino acid change in a cathepsin L protease that is secreted by the trematode Fasciola hepatica altered this enzyme’s substrate specificity (Smooker et al. 2000). The altered protein is able to cleave substrates with a proline at the cleavage site, whereas the native protein cannot. In another example, a single amino acid substitution in an aspartate aminotransferase from E. coli changed the pH dependence of the enzyme (Jeffery et al. 2000). This change alters the kinetic parameters of the enzyme, allowing the enzyme to recognize its aspartate substrate at a different pH. In both of these examples, the amino acid substitution responsible for the change in enzymatic activity is located outside of the catalytic site. Importantly, both substitutions result in a change in the overall conformation of the protein. This suggests that the overall protein conformation is more important for function than the protein sequence. One could argue that PDT enzymes themselves are an example of this hypothesis. For example, the PDT enzymes from C. tepidum and S. aureus display very similar overall conformation and function although their sequences are very distinct (only 27.3% identical (Tan et al. 2008). Therefore, a substitution that does not affect the catalytic site directly can substantially alter enzyme catalysis if it results in an overall conformational change that transfers the effects of the substitution to the catalytic site.
4.2 The importance of the ACT domain

The Phe341Leu substitution that conferred PDT function to ADT5 is located in the N-terminal region of the ACT domain. This was surprising since it is more intuitive that the amino acid required for prephenate recognition would be located in the catalytic domain, rather than the ACT domain. Characteristically, ACT domains are ligand-binding domains that can be found in a diverse group of metabolic enzymes that are regulated by allosteric binding of amino acids to this domain (Grant 2006; Liberles et al. 2005). The effects of ligand binding to the ACT domain are transferred to the catalytic site and can either inhibit or activate enzymatic activity (Grant 2006).

The enzymatic activity of the bacterial PDTs is allosterically inhibited by Phe binding to the ACT domain. The site of Phe binding in PDTs is in the N-terminal region of the ACT domain, and the amino acids that form the hydrophobic Phe-binding pocket are highly conserved among these enzymes (Tan et al. 2008). The two most highly conserved amino acid motifs in this region are the GALV and ESRP motifs, which are essential to Phe binding and feedback inhibition (Pohnert et al. 1999). These amino acids contribute directly to interactions with the Phe ligand as well as the formation of the closed ACT domain conformation. These amino acids correspond to G209SLF and E229SRP in Ct-PDT (Figure 27A). The GALV and ESRP motifs are also very highly conserved within the plant ADTs (Tan et al. 2008). These amino acids correspond to G297PGVL and E319SRP in Arabidopsis ADT2 (Figure 27B) and G328TSVL and E350SRP in ADT5 (Figure 27C) and are essential for Phe binding. In fact, a substitution of S320A in Arabidopsis ADT2 substantially decreased Phe-induced feedback inhibition of enzymatic function and increased the free Phe levels 160-fold (Huang et al. 2010). An S298I substitution in the ESRP motif in an Oryza sativa ADT also decreased Phe-induced inhibition of this enzyme (Yamada et al. 2008). This indicates that these motifs are required for Phe binding in plant ADTs, and that changes in this area of the ACT domain can indirectly affect the catalytic site in these enzymes as well. Although the Phe341Leu substitution in ADT5 is not
part of the GALV or ESRP motif it is located right in between them (Figure 27B). Therefore, it is understandable that this amino acid change could also indirectly affect substrate recognition.

Due to a lack of structural information, the mechanism of allosteric Phe inhibition in the plant ADTs is unknown. However, this mechanism has been determined in two bacterial PDTs (Tan et al. 2008). Crystal structures of the PDTs from *C. tepidum* and *S. aureus* indicate that Phe binding in the ACT domain results in an overall conformational change in these proteins, reducing the access of prephenate to the catalytic cleft in the catalytic domain.

These PDT enzymes form homodimers in which the two catalytic domains and the two ACT domains are closely aligned (Figure 4; Tan et al. 2008). An extended catalytic cleft, which hosts the active site, forms between the two catalytic domains. Interestingly, the N-termini of several α-helices from both monomers point into this cleft, making it highly positive (about 3.0 positive charges; (Hol 1985). This attracts prephenate, which is negatively charged, to the active site. When Phe is not bound to the ACT domain the dimer exists in an open conformation in which prephenate can access the catalytic cleft. However, when two Phe molecules simultaneously bind to two hydrophobic pockets at the ACT dimer interface, the ACT domains shift closer together. This movement is transferred throughout the entire protein, resulting in an overall change in conformation that splits the catalytic cleft and reduces the access of prephenate to the active site. Figure 5 shows a schematic of this conformational change. During this change, several of the positively charged α-helix N-termini that point into the cleft are rotated, decreasing the attractive positive charge near the active site. Therefore, it has been conclusively shown that amino acids in the region of the *Ct*-PDT and *Sa*-PDT ACT domain that correspond to the region of the Phe341Leu substitution in ADT5 can cause an overall change in protein conformation and indirectly affect the ability of prephenate to bind at the active site.
Figure 27. GALV and ESRP motifs in Ct-PDT, ADT2 and ADT5.
The amino acid site corresponding to Leu317 in ADT2 is in close proximity to highly conserved GALV and ESRP motifs in the Ct-PDT protein and the ADT2 and ADT5 homology models. (A) The GALV and ESRP motif in the Ct-PDT dimer, which correspond to G209SLF and E229SRP in each monomer, respectively. The motifs from monomer A (blue) are shown in light blue and the motifs from monomer B (red) are shown in pink. Leu220 from each monomer is indicated in green. (B) The GALV and ESRP motifs (P298GVL and E319SRP, respectively) and Leu317 in the ADT2 homology model. The motifs are shown in blue, and Leu317 is shown in green. (C) The GALV and ESRP motifs (G328TSVL and E350SRP, respectively) and Phe341, the amino acid that corresponds to Leu317 in ADT2, in the ADT5 homology model. The motifs are shown in pink, and Phe341 is shown in green.
Although there is no structural information available for the plant ADTs, the protein conformation of these enzymes can be predicted with a 3D homology modeling computer program such as UCSF Chimera. This program can model a protein sequence with unknown structure against a known protein structural template. Using the PDT from *C. tepidum* as a template, the overall conformation of *Arabidopsis* ADT2 and ADT5 was predicted in order to determine whether the Phe341Leu substitution in ADT5 has an impact on its surrounding amino acids or the ADT5 overall protein conformation (section 3.8). The predictions made by the UCSF Chimera program did not indicate that the Phe341Leu substitution in ADT5 has any substantial effects on interactions with its nearby amino acids or the overall protein conformation. However, since these interactions are predicted based on data from the PDT protein, they may not accurately reflect the actual ADT amino acid interactions or protein conformation *in vivo*. In addition, since the program can only model a single chain or subunit at one time, only interactions within an ADT monomer could be predicted.

Dimerization is required for both catalytic function and Phe feedback inhibition in the bacterial PDTs (Tan *et al.* 2008). Interestingly, it has been shown that all six of the *Arabidopsis* ADTs can form homo and heterodimers *in planta* (Styranko 2011). However, whether or not this dimerization is required for ADT catalytic function or Phe inhibition in the ADTs is currently unknown. Therefore, we cannot yet predict whether the Phe341Leu substitution in the ACT domain of ADT5 results in an overall change in protein dimer conformation similar to the change described for the bacterial PDTs (Tan *et al.* 2008).

4.3 Future directions

Using a *pha2* yeast complementation assay, it was identified that the Phe341Leu substitution in ADT5 introduced PDT function to this enzyme. To ensure that this substitution can confer PDT function to the other solely arogenate-accepting ADTs, future research should substitute this amino acid position in ADT3, ADT4 and ADT6 as well. As well, the Phe at this position in the other arogenate-only ADTs could be substituted to one of the other amino acids
found at this position in the bacterial PDTs (Trp, Thr, Asn Arg), to determine if this confers PDT function. In addition, since the \textit{in silico} analysis indicated it may be the lack of the Phe at this position that confers PDT function as opposed to a presence of Leu specifically, Leu317 should be substituted for Phe in ADT2. If a Phe at this position is in fact responsible for removing PDT function, it is expected that the Leu317Phe substituted ADT2 would lose PDT function and therefore not be able to complement the \textit{pha2} phenotype.

However, the \textit{pha2} complementation analysis can only determine if a protein has PDT activity, it cannot be used to test whether the altered ADT proteins have retained or lost their ADT activity. Therefore, it is important that all of the altered proteins are tested through biochemical analysis to determine whether these proteins can still recognize arogenate as a substrate.

The lack of structural information available for the plant ADTs greatly hinders our ability to understand substrate recognition and Phe inhibition in these proteins. Future research should concentrate on determining crystal structures for both the unaltered ADT proteins as well as those containing the ACT domain substitution that alter substrate recognition. This structural information will reveal whether the catalytic and allosteric inhibition mechanisms described for the bacterial PDTs (Tan \textit{et al.} 2008) are similar in the \textit{Arabidopsis} ADTs.

4.4 Conclusions

This study represents the first characterization of an amino acid that can potentially distinguish between ADT and PDT function. Through the generation of domain swapping ADT chimeras and a series of \textit{pha2} yeast complementation assays, I have identified that Phe341Leu, a single amino acid substitution in the ACT domain of ADT5, results in PDT function to this arogenate-only ADT. This study is an example of the ability of a single amino acid outside of the catalytic site to substantially impact enzymatic activity. These results also emphasize the importance of the ACT domain and the need to understand its role in indirectly affecting substrate recognition and protein function. As Phe is a building block of protein synthesis and a precursor to a wide variety of important secondary
metabolites in plants, understanding the enzymes involved in its biosynthesis pathway is important to regulating and manipulating the synthesis of this essential amino acid.
5 References


Bonawitz ND, Chapple C (2010) The genetics of lignin biosynthesis: Connecting genotype to phenotype. 44: 337-363


Cotton RGH, Gibson F (1965) The biosynthesis of phenylalanine and tyrosine; enzymes converting chorismic acid into prephenic acid and their relationships to prephenate dehydratase and prephenate dehydrogenase. BBA - General Subjects 100: 76-88


Fazel AM, Bowen JR, Jensen RA (1980) Arogenate (pretyrosine) is an obligatory intermediate of L-tyrosine biosynthesis: confirmation in a microbial mutant. Proc Natl Acad Sci U S A 77: 1270-1273


Mendonça LMF, Marana SR (2011) Single mutations outside the active site affect the substrate specificity in a β-glycosidase. Biochimica et Biophysica Acta - Proteins and Proteomics 1814: 1616-1623


6 Appendices

Appendix A.

The number of mutants that must be generated and screened in order to exhaustively search a randomly mutated library can be calculated with the following equation:

\[ \# \text{ mutants} = b^k \times \frac{n!}{(n-k)!} \]

“b” represents the number of different bases the wild type nucleotide could be changed to. “b” will always equal 3, because one of the four nucleotides (A,T,G,C) must be substituted to a different nucleotide to result in a mutation. “k” represents the number of nucleotide mutations that must be introduced in order to alter gene function and “n” represents the length of the sequence to be mutated (in nucleotides).

For example, in order to introduce 4 nucleotide mutations to a sequence that is 10 nucleotides in length through random mutagenesis:

\[ \# \text{ mutants} = 3^4 \times \frac{10!}{(10-4)!} \]
\[ \# \text{ mutants} = 408240 \]

408240 mutants would need to be generated and screened. For a full-length ADT gene of approximately 1200 bp, this number is essentially infinite. Therefore, it was deemed that random mutagenesis is not a practical method for determining the sequences required for PDT function in the prephenate-accepting Arabidopsis ADTs.
7 Curriculum Vitae

Research Experience

MSc Candidate & Researcher (Plant Genetics)  
UNIVERSITY OF WESTERN ONTARIO – London, ON  
09/12 – 08/14

- Has successfully applied analytical and creative problem solving to alter the substrate specificity of a family of enzymes required for phenylalanine biosynthesis in Arabidopsis thaliana.
- Maintains a strong technical understanding of molecular genetics in silico and in vivo laboratory practices.
- Supervises a team of four undergraduate student researchers and budgets laboratory expenses.

Research Assistant (Plant-Virus Interactions)  
AGRICULTURE & AGRI-FOOD CANADA (Dr. Aiming Wang) – London, ON  
09/11 – 04/12

- Supported senior researcher Dr. Aiming Wang with his research concerning the resistance-breaking mechanisms of the Soybean Mosaic Virus (SMV) to evaluate the genetic response to SMV in Ontario-grown soybean varieties.

Research Assistant (Fish Functional Genomics)  
UNIVERSITY OF WESTERN ONTARIO – London, ON  
09/10 – 04/11

- Worked with senior researcher Dr. Bryan Neff to characterize the genetics of mating tactics in bluegill sunfish. Identified presence of genes involved in courtship and kin recognition and wrote successful grant proposals to fund Dr. Neff’s research.

Research Assistant (Literature Review)  
CENTRE FOR RESEARCH ON INNER CITY HEALTH, ST. MICHAELS HOSPITAL – Toronto, ON  
05/09 – 08/09

- Worked with a team of university researchers lead by senior scientists Dr. Richard Glazier to apply a large scale data mining approach to update a series of systematic reviews regarding medical care for patients living with HIV/AIDS.
- Presented results to a review panel of inner city health research specialists. Reviews were published in the Cochrane Collaboration.

Volunteer Experience

Commercialization & Innovation Market Research Volunteer  
IVEY INTERNATIONAL CENTRE FOR HEALTH INNOVATION (ICHI) – London, ON  
12/13 – Present

- Part of an industry-academic collaboration to identify the success and challenges of implementing a new model of pharmaceutical care in a large-scale grocery chain industry partner in Ontario.
- Analyzes and presents data regarding the current quality of patient care and level of patient satisfaction, as well as the feasibility of expanding the program to stores across Canada.

Social Recreation Volunteer  
ALZHEIMER SOCIETY OF LONDON AND MIDDLESEX – London, ON  
09/08 – 04/14

- Works one-on-one with Alzheimer’s patients in recreation programs aimed at socializing patients and establishing an inclusive, supportive caregiver community.

Leadership Experience

Genetics Teaching Assistant  
UNIVERSITY OF WESTERN ONTARIO – London, ON  
09/12 – 04/14

- Lead tutorials of 30-90 students, writes assignments and marks assignments/exams for undergraduate genetics courses. Communicates course concepts and challenging students to apply them to solve molecular genetics problems.
- Was personally selected by Dr. Brenda Murphy, current laboratory director at Credit Valley Hospital, to TA her Advanced Genetics (Bio3595a) course as a result of previously receiving the highest mark in her fourth year Genes & Genomes (Bio4562b) course.
Team Leader
10/13 – 04/14
ALTERNATIVE SPRING BREAK/CAMP RESTORE – London, ON / New Orleans, LA
• Lead 36 students on an international service-learning trip to New Orleans, LA in February 2014.
• Selected applicants for participation and built a cohesive team with a diverse but complementary skill set. Communicated with our international partners to coordinate our volunteer efforts with the needs of the community.

Education and Training

Master of Science – Molecular Biology 2012 – 2014
UNIVERSITY OF WESTERN ONTARIO, Molecular Genetics Unit – London, ON
“Converting the substrate specificity of arogenate dehydratases (ADTs) in Arabidopsis thaliana.”
• Supervisor: Dr. Susanne Kohalmi, GPA: 4.0 (2012-2014)

Bachelor of Science – Honors Specialization in Genetics 2008 – 2012
UNIVERSITY OF WESTERN ONTARIO – London, ON
• Dean’s Honors List (2009-2012), GPA: 3.7 (2010-2012)
• Received the highest grade in Genes & Genomes (Bio4562b), an advanced fourth year genetics course.

Awards and Honors

• Offered: Western University’s Science Entrance Scholarship (in the event of transfer to the Doctoral program) (2014)
• Western University’s Biology Graduate Travel Award (2014)
• The Canadian Society of Plant Biologist’s George H. Duff Award (2014)
• Western University’s Global Opportunities Award (2014)

• Western University’s Western Science Graduate Entrance Scholarship (2012)
• The University of Toronto’s Scholarship for Dependents of Faculty Members (2008-2012)
• Western University’s Entrance Scholarship for High Academic Achievement (2008)

Research Presentations

Poster Presenter
07/14
CANADIAN & AMERICAN SOCIETY OF PLANT BIOLOGISTS ANNUAL MEETING – Portland, OR
• “Converting the substrate specificity of arogenate dehydratases (ADTs) from Arabidopsis thaliana.” Megan Smith-Uffen & Susanne E. Kohalmi (2014).

Oral Presenter
12/13
CANADIAN SOCIETY OF PLANT BIOLOGISTS EASTERN REGIONAL CONFERENCE – Mississauga, ON

Poster Presenter
06/13
CANADIAN SOCIETY OF PLANT BIOLOGISTS ANNUAL NATIONAL CONFERENCE – Quebec City, QC
• “Identifying the amino acids responsible for substrate specificity in arogenate dehydratases (ADTs) through the generation of gene chimeras.”