MONOLIGNOL AND LIGNAN BIOSYNTHETIC STUDIES:
FROM REACTION MECHANISMS TO
NEXT GENERATION SEQUENCING

By

CHOONSEOK LEE

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

DECEMBER 2012
To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of CHOONSEOK LEE find it satisfactory and recommend that it be accepted.

__________________________
Mechthild Tegeder, Ph.D., Chair

__________________________
Norman G. Lewis, Ph.D., Co-chair

__________________________
Gerald E. Edwards, Ph.D.

__________________________
John Browse, Ph.D.

__________________________
Andrew McCubbin, Ph.D.
I would like to heartily thank my advisor, Dr. Norman G. Lewis, for endless support during Ph.D. study at Washington State University and I also want to sincerely thank my committee members, Dr. Mechthild Tegeder, Dr. Gerald Edwards, Dr. John Browse and Dr. Andrew McCubbin for scientific advice. I want to give thanks to Dr. Laurence B. Davin for her help with scientific assistance and editing my thesis.

I want to show my gratitude to Diana Bedgar, Dr. Michael Costa, Dr. Joaquim V. Marques, Dr. Sung-Jin Kim, Dr. Syed Moinuddin and Dr. Kye-Won Kim for their help. I want to thank all Dr. Lewis Lab. members for their support. I also want to express my thanks to Dr. Min-Jeong Kim for his help to analyze transcriptome data. I want to thank Dr. ChulHee Kang and Mark Nissen, a technician in Dr. ChulHee Kang’s Lab, for use of the isothermal titration calorimetry machine.

I want to show sincere gratitude to my parents, Nanhyeong, a sister, and Yangseok, a brother. I also want to express my sincere gratitude to my lovely wife, Soohee, and my lovely children, Joshua, and Katherine.
Three objectives were studied: determination of the biochemical mechanisms of an *Arabidopsis thaliana* allylic double bond reductase (AtDBR1) producing stress/defense metabolites, dihydrocinnamyl aldehydes, and the *Arabidopsis* cinnamyl alcohol dehydrogenase (AtCAD5), as well as exploring the potential of utilizing next-generation sequencing to help identify unknown biochemical steps in formation of the lignan, podophyllotoxin, in *Podophyllum* species.

Apo and binary structures of AtCAD5 were solved at 2.0 and 2.6 Å resolution, respectively, and ternary complexes were modeled with *p*-coumaryl aldehyde. A putative proton shuttle mechanism for AtCAD5 involving Thr49, His52 and Asp57 was evaluated, based on a proposed comparable mechanism for horse liver alcohol dehydrogenase. Site-directed mutants of each were prepared with corresponding mutant proteins characterized by kinetic and isothermal titration calorimetry (ITC) analyses. It was established that Thr49 was important in overall
catalysis, whereas His52 and Asp57 were not. No evidence was obtained for a putative extended proton relay mechanism in AtCAD5.

Apo, binary and ternary complexes of AtDBR1 were obtained at 2.5 (apo) and 2.8 (binary and ternary) Å resolution, respectively. Analysis of the ternary structure indicated a concerted catalytic mechanism involving hydride transfer to C-7 of p-coumaryl aldehyde (C-3 in 4-HNE), with the Tyr260 hydrogen-bonded to the aldehydic group of p-coumaryl aldehyde and the 2′-OH of nicotine amide ribose. Site-directed mutation of the Tyr260 residue further confirmed an essential role in catalysis through kinetic and ITC analyses.

Illumina-based short reads/next-generation sequencing and bioinformatics analyses, together with a targeted metabolomics approach, of *Podophyllum hexandrum* and *P. peltatum*, were used to explore the potential of these technologies to deduce unknown biosynthetic steps to podophyllotoxin. Genes encoding steps in shikimate/chorismate, aromatic amino acid pathways, phenylpropanoid and lignan pathways were assembled, and putative enzymes catalyzing methylene-dioxy bridge formation were identified. Of these, recombinant proteins encoding by two genes CYP719A23 (*P. hexandrum*) and CYP719A24 (*P. peltatum*) were capable of converting (−)-matairesinol into (−)-pluviatolide.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Background and Significance</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Attributions</td>
<td>23</td>
</tr>
<tr>
<td>1.3. References</td>
<td>25</td>
</tr>
<tr>
<td>2. CRYSTAL STRUCTURES AND CATALYTIC MECHANISM OF THE</td>
<td>36</td>
</tr>
<tr>
<td><em>Arabidopsis</em> CINNAMYL ALCOHOL DEHYDROGENASES</td>
<td></td>
</tr>
<tr>
<td>AtCAD5 AND AtCAD4</td>
<td>36</td>
</tr>
<tr>
<td>Abstract</td>
<td>36</td>
</tr>
<tr>
<td>2.1. Introduction</td>
<td>37</td>
</tr>
<tr>
<td>2.2. Results and Discussion</td>
<td>40</td>
</tr>
<tr>
<td>2.3. Conclusions</td>
<td>57</td>
</tr>
<tr>
<td>2.4. Experimental</td>
<td>58</td>
</tr>
<tr>
<td>2.5. Acknowledgements</td>
<td>65</td>
</tr>
<tr>
<td>2.6. References</td>
<td>66</td>
</tr>
</tbody>
</table>
3. ASSESSMENT OF A PUTATIVE PROTON RELAY IN Arabidopsis

CINNAMYL ALCOHOL DEHYDROGENASE CATALYSIS .......................... 70

Abstract .................................................................................................................. 70

3.1. Introduction ........................................................................................................ 70

3.2. Results and Discussion .................................................................................. 76

3.3. Conclusions ...................................................................................................... 83

3.4. Experimental ................................................................................................... 83

3.5. Acknowledgements ........................................................................................ 87

3.6. References ...................................................................................................... 88

4. MECHANISTIC AND STRUCTURAL STUDIES OF APO-FORM,
   BINARY AND TERNARY COMPLEXES OF THE Arabidopsis

ALKENAL (DOUBLE BOND) REDUCTASE (At5g16970) .............................. 91

Abstract .................................................................................................................. 91

4.1. Introduction ........................................................................................................ 92

4.2. Experimental Procedures ............................................................................... 98

4.3. Results .............................................................................................................. 106

4.4. Discussion ........................................................................................................ 118

4.5. Concluding Remarks ...................................................................................... 121

4.6. Acknowledgements ......................................................................................... 124

4.7. References ...................................................................................................... 125

5. NEXT GENERATION SEQUENCING IN PREDICTING GENE FUNCTION IN
   PODOPHYLLOTOXIN BIOSYNTHESIS ......................................................... 131

Abstract .................................................................................................................. 131
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>132</td>
</tr>
<tr>
<td>5.2</td>
<td>Experimental procedures</td>
<td>136</td>
</tr>
<tr>
<td>5.3</td>
<td>Results and Discussion</td>
<td>146</td>
</tr>
<tr>
<td>5.4</td>
<td>Conclusion</td>
<td>162</td>
</tr>
<tr>
<td>5.5</td>
<td>Acknowledgements</td>
<td>162</td>
</tr>
<tr>
<td>5.6</td>
<td>References</td>
<td>163</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.</td>
<td>Crystallographic data for AtCAD5</td>
<td>63</td>
</tr>
<tr>
<td>3.1.</td>
<td>Sequences of primers used for site-directed mutagenesis</td>
<td>76</td>
</tr>
<tr>
<td>3.2.</td>
<td>Kinetic parameters for AtCAD5 and its site-directed mutants with ( p )-coumaryl aldehyde as substrate in presence of NADPH</td>
<td>77</td>
</tr>
<tr>
<td>3.3.</td>
<td>Thermodynamic parameters for NADPH binding to AtCAD5 and its mutants</td>
<td>79</td>
</tr>
<tr>
<td>4.1.</td>
<td>Comparison of amino acid similarity and identity for various alkenal/alkenone reductases</td>
<td>97</td>
</tr>
<tr>
<td>4.2.</td>
<td>Crystallographic data for the AtDBR1 apo- and binary and ternary complex forms</td>
<td>106</td>
</tr>
<tr>
<td>4.3.</td>
<td>Kinetic parameters for AtDBR1</td>
<td>109</td>
</tr>
<tr>
<td>5.1.</td>
<td>Characteristic ions from detected lignans in positive mode mass spectral analyses</td>
<td>147</td>
</tr>
<tr>
<td>5.2.</td>
<td>Blast results for known protein sequences from ( P. ) ( peltatum ) and ( P. ) ( hexandrum )</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Supplementary Table 5.1.     Shikimate, phenylpropanoid, and downstream lignan biosynthetic gene homologs search results for ( P. ) ( peltatum ) and ( P. ) ( hexandrum )</td>
<td>169</td>
</tr>
<tr>
<td></td>
<td>Supplementary Table 5.2.     CyP450 genes selected for cloning</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>Supplementary Table 5.3.     (^{1}H ) and (^{13}C ) NMR spectroscopic assignments for enzymatically produced (-)-Pluviatolide</td>
<td>173</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Podophyllotoxin</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Phenylpropanoid pathway in vascular plants</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>Various derivatives of phenylpropanoid/phenylpropanoid-acetate pathways</td>
<td>3</td>
</tr>
<tr>
<td>1.4</td>
<td>H, G and S monomers of lignin</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Various dihydrophenylpropanoids</td>
<td>8</td>
</tr>
<tr>
<td>1.6</td>
<td>Various lignans</td>
<td>8</td>
</tr>
<tr>
<td>1.7</td>
<td>Various allyl/propenyl phenols</td>
<td>8</td>
</tr>
<tr>
<td>1.8</td>
<td>Podophyllotoxin and its semi-synthetic derivatives</td>
<td>9</td>
</tr>
<tr>
<td>1.9</td>
<td>Enzymatic reactions catalyzed by AtCAD5</td>
<td>12</td>
</tr>
<tr>
<td>1.10</td>
<td>Enzymatic reactions catalyzed by AtDBR1 and/or PPDBR</td>
<td>15</td>
</tr>
<tr>
<td>1.11</td>
<td>Proposed biosynthetic pathway to podophyllotoxin in <em>Podophyllum petatum</em></td>
<td>16</td>
</tr>
<tr>
<td>1.12</td>
<td>(-)-Pinoresinol-specific reactions catalyzed by (-)-PLR in <em>T. plicata</em></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>and <em>L. usitatissimum</em></td>
<td></td>
</tr>
<tr>
<td>1.13</td>
<td>Enzymatic reactions catalyzed by CYP81Q1 in <em>Sesamum indicum</em></td>
<td>18</td>
</tr>
<tr>
<td>1.14</td>
<td>Various alkaloids</td>
<td>19</td>
</tr>
<tr>
<td>1.15</td>
<td>Conversion of (-)-matairesinol into (-)-pluviatolide by CYP719A23 and</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>CYP719A23</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Cinnamyl alcohol dehydrogenase substrates and products</td>
<td>38</td>
</tr>
<tr>
<td>2.2</td>
<td>Crystal structure of AtCAD5 homodimer and energy minimized model</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>of AtCAD4</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Molecular mass determination of AtCAD5 and AtCAD4</td>
<td>43</td>
</tr>
</tbody>
</table>
2.4. Amino acid sequence comparisons of plant CADs................................................................. 45
2.5. Superimposed views of AtCAD5 in its apo- and binary complex forms and
structure of the substrate-binding pocket of the NADP⁺ binary form of AtCAD5........ 48
2.6. Difference Fourier maps for NADP⁺ binary complexes of AtCAD5......................... 51
2.7. Proposed proton shuttle mechanism during the reduction process in the
active site of the AtCAD5........................................................................................................ 54
3.1. Enzyme reactions catalyzed by AtCAD5 in presence of NADPH ......................... 71
3.2. Two proposed catalytic mechanisms for alcohol dehydrogenases ..................... 73
3.3. Putative proton shuttle mechanism for AtCAD5......................................................... 75
3.4. Isothermal titration calorimetry analyses of AtCAD5, AtCAD5 T49A, AtCAD5
H52A and AtCAD5 D57A with NADPH............................................................................. 80
3.5. Isothermal titration calorimetry analysis of AtCAD5.................................................. 82
4.1. Various dihydrophenylpropanoids and phlorizin................................................... 94
4.2. Reactions catalyzed by NADPH-dependent alkenal/alkenone/α,β double
bond reductase in vitro......................................................................................................... 95
4.3. Amino acid sequence alignment of AtDBR1 with other oxidoreductases .......... 97
4.4. Phenylpropanoids and dihydrophenylpropanoids.................................................. 108
4.5. Ternary complex of AtDBR1 homodimer with bound NADP⁺/p-coumaryl
aldehyde and NADP⁺/4-HNE .......................................................................................... 111
4.6. Molecular mass of AtDBR1 in solution.................................................................... 112
4.7. Stereo-view of the substrate-binding pocket of AtDBR1..................................... 116
4.8. Observed potential interactions in the ternary complex of AtDBR1 with NADP+/p-coumaryl aldehyde, NADP+/4-HNE, and their corresponding schematic reaction mechanisms.................................................................................................................. 122

5.1. (-)-Podophyllotoxin and its derivatives teniposide, etopophos and etoposide, used in cancer treatment......................................................................................................................................... 134

5.2. Possible biosynthetic pathway and/or grid leading to (-)-podophyllotoxin and related lignans ...................................................................................................................................................... 135

5.3 Lignans tested as putative substrates in assays for CYP719A23 and CYP719A24 methylenedioxy bridge formation and (-)-haplomyrfolin ..................................................................................................................... 138

5.4. Relative lignan contents in different tissues of *Podophyllum hexandrum* and *P. peltatum* .................................................................................................................................................. 147

5.5. Sequence alignment of methylenedioxy bridge-forming cytochrome P450s .............. 155

5.6. Ultra-performance liquid chromatography-mass spectrometry analysis of enzymatic assays .......................................................................................................................................................... 157

5.7. Fragmentation pattern of (-)-pluviatolide and (-)-haplomyrfolin ......................... 158

5.8. Kinetic parameters for CYP719A23 and CYP719A24 ................................................ 160

5.9. Phylogenetic analysis of cloned and known cytochrome P450 enzymes illustrating the similarity of the cloned genes to known alkaloid biosynthetic enzymes ...................... 161

Supplementary Figure 5.1. (-)-Pluviatolide (CDCl3) 600MHz-1H-NMR.................. 174

Supplementary Figure 5.2. (-)-Pluviatolide (CDCl3) 600MHz-13C-NMR .................. 175

Supplementary Figure 5.3. (-)-Pluviatolide (CDCl3) 600MHz-1H-13C-NMR ............. 176
CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE

The study of various enzymatic steps in the biosynthesis of monolignols, dihydrophenylpropanoids, and the medicinally important lignan, podophyllotoxin (1, Figure 1.1), is the focus of this dissertation, these being products of the phenylpropanoid pathway (Figure 1.2).1-12 This pathway and that of phenylpropanoid-acetate metabolism, which lead to (so-called) secondary metabolites, including lignins, lignans, dihydrophenylpropanoids7-9, allyl/propropyl phenols, flavonoids, as well as the aromatic components of cutin and suberins (Figure 1.3).1,2,4,5

Figure 1.1. Podophyllotoxin (1)
Figure 1.2. Phenylpropanoid pathway in vascular plants: Phenylalanine ammonia-lyase (PAL), tyrosine ammonia-lyase (TAL; predicted to exist in grasses as a bifunctional PAL/TAL), cinnamate-4-hydroxylase (C4H) and its related reductase, 4-coumarate:CoA ligase (4CL), cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), p-coumarate 3-hydroxylase (pC3H) and its related reductase, caffeoyl-CoA O-methyltransferase (CCOMT), caffeate O-methyltransferase (COMT), ferulate 5-hydroxylase (F5H) and its associated reductase, hydroxycinnamoyl-CoA shikimate/quinate hydroxyphenyl transferase (HQT), and hydroxycinnamoyl quinate/hydroxycinnamoyl-CoA transferase (HQT).
Figure 1.3. Various derivatives of phenylpropanoid/phenylpropanoid-acetate pathways. Adapted, modified and redrawn from Vogt\textsuperscript{13}.

Of these, the lignins are the second most abundant substances in vascular plants, next to cellulose.\textsuperscript{1} The lignins are major constituents of secondary cell walls in the vasculature, thereby helping provide mechanical reinforcement and thus enabling vascular plants to be able to grow tall and to conduct water/nutrients from the roots to aerial tissues without significant loss because of the hydrophobicity of lignin.\textsuperscript{1} Cutin and suberins, whose aromatic constituents are also derivatives of the phenylpropanoid pathway, help to provide a protective coat thereby preventing loss of water from the surfaces of plants.\textsuperscript{13} In addition, many components of the phenylpropanoid/phenylpropanoid-acetate pathways are involved in plant defense via several
genes being significantly induced by bacterial or fungal infection. For example, phenylalanine ammonia-lyase (PAL), cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD), in Linum usitatissimum, were also induced by either Fusarium culmorum or F. oxysporum infection: The highest expression of these three genes occurred circa 48 h after infection of F. culmorum or F. oxysporum; of these genes, PAL levels increased ~17.2 fold at 48 h after infection of F. oxysporum, as compared to the control.

Phenylpropanoid/phenylpropanoid-acetate metabolism starts mainly from deamination of phenylalanine (2) or, to a lesser extent tyrosine (4) (e.g. in monocots) by the action of phenylalanine (2) ammonia-lyase (PAL) or tyrosine (4) ammonia-lyase (TAL). The PAL activity for conversion of phenylalanine (2) into cinnamic acid (3) was first reported in Hordeum vulgare, and later 4 PAL homologues, annotated after the release of Arabidopsis thaliana genome sequence, were characterized: AtPAL1, 2 and 4 exhibited similar affinity with phenylalanine (2), \( K_m \) values from 64 – 71 µM, except for AtPAL3, and had phenylalanine (2) specificity in contrast to low activities with tyrosine (4). Further analyses of site-directed mutants, including F114H in Arabidopsis and H89F in Rhodobacter sphaeroides, exhibited a change in substrate specificity, i.e. from phenylalanine (2) → tyrosine (4) in F114H in Arabidopsis, and tyrosine (4) → phenylalanine (2) in H89F in R. sphaeroides.

Cinnamic acid (3), generated by PAL activity, is converted into \( p \)-coumaric acid (5) by C4-hydroxylation by cinnamate-4-hydroxylase (C4H). C4H activity was first reported in Pisum sativum. The gene encoding C4H was later established to be a CYP73, a member of a family of cytochrome P450s (CYP450s). The genes encoding C4H were reported from Helianthus tuberosus, Vigna radiata and Medicago sativa, respectively. As reviewed in
Davin *et al.*⁴, C4H is able to give entry to various metabolites derived from the phenylpropanoid/phenylpropanoid-acetate pathways.⁴

The other two CYP450s, including *p*-coumarate-3-hydroxylase (*p*C3H) for C3-hydroxylation³¹,³² and ferulate-5-hydroxylase (F5H) for C5-hydroxylation³³,³⁴, are also involved in the phenylpropanoid pathway,³¹-³⁴ and especially, as reviewed in Davin *et al.*⁴, *p*C3H and F5H are importantly involved in formation of coniferyl (22) and sinapyl (24) alcohols. *p*C3H activity, which converts *p*-coumaroyl shikimate (27) into caffeoyl shikimate (28), but not *p*-coumaric acid (5) into caffeic acid (6), was discovered using assays with a microsomal preparation from parsley cell suspension cultures.³¹ The gene encoding *p*C3H was later reported as *CYP98A3*, a CYP450 gene, from *Arabidopsis thaliana* and its recombinant protein, CYP98A3, where expressed in yeast, was able to convert *p*-coumaroyl shikimate (27) into caffeoyl shikimate (28), and *p*-coumaroyl quinate (25) into caffeoyl quinate (chlorogenic acid, 26).³² Later, two CYP98As, including CYP98A35 and CYP98A36, were cloned and characterized from green beans of *Coffea canephora.*³⁵ These two *p*C3Hs of *C. canephora* exhibited much lower catalytic efficiency than *A. thaliana* CYP98A3, but showed their own distinct catalytic activities, i.e. CYP98A35 converts *p*-coumaroyl shikimate (27) and *p*-coumaroyl quinate (25) into caffeoyl shikimate (28) and caffeoyl quinate (chlorogenic acid, 26), whereas CYP98A36 catalyzes only *p*-coumaroyl shikimate (27) into caffeoyl shikimate (28). These results supposed the roles of CYP98As of *C. canephora* to be involved in formation of chlorogenic acid (26), as well as in formation of monolignols, including coniferyl (22) and sinapyl (24) alcohols.³⁵

The CYP450 F5H was also discovered in assays with a microsomal fraction prepared from young stems of *Populus × euramericana,*³³ with a gene encoding a F5H later reported as a *CYP84* in *A. thaliana.*³⁴ Ferulic acid (7) was first considered as a substrate of F5H,³³ but it was
later reported that, as substrates, F5H preferentially utilized coniferyl alcohol (22) and coniferyl aldehyde (17) than ferulic acid (7). Patten et al. exhibited that A. thaliana fah 1-2, a knock-out mutant of F5H, had no detectable S unit (38, Figure 1.4), but significant increase of G unit (39) than the wild type, whereas the F5H over-expression mutant of A. thaliana showed significant increase of S unit (38), but severe decrease of G unit (39), indicating the involvement of F5H in formation of sinapyl alcohol (24).

\[ \text{Figure 1.4. } \text{H, G and S monomers of lignin} \]

\[ \text{p-Coumaric acid (5) is converted into p-coumaroyl CoA (10) by the action of 4-coumarate CoA ligase (4CL). Gross and Zenk provided the first indirect evidence for CoA esters being involved as intermediates in the phenylpropanoid/phenylpropanoid-acetate pathways, the 4CL activity was further confirmed in Petroselinum hortense, Salix alba and Forsythia sp. The 4CL gene was cloned using parsley, but this was incorrect. The first true 4CL gene was cloned from Glycine max by Lindermayr et al. Following release of the Arabidopsis genome sequence, 14 putative At4CLs were annotated, but only 4 of these, namely At4CL1, At4CL2, At4CL3 and At4CL5, were biochemically active as such. Interestingly, At4CL1, At4CL2 and At4CL3 can utilize cinnamic (3), p-coumaric (5), caffeic (6), ferulic (7) } \]
and 5-hydroxyferulic (8) acids as substrates, whereas At4CL5 can employ p-coumaric (5), caffeic (6), ferulic (7), 5-hydroxyferulic (8) and sinapic (9) acids as substrates.\textsuperscript{40} At4CL5 is the only enzyme which can use sinapic acid (9) as a substrate.\textsuperscript{40} For example, \textit{At4CL1} was strongly expressed in both lignifying – leaf and root veins, trichomes and hydathodes – and lignified – the reproductive organ – tissues; At4CL5 was also expressed in lignifying tissues, as in \textit{At4CL1}.\textsuperscript{4}

\textit{p}-Coumaroyl CoA (10), generated from \textit{p}-coumaric acid (5) by 4CL, is a pathway intermediate to a very diverse group of phytochemicals, including flavonoids, isoflavonoids, stilbenes, coumarins, lignins, lignans, dihydrophenylpropanoids and allyl/propanenyl phenols.\textsuperscript{1,4,13} Hydroxycinnamyl aldehydes, including \textit{p}-coumaryl (15), coniferyl (17) and sinapyl (19) aldehydes, are subsequently formed by reduction of hydroxycinnamoyl CoA esters. This can then be converted into monolignols by the action of cinnamyl alcohol dehydrogenase (CAD)\textsuperscript{47} or into dihydrophenylpropanoids (Figure 1.5) by allylic double bond reductase catalysis.\textsuperscript{8,9} Coniferyl alcohol (22), a monolignol, can be utilized to afford lignans (Figure 1.6), such as podophyllotoxin (1)\textsuperscript{11,12}, secoisolariciresinol diglucoside (41, SDG)\textsuperscript{48} and (+)-sesamin (42)\textsuperscript{49}, and allyl/propanenyl phenols (Figure 1.7), including eugenol (43)\textsuperscript{50-52}. Interestingly, some of these derivatives of the phenylpropanoid pathway show medicinal properties, including anti-oxidant, cytoprotective, anti-cancer and cardioprotection.\textsuperscript{10,53-59} For example, dihydroconiferyl alcohol (44), dihydrophenylpropanoids, first found in nematode (\textit{Bursaphelenchus xylophilus})-infected \textit{Pinus densiflora} and \textit{P. thunbergii}\textsuperscript{60}, showed anti-oxidant and cytoprotective activities.\textsuperscript{10,61} Podophyllotoxin (1), an aryltetralin lignan extracted from \textit{Podophyllum hexandrum} and \textit{P. peltatum}, is utilized to make its semi-synthetic derivative forms (Figure 1.8), including etoposide (45), teniposide (46) and Etopophos\textsuperscript{®} (47), widely used for cancer treatment\textsuperscript{62-65}. The biochemical pathway to podophyllotoxin (1) is a focus of this dissertation.
Various dihydrophenylpropanoids

- Podophyllotoxin (I)
- Secoisolariciresinol diglucoside (SDG, 41)
- (-)-Trachelogenin (51)
- Plicatic acid (52)
- (+)-Piperitol (53)
- (+)-Sesamin (42)

Various lignans

- R=H, R₁=H, R₂=H, Anethole (57)
- R=CH₃, R₁=H, R₂=H, Methylchavicol (60)

Various allyl/prophenyl phenols.

R₁ = CHO, R₂ = R₃ = H, p-Dihydrocumarin aldehyde (48)
R₁ = CHO, R₂ = OH, R₃ = OCH₃, R₄ = H, Dihydroconiferyl aldehyde (49)
R₁ = CH₂OH, R₂ = R₃ = H, p-Dihydrocumarin alcohol (50)
R₁ = CH₂OH, R₂ = H, R₃ = OCH₃, Dihydroconiferyl alcohol (44)
Figure 1.8. Podophyllotoxin (1) and its semi-synthetic derivatives.

$p$-Coumaroyl CoA (10) is also converted into caffeoyl CoA (11), a pathway intermediate to the monolignols, coniferyl (22) and sinapyl (24) alcohols, via sequential catalysis by hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT) and $p$C3H.$^{4-6}$ That is, via addition of shikimate (30) to $p$-coumaroyl CoA (10), and then hydroxylation by $p$C3H and subsequent removal of the shikimate (30) moiety from caffeoyl shikimate (28) by HCT.$^{4-6}$ A gene encoding HCT was cloned from \textit{Nicotiana tabacum} and its recombinant HCT was also characterized.$^{66}$ Knock-down of \textit{HCT} by RNAi in \textit{N. benthamiana}$^{67}$ and \textit{N. tabacum}$^{6}$ caused severe decrease of the contents of G (39) and S (38) units and significant increase of the contents of H unit (40), this being indicated that HCT is involved in production of coniferyl (22) and sinapyl (24) alcohols, methoxylated monolignols. As in the enzymatic reactions by HCT$^{66}$, hydroxycinnamoyl CoA:quinate hydroxycinnamoyl transferase (HQT), cloned and characterized from \textit{N. tabacum} and \textit{Lycopersicon esculentum}, utilizes $p$-Coumaroyl CoA (10) as a substrate.
However, Cardenas-Ardila reported through knock-down of HQT by RNAi in *N. tabacum* that HQT was involved in chlorogenic acid (26) formation, but not in monolignols formation.  

Caffeoyl CoA (11) is converted into feruloyl CoA (12) by caffeoyl CoA (11) O-methyltransferase (CCOMT).  

Pakusch et al. first detected the CCOMT activity from parsley (*Petroselinum crispum*)

5 putative AtCCOMTs were also annotated in The Arabidopsis Information Resource (TAIR). Of these, AtCCOMT1 and AtCCOMT2 were characterized (Takahashi et al., unpublished): AtCCOMT1 was preferentially able to catalyze caffeoyl (11) and 5-hydroxyferuloyl (13) CoA; AtCCOMT2 utilized quercetin (33) as more efficient substrate; AtCCOMT1 and AtCCOMT2 did not, however, methylate caffeic (6) and 5-hydroxyferulic (8) acids (Takahashi et al., unpublished). A CCOMT-deficient alfalfa (*Medicago sativa*) putatively exhibited a similar lignin structure to the control, but ~20% decrease in lignin content than the control.

Finkle and Nelson reported the in vitro conversion of caffeic acid (6) into ferulic acid (7), i.e. the activity of caffeic acid (6) O-methyltransferase (COMT), using cell-free extracts from apple cambial tissue, and then, Hess reported bifunctional COMT activity, which converted both caffeic (6) and 5-hydroxyferulic (8) acids into ferulic (7) and sinapic (9) acids, from *Rosa* sp., *Petunia hybrid*, *Avena sativa*, *Streptocarpus hybrida* and *Gaillardia picta sanguinea*, respectively. These supposed bifunctional COMTs were also cloned and characterized from *M. sativa* and *Populus tremuloides*. Moreover, of the 17 putative AtCOMTs annotated in the TAIR, only AtCOMT1 could methylate caffeic (6)/5-hydroxyferulic (8) acids, caffeyl (16)/5-hydroxyconiferyl (18) aldehydes and caffeyl (21)/5-hydroxyconiferyl (23) alcohols. Of these substrates, caffeyl (16) and 5-hydroxyconiferyl (18) aldehydes were preferentially catalyzed by
AtCOMT1. In addition, further analyses of G (39) and S (38) thioacidolysis monomers in an Atcomt1 mutant established that levels of the S-derived monomer (38) were undetectable and the G-derived monomer (39) also decreased ~20 % than that of wild type. This indicated that AtCOMT1 was involved in S (38)- and not G (39)-derived monomer formation, thereby differing from the results of the in vitro assays, either. In addition, in vitro assays of AtCOMT1 indicated preferential use of caffeoyl (16) and 5-hydroxyconiferyl (18) aldehydes as putative substrates, relative to caffeic (6)/5-hydroxyferulic (8) acids and caffeoyl (21)/5-hydroxyconiferyl (23) alcohols.

\[
p\text{-Coumaroyl (10), feruloyl (12), 5-hydroxyferuloyl (13) and sinapoyl (14) CoAs, as well as caffeoyl CoA (11) mentioned above, are converted into the corresponding aldehydes by CCR.} \]

The gene encoding CCR was first cloned from Eucalyptus gunnii after confirmation of CCR activity from cell-free extracts of Salix alba. 11 putative AtCCRs were also annotated in the TAIR, and, of them, AtCCR1, 2 and 8 were putatively biochemically characterized, (Guerra et al., unpublished). Lauvergeat et al. indicated that AtCCR1 was 5 times more efficient for feruloyl-(12) and sinapoyl-(14) CoA than AtCCR2. Patten et al. further reported that feruloyl-CoA (12) was the most efficient substrate for AtCCR1, following by 5-hydroxyferuloyl-CoA (13), through kinetic analyses of native AtCCR1. Interestingly, AtCCR8 had very low CCR activity, but could catalyze p-coumaryl (15), caffeoyl (16) and sinapyl (19) aldehydes into the corresponding alcohols (Guerra et al., unpublished).

Cinnamyl alcohol dehydrogenase (CAD, EC 1.1.1.195), the last step to the monolignols, catalyzes conversion of hydroxycinnamyl aldehydes into hydroxycinnamyl alcohols (Figure 1.9). CAD activity was first detected in Salix alba, and the gene encoding CAD was first cloned from N. tabacum. After the genome sequencing of Arabidopsis thaliana, a model plant, was
completed. However, based on homology analysis with \textit{bona fide} \textit{N. tabacum} and \textit{Pinus taeda} CADs, 8 of these 17 putative CAD homologues were not considered to be putative CAD homologues due to their very low homology (0.9 – 1.6 % similarity) to \textit{bona fide} CADs from \textit{N. tabacum} and \textit{P. taeda} and also their lacked catalytic zinc. The remaining 9 CAD homologues showed relatively high similarity (57.1 – 82.9 %) and identity (46.0 – 76.5 %) with \textit{N. tabacum} and \textit{Pinus taeda} CADs, and further \textit{in vitro} assays of them established that AtCAD4 and AtCAD5, with highest homology to \textit{bona fide} \textit{N. tabacum} and \textit{Pinus taeda} CADs, had the highest catalytic activity. \textit{Figure 1.9.} Enzymatic reactions catalyzed by AtCAD5.

Based on the crystal structure of horse liver alcohol dehydrogenase (HADH, EC 1.1.1.1), the structure of \textit{Eucalyptus gunii} CAD2 (EgCAD2) was modeled based on the crystal structure of horse liver alcohol dehydrogenase (HADH, EC 1.1.1.1). The modeled structure of EgCAD2 indicated that five conserved features were shared by all ADHs, including catalytic zinc-binding residues, structural zinc-binding residues, Rossmann fold (GXGXXG motif), nicotinamide-
neighboring residues and dimer interface residues. Especially, residues of Ser49 and His52, corresponding to Ser48 and His51 in HADH, were also putatively involved in the proton relay, as proposed in HADH.\textsuperscript{83} However, two putative catalytic mechanisms, including proton relay\textsuperscript{84} and penta-coordinated zinc system\textsuperscript{85}, were also postulated as being involved in alcohol dehydrogenases. In the putative proton relay in HADH, a proton is transferred into a water molecule in solution via Ser48 and His51, and the catalytic zinc is tetra-coordinated during catalysis.\textsuperscript{84} The putative penta-coordinated zinc involvement, by contrast, does not need a proton relay for catalysis; instead, there are two penta-coordinated transient complexes, including the first transient complex formed by addition of a water molecule and the second transient complex formed by deletion of Glu60 and further addition of substrate.\textsuperscript{85} In order to investigate putative catalytic mechanisms of CADs, the crystal structures, including apo-, binary and ternary, of CADs should be determined. After the determination of crystal structures of CADs, residues involved in the putative proton relay need also to be evaluated through site-directed mutagenesis. This is the focus of our studies on CAD. Specifically, AtCAD5 and AtCAD4 were chosen to examine herein due to its importance for lignin biosynthesis\textsuperscript{86,87} and having the highest catalytic activities \textit{in vitro}\textsuperscript{47}. Therefore, in CHAPTER TWO in this dissertation, the crystal structures of AtCAD5 and AtCAD4 were investigated. The putative proton relay mechanism in AtCAD5 was further evaluated through site-directed mutagenesis with kinetic and isothermal titration calorimetry (ITC) analyses in CHAPTER THREE in this Ph.D. thesis.

After conversion of coniferyl aldehyde (17) into both coniferyl alcohol (22) and dihydroconiferyl alcohol (44) was detected during \textit{Saccharomyces cerevisiae} fermentation\textsuperscript{88}, Savidge and Förster also reported an \textit{in vitro} conversion of coniferyl aldehyde (17) into both coniferyl alcohol (22) and dihydroconiferyl alcohol (44) through assays using microsomes,
prepared from lignifying xylem of *Pinus strobus*, with NADPH\(^7\). By contrast, the actual gene, named *phenylpropenal double bond reductase (PPDBR)*, involved in reduction of double bond, was first cloned and characterized by us from *Pinus taeda*.\(^8\) PPDBR can also catalyze conversion of dehydrodiconiferyl (61) and coniferyl (17) aldehydes into dihydrodehydrodiconiferyl (62) and dihydroconiferyl (49) aldehydes (Figure 1.10, A and C).\(^8\) In *A. thaliana*, the allylic double bond reductase, namely AtDBR1 (homologous to PPDBR), was reported.\(^8,9\) AtDBR1 is also able to catalyze conversion of coniferyl aldehyde (17), as in the reaction of PPDBR, into dihydroconiferyl aldehydes (49), as well as 4-hydroxy-(2E)-nonenal (4-HNE, 63) into 4-hydroxynonanal (4-HNA, 64), but not dehydrodiconiferyl aldehydes (61) into dihydrodehydrodiconiferyl aldehyde (62) as distinct from PPDBR (Figure 1.10, A and B).\(^9\) Interestingly, knockdown of *Pinus radiata CAD* gene by RNAi caused significant accumulation of *p*-dihydrocoumaryl (50) and dihydroconiferyl (44) alcohols, indicating the correlation between decrease of CAD activity and relatively high frequency of use of *p*-coumaryl (15) and coniferyl (17) aldehydes by putative enzymes involved in dihydrophenylpropanoids biosynthesis.\(^8,9\) This dissertation focuses upon the catalytic mechanism and structures of AtDBR1. Therefore, the structures of AtDBR1 were examined in CHAPTER FOUR in this Ph.D. thesis, in order to gain insight into the catalytic mechanism of AtDBR1, supposed to be helpful for the biotechnological production of dihydrophenylpropanoids.

Podophyllotoxin (1), also focused upon in this Ph.D. thesis, is a natural plant product, identified in *Podophyllum hexandrum* and *P. peltatum*.\(^54\) It was reported that podophyllotoxin (1) prevents microtubule assembly, causing death of cancer cells.\(^53\) Its semi-synthetic derivatives, including etoposide (45), teniposide (46) and Etopophos\(^\circledR\) (47), are widely used for treatment of various cancers, including lung cancer, leukemia and other solid tumors.\(^64,65,90\) Interestingly, the
addition of the bulky glucoside moiety into podophyllotoxin (1) by semi-synthesis changed their target from microtubule assembly into DNA topoisomerase II.\textsuperscript{62-65} \textit{P. hexandrum}, however, is endangered because of its difficulty to cultivate and over-harvesting from its wild habitat, whereas \textit{P. peltatum} is currently not utilized due to its lower podophyllotoxin (1) content when compared to \textit{P. hexandrum}.\textsuperscript{54, 65, 91-93} Thus, an alternative method, i.e., biotechnological production methods including plant cell culture, is needed for production of podophyllotoxin (1).\textsuperscript{65, 94, 95}

**Figure 1.10.** Enzymatic reactions catalyzed by AtDBR1 and/or PPDBR. A. Reactions by AtDBR1 and PPDBR, B. Reaction by AtDBR1, C. Reaction by PPDBR. Adapted, modified and redrawn from Kasahara \textit{et al}.\textsuperscript{8} and Youn \textit{et al}.\textsuperscript{9}
Xia et al. elucidated that podophyllotoxin (1) is biosynthesized from coniferyl alcohol (22) via involvement of a (+)-pinoresinol (65)-forming dirigent protein (DP) in *P. peltatum* (Figure 1.11).\(^{11,96}\) This was done after the corresponding (+)-pinoresinol (65)-forming DP was characterized in *Forsythia intermedia*\(^{96}\), and the homologue was later cloned from a *P. peltatum* cDNA library\(^{11}\). It had ~60.4 % identity with the *F. intermedia* (+)-pinoresinol (65)-forming DP.\(^{11}\) Recently, a (-)-pinoresinol (66)-forming DP was also reported in *A. thaliana*\(^{97,98}\).

The second enzymatic step in podophyllotoxin (1) biosynthesis involves (+)-pinoresinol (65)/(+)-lariciresinol (67) reductase [(+)−PLR], which catalyzes the NADPH-dependent conversion of (+)-pinoresinol (65) into (-)-secoisolariciresinol (68) via (+)-lariciresinol (67).\(^{11,99}\) The corresponding gene was cloned and characterized initially in *Forsythia intermedia*.\(^{100}\) Later, the genes encoding (-)-pinoresinol (66)/(-)-lariciresinol (69) reductases [(-)-PLR], which catalyze the conversion of (-)-pinoresinol (66) into (+)-secoisolariciresinol (70) via (-)-lariciresinol (69), were cloned and characterized from *Thuja plicata*\(^{101}\) and *Linum usitatissimum*\(^{102}\) (Figure 1.12).

![Figure 1.11](image-url)  
*Figure 1.11.* Proposed biosynthetic pathway to podophyllotoxin (1) in *Podophyllum peltatum*. (+)-PLR: (+)-pinoresinol/(+)-lariciresinol reductase, SDH: secoisolalicresinol dehydrogenase. Adapted and redrawn from Xia *et al.*\(^{12}\) and Moinuddin *et al.*\(^{103}\)
The next step in podophyllotoxin (1) biosynthesis involves enantio-specific conversion of (-)-secoisolariciresinol (68) into (-)-matairesinol (71). This reaction is catalyzed by secoisolariciresinol dehydrogenase (SDH) with the corresponding gene cloned and characterized from the P. peltatum cDNA library. The P. peltatum SDH catalyzes the enantio-specific conversion of (-)-secoisolariciresinol (68) into (-)-matairesinol (71) via the (-)-lactol (72). However, there are still unknown steps in the biosynthetic pathways of podophyllotoxin (1), including possible reactions for methylene-dioxy bridge formation, hydroxylation, intramolecular C-C coupling and methylation. Various genes for some of these possible reactions were already known in alkaloid or lignan biosynthesis.

The gene encoding methylene-dioxy bridge formation was reported as CYP81Q1 for furofuran lignan biosynthesis from Sesamum indicum, this step having been discovered in our laboratory. CYP81Q1 catalyzes conversion of (+)-pinoresinol (65) into (+)-sesamin (42) via (+)-piperitol (53) by two subsequent methylene-dioxy bridge formation (Figure 1.13). In alkaloid biosynthesis, CYP719As catalyze various reactions of methylene-dioxy bridge formation, including (S)-canadine (73, Figure 1.14) (CYP719A1, Coptis japonica), (S)-stylopine (74) (CYP719A2 and CYP719A3, Eschscholzia californica; CYP719A13, Argemone Mexicana).
and (S)-cheilanthifoline (75) (CYP719A14, *A. Mexicana*) formation. CYP719 is Ranunculales specific family, and, interestingly, *Podophyllum hexandrum* and *P. peltatum* belong to Ranunculales, indicating that these *Podophyllum* species might have CYP719 genes.

![Enzymatic reactions catalyzed by CYP81Q1 in *Sesamum indicum*.](image)

**Figure 1.13.** Enzymatic reactions catalyzed by CYP81Q1 in *Sesamum indicum*. Adapted, modified and redrawn from Ono *et al.*

CYP719B1 was reported to be involved in salutaridine (76) formation by the C-C phenol coupling in *Papaver somniferum*. CYP80G2 was also characterized as a C-C phenol coupling enzyme in *C. japonica*, catalyzing conversion of (S)-reticuline (77b) into (S)-corytuberine (78). Several CYP450s were further characterized as hydroxylases in alkaloid biosynthesis: CYP80B1 catalyzes 3'-hydroxylation of (S)-N-methylcoclaurine (79); CYP71D12 is involved in 16-hydroxylation of tabersonine (80); CYP71BJ1 adds hydroxyl function at 19-position of tabersonine (80) and lochnericine (81). Therefore, these indicate that the analysis of CYP450s is useful to get target gene for each possible reaction due to involvement of CYP450s in various reactions in natural product biosynthesis.

In addition, several *O*-methyltransferases (OMTs), including (R,S)-norcoclaure (82 a/b) 6-OMT, (R,S)-reticuline (77 a/b) 7-OMT and 3'-hydroxy-N-methylcoclaurine (83 a/b) 4'-OMT, were also reported in alkaloid biosynthesis. Although various other *O*-methyltransferases
were cloned and characterized in planta\textsuperscript{75, 118-125}, OMTs in alkaloid biosynthesis were selected because of structural similarity between podophyllotoxin (1) and alkaloids, including (R,S)-norcoclaurine (78), (R,S)-3′-hydroxy-N-methylcoclaurine (83 a/b) and (R,S)-reticuline (77 a/b).\textsuperscript{11, 12, 111, 112}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{alkaloids}
\caption{Various alkaloids}
\end{figure}

As mentioned above, multiple steps are unknown in the biosynthetic pathway of podophyllotoxin (1). Towards the elucidation of these unknown steps is one of the goals in this dissertation. As done by Xia et al.\textsuperscript{11} and Xia et al.\textsuperscript{12}, transcriptome from tissues in which podophyllotoxin (1) accumulates is the base for the elucidation of unknown steps, but
conventional methods, including PCR and colony hybridization, are not efficient tools for massive works\textsuperscript{11, 12}. For example, \textit{A. thaliana} has 245 CYP450s after excluding predicted pseudogenes.\textsuperscript{116} However, \textit{P. hexandrum} and \textit{P. peltatum} are non-model plants with little information of their genomes. Currently, next-generation DNA sequencing, including 454, Illumina, SOLiD and HeliScope, is widely being used for various purposes, including genome sequencing, transcriptome sequencing and small RNA sequencing.\textsuperscript{126, 127} Next-generation DNA sequencing is able to sequence massively, compared to Sanger sequencing, but it generates relatively short sequences (30 – 40 bp), except for 454 generating relatively long sequence (200 – 300 bp), followed by assembly.\textsuperscript{126-128} Therefore, next-generation sequencing can be an efficient tool for studying transcriptomes of non-model plants.\textsuperscript{122, 123} Moreover, if a CYP450 database can be established, further homology analyses against reference genes can be easily done. In CHAPTER FIVE in this dissertation, transcriptome data of \textit{P. hexandrum} and \textit{P. peltatum} were investigated through Illumina sequencing, next-generation sequencing, following by metabolomics profiling analysis, and the unknown steps in podophyllotoxin (1) biosynthesis were elucidated through transcriptome analysis.

Summarized herein are the main findings of each chapter in this dissertation. In CHAPTER TWO, the crystal structures of the apo and binary (with NADP\textsuperscript{+}) complex of \textit{A. thaliana} cinnamyl alcohol dehydrogenase 5 (AtCAD5) are reported.\textsuperscript{129} This enzyme has the highest catalytic activity in AtCADs, and catalyzes conversion of hydroxycinnamyl aldehydes into hydroxycinnamyl alcohols, \textit{in vitro} and \textit{in vivo} CADs. Analysis of AtCAD5 established that it was a homodimer containing two zinc ions, are catalytic and structural. A binary complex structure with NADP\textsuperscript{+} was obtained, and a ternary complex was modeled with \textit{p}-coumaryl aldehyde (15). A putative proton shuttle mechanism, i.e., an extended proton relay system, was
then provisionally proposed for AtCAD5 from the modeled ternary complex and by comparison with the proposed proton relay mechanism for horse liver alcohol dehydrogenase. AtCAD4, which exhibits the next highest catalytic activity to AtCAD5, was also modeled based on the crystal structures of AtCAD5, and its modeled structure was not significantly changed, compared to AtCAD5.

Furthermore, in CHAPTER THREE, an extended proton relay system was evaluated through kinetic and isothermal titration calorimetry (ITC) analyses of each site-directed mutant of Thr49, His52 or Asp57, proposed as essential residues in the putative catalytic mechanism of AtCAD5. The Thr49 residue was confirmed to be essential for overall catalysis because the Thr49Ala mutation gave a near abolition of catalysis. On the other hand, His52 and Asp57 had little effect on catalysis. In addition, the thermodynamic data for the Thr49Ala mutant exhibited a negative enthalpic change (ΔH) and a significant decrease in binding affinity with NADPH, compared to the wild type AtCAD5. ITC analyses also suggested that during catalysis, NADPH was bound to AtCAD5 first followed by p-coumaryl aldehyde (15).

In CHAPTER FOUR, the crystal structures of apo, binary and ternary complexes were determined from the Arabidopsis allylic (double bond) reductase 1 (AtDBR1), which catalyzes conversion of hydroxycinnamyl aldehydes into dihydrohydroxycinnamyl aldehydes. 4-Hydroxy-2E-nonenal (4-HNE, 63), a toxic substance, could be also reduced by AtDBR1. The ternary complex structures exhibited that Tyr260 was hydrogen-bonded to O2´ of nicotinamide ribose in NADP+ and the aldehydic oxygen of p-coumaryl aldehyde (15) or 4-HNE (63), and was proposed to be involved in hydride transfer as a general acid/base.

In CHAPTER FIVE, the regio-specific methylenedioxy bridge-forming CYP450s were elucidated in podophyllotoxin (1) biosynthesis through Illumina sequencing, a next-generation
sequencing method, and metabolomics analyses. CYP719A23 from *Podophyllum hexandrum* and CYP719A24 from *P. peltatum* could convert (−)-matairesinol (71) into (−)-pluviatolide (84) catalyzing methylenedioxy bridge formation (Figure 1.15). CYP719A23 and CYP719A24 are included in subfamily A of CYP719 family, which shows the methylenedioxy bridge-forming activity in alkaloid biosynthesis, whereas two CYP719s are different to CYP450s, i.e. CYP81Qs, involved in lignan biosynthesis.

![Chemical structures](image.png)

**Figure 1.15.** Conversion of (−)-matairesinol (71) into (−)-pluviatolide (84) by CYP719A23 (*P. hexandrum*) and CYP719A24 (*P. peltatum*).
1.2 ATTRIBUTIONS

Part of the work mentioned in this dissertation was carried out by the collaboration with members of Dr. Norman G. Lewis’ research group and other research groups. All chapters were completed under Dr. Norman G. Lewis’ critical review.

For CHAPTER TWO, Choonseok Lee conducted the generation and characterization of the Glu70Ala mutant, a site-directed mutant, of Arabidopsis thaliana cinnamyl alcohol dehydrogenase 5 (AtCAD5). Choonseok Lee purified the recombinant proteins of AtCAD5 and AtCAD5 E70A and confirmed the activity of the Glu70Ala mutant of AtCAD5.

For CHAPTER THREE, Choonseok Lee wrote this draft section and carried out work to generate AtCAD5 T49A, AtCAD5 H52A and AtCAD5 D57A, the site-directed mutants of AtCAD5, followed by the expression and purification of AtCAD5 and its site-directed mutants. Ms. Diana Bedgar carried out the bulk purification of AtCAD5 and its site-directed mutants and conducted the enzymatic assays for the kinetic analyses. The quantification of the products in each assay reaction was performed by Choonseok Lee using Ultra Performance Liquid Chromatography® (Waters). All data were then processed by both Choonseok Lee and Dr. Laurence B. Davin with calculation of $V_{\text{max}}$ and $K_m$ values obtained using Origin software. Choonseok Lee carried out the isothermal titration calorimetry (ITC) experiments and the analysis of the ITC data.

For CHAPTER FOUR, Choonseok Lee generated and characterized the site-directed mutants – AtDBR1 Y81A, Y260A and S287A – of the A. thaliana double bond reductase 1 (AtDBR1), as characterized the corresponding recombinant proteins. For the latter, Choonseok Lee carried out the purifications of the recombinant enzymes – AtDBR1 and its site-directed mutants, and performed all kinetic analyses for AtDBR1 and its site-directed mutants.
For CHAPTER FIVE, Choonseok Lee prepared total RNA for transcriptome profiling, and then with the assemblies provided by the National Center for Genome Resources, and used a bioinformatics and BLAST analysis to identify homologues of genes in the shikimate, phenylalanine, phenylpropanoid and lignan biosynthetic pathways with *Podophyllum hexandrum* and *P. peltatum*. 
1.3 REFERENCES


53. J. D. Loike and S. B. Horwitz, Biochemistry, 1976, 15, 5435-5443.


76. R. Bugos, V. C. Chiang and W. Campbell, Plant Molecular Biology, 1991, 17, 1203-1215.


120. A. Berim, B. Schneider and M. Petersen, Plant Molecular Biology, 2007, 64, 279-291.


CHAPTER TWO

CRYSTAL STRUCTURES AND CATALYTIC MECHANISM OF THE *Arabidopsis* CINNAMYL ALCOHOL DEHYDROGENASES AtCAD5 AND AtCAD4

Buhyun Youn, Roy Camacho, Syed G.A. Moinuddin, Choonseok Lee, Laurence B Davin, Norman G. Lewis, and ChulHee Kang


Copyright 2006

Abstract

The cinnamyl alcohol dehydrogenase (CAD) multigene family in planta encodes proteins catalyzing the reductions of various phenylpropenyl aldehyde derivatives in a substrate versatile manner, and whose metabolic products are the precursors of structural lignins, health-related lignans, and various other metabolites. In *Arabidopsis thaliana*, the two isoforms, AtCAD5 and AtCAD4, are the catalytically most active being viewed as mainly involved in formation of guaiacyl/syringyl lignins. In this study, we determined the crystal structures of AtCAD5 in the apo-form and binary complex with NADP⁺, respectively, and modeled that of AtCAD4. Both AtCAD5 and AtCAD4 are dimers with two zinc ions per subunit and belong to the Zn-dependent medium chain dehydrogenase/reductase (MDR) superfamily, on the basis of their overall 2-domain structures and distribution of secondary structural elements. The catalytic Zn²⁺ in both enzymes are tetrahedrally coordinated, but differ from horse liver alcohol dehydrogenase since the carboxyl side-chain of Glu70 is ligated to Zn²⁺ instead of water. Using AtCAD5, site-
directed mutagenesis of Glu70 to alanine resulted in loss of catalytic activity, thereby indicating that perturbation of $\text{Zn}^{2+}$ coordination was sufficient to abolish catalytic activity. The substrate binding-pockets of both AtCAD5 and AtCAD4 were also examined, and found to be significantly different and smaller compared to that of a putative aspen sinapyl alcohol dehydrogenase (SAD) and a putative yeast CAD. While the physiological roles of the aspen SAD and yeast CAD are uncertain, they nevertheless have high similarity in overall 3D structures to AtCAD5 and 4. With the bona fide CAD’s from various species, nine out of the twelve residues which constitute the proposed substrate-binding pocket were, however, conserved. This is provisionally considered as indicative of a characteristic fingerprint for the CAD family.

Abbreviations: ADH, alcohol dehydrogenase; CAD, cinnamyl alcohol dehydrogenase; MDR, medium chain dehydrogenase/reductase; PDB, protein data bank; SAD, sinapyl alcohol dehydrogenase.

2.1 INTRODUCTION

Cinnamyl alcohol dehydrogenases (CAD’s, EC 1.1.1.195) catalyze the substrate versatile, NADPH-dependent, conversion of $p$-hydroxycinnamyl aldehydes 1-5 in vitro into the corresponding alcohols 7-11,1, 2 (Figure 2.1). The latter metabolites, in turn, are obligatory precursors of structural cell-wall lignins3, 4 and health-related lignans5-10 e.g., including the antiviral agent podophyllotoxin (13) from Podophyllum species.8 Podophyllotoxin (13) finds application in the semi-synthesis of the widely employed anticancer compounds, teniposide (14), etoposide (15) and Etopophos® (etoposide phosphate, 16).11 CAD is also an essential step on the
pathway to other lignans, such as matairesinol (17)⁸ and secoisolariciresinol (18),⁵,⁶ which can serve as plant precursors of the “mammalian” lignans, enterolactone (19) and enterodiol (20).¹²,¹³

**Figure 2.1.** Cinnamyl alcohol dehydrogenase substrates and products.
In Arabidopsis, there are nine Zn\(^{2+}\)-dependent CAD’s or homologues thereof\(^{14}\) belonging to the medium chain dehydrogenase/reductase (MDR) superfamily, with two Zn\(^{2+}\) ions per subunit.\(^{15}\) This MDR superfamily is found in a variety of organisms, such as bacteria, fungi, plants, cephalopods and vertebrates, of which the best known example is horse liver alcohol dehydrogenase.\(^{16}\) In Arabidopsis, two isoforms of the CAD family [AtCAD4 (At3g19450) and AtCAD5 (At4g34230)] are catalytically the most active in vitro with \(p\)-hydroxycinnamyl aldehydes 1-5, albeit differing in relative abilities to process sinapyl aldehyde (5); AtCAD5 is evidently ~270-fold more efficient than AtCAD4.\(^{14}\) Overall, however, \(p\)-hydroxycinnamyl aldehyde (1) is the preferred substrate for both isoforms, as well as for the catalytically less active AtCAD2, 3, 7 and 8. The remaining 3 putative homologues (AtCAD1, 6 and 9) displayed no activities when incubated with substrates 1-5.

In terms of precise physiological function, a double mutant of both AtCAD4 and 5 in Arabidopsis\(^{17}\) markedly reduced the ability to synthesize either coniferyl (9) or sinapyl (11) alcohols, and hence formation of the corresponding guaiacyl (G) and syringyl (S) lignins. This suggests therefore that in Arabidopsis AtCAD4 and 5 are largely responsible for the formation of 9 and 11 in lignifying tissues.

Much of our current understanding of alcohol dehydrogenases (ADH’s) stems from comprehensive studies on the aliphatic horse liver ADH crystal structure,\(^{18}\) and this knowledge has been applied to many other aliphatic ADH’s. By contrast, little is known about aromatic alcohol dehydrogenases, such as CAD’s, in spite of their enormous importance in vascular plant formation, in health-related areas, in biotechnology and so forth. There are useful reports, however, of crystal structures of a putative CAD in yeast (Saccharomyces cerevisiae)\(^{19}\) and an aspen sinapyl alcohol dehydrogenase (SAD).\(^{20}\) These enzymes, like horse liver ADH, show
some similarity to AtCAD5 and AtCAD4 in terms of amino acid sequences, even though their precise biochemical/physiological significance is uncertain. *S. cerevisiae* lacks a biochemical pathway to either monolignols 7-11 or cinnamyl alcohol 12, and the biochemical/physiological role of the putative SAD is also uncertain. Furthermore, current evidence of the biochemical/physiological properties of the *Arabidopsis* CAD isoforms in vitro14 and in vivo17 has given no indication that there is a SAD-specific enzyme as previously reported.21 Indeed, from an enzymological perspective, the aspen SAD displays considerable substrate versatility for aldehydes 1-5 in vitro, suggesting that there is no role specific to a particular substrate. The analysis of the *Arabidopsis* AtCAD4 and AtCAD5 double mutant also indicated that there was not a specific SAD responsible solely for sinapyl alcohol (11) formation in lignification.17

In this study, we examined the crystal structures of the two bona fide CAD’s, AtCAD4 and AtCAD5, both of which have established roles in monolignol, lignan and lignin formation in vivo. In addition, we considered it instructive to conduct a detailed comparison of AtCAD4 and AtCAD5 with the putative aspen SAD and yeast CAD, particularly with respect to their binding site geometries and amino acid compositions.

2.2 RESULTS AND DISCUSSION

Overall structure

Recombinant AtCAD5 was first crystallized in both its apo form and binary complex, with crystals of the latter obtained by mixing with NADP+. The structure of the apo-form was determined at 2.0 Å resolution by molecular replacement using coordinates of a yahK, a zinc-type alcohol dehydrogenase-like protein (1UUF) from *Escherichia coli*, which at the initiation of our studies had the highest sequence similarity to AtCAD5 in the Protein Data Bank (PDB). The
binary complex structure of AtCAD5 (with NADP\(^+\)) was determined at 2.6 Å resolution, using the coordinates of the deduced structure of the apo-form.

The asymmetric unit of the AtCAD5 crystal is composed of one molecule and the crystallographic 2-fold axis produces a dimer tightly associated through two 2-fold related β strands (βF). Consequently, its dimer forms an extended 12 stranded β-sheet with 6 strands (βA-βF) from each subunit (Figure 2.2). Additionally, with the differences in substrate preferences noted for AtCAD5 and AtCAD4 regarding sinapyl aldehyde (5),\(^1\) the structure of AtCAD4 was also modeled using the refined coordinates of AtCAD5 followed by the process of energy minimization. As expected, since AtCAD4 has 77.5% identity with AtCAD5, the overall backbone structure of the modeled AtCAD4 had no significant changes (Figure 2.2, inset).

AtCAD5 and AtCAD4 were studied by light scattering to determine their tendencies to form oligomers. Both static and dynamic light-scattering experiments further confirmed that most of the population of AtCAD5 and AtCAD4 were in dimer form in both phosphate buffered saline (pH 7.0) and 20 mM Tris-HCl buffer (pH 8.0) (Figure 2.3). This quaternary structure of AtCAD5 and AtCAD4 is similar to other Zn-containing ADH’s from higher plants and mammals. In addition, the elution volumes of both AtCAD4 and AtCAD5 reflected their extended hydrodynamic volumes in agreement with the asymmetric shape of dimers shown in Figure 2.2.
Figure 2.2. Crystal structure of AtCAD5 homodimer and energy minimized model of AtCAD4 (inset). The catalytic and nucleotide-binding domains are colored in light blue and violet for the bottom subunit, and green and dark orange for the upper subunit, respectively. Catalytic and structural Zn2+ are depicted as red dots. Secondary structural elements have been numbered sequentially as $\alpha_1$-$\alpha_6/\alpha_A$-$\alpha_E$ and $\beta_1$-$\beta_{12}/\beta_A$-$\beta_F$ for the $\alpha$-helices and $\beta$-strands, respectively.
Figure 2.3. Molecular mass determination of AtCAD5 and AtCAD4. (a) Multiangle laser light-scattering elution profile of AtCAD5 (red) and AtCAD4 (blue) (~2 mg cm$^{-3}$ each). Elution profile is shown as molecular weight versus elution time. The thin solid lines represent changes in refractive index on an arbitrary scale that is proportional to protein concentration. The thick solid lines indicate calculated molecular masses (thick green line: AtCAD5, thick purple line: AtCAD4). (b) Dynamic light scattering data of AtCAD5 (red) and AtCAD4 (blue) (~2 mg cm$^{-3}$ each). The calculated molecular radius and molecular weight are 3.86 nm and 82 kDa for AtCAD5, and 3.94 nm and 83 kDa for AtCAD4, respectively.
Each AtCAD5 subunit is composed of two distinct domains, namely a Rossmann fold forming the nucleotide-binding domain (residues 163–301) and a catalytic domain (residues 1–162 and 302–357) (Figure 2.4). Interestingly, AtCAD5 has a similar overall fold to other ADH’s that belong to the same MDR superfamily in spite of its relatively low level of sequence similarity (20-23%). Hence, in Fig. 2.2 and 2.4, the secondary structural elements of AtCAD5 can be conveniently described by comparison with horse liver ADH (8ADH) nomenclature. The nucleotide-binding domain is thus composed of a β-pleated sheet of six parallel strands (βA to βF) flanked by five helices (αA to αE), whereas the catalytic domain consists mainly of a core of antiparallel β-strands (β1 to β9) with six helical segments (α1 to α6) at the surface of the molecule. As with mammalian ADH’s, there are two Zn$^{2+}$ ions in AtCAD5, one catalytic and the other structural, with both being coordinated by residues solely from the catalytic domain (Figure 2.5b-2.5d). The catalytic Zn$^{2+}$ is located inside a cleft formed between the two domains and is positioned at the bottom of the hydrophobic substrate-binding pocket, coordinated by four residues, Cys47, His69, Glu70 and Cys163, in an approximate tetrahedral symmetry (Figure 2.5b). Notably, the O$^{ε1}$ atom of the carboxyl side-chain of Glu70 coordinates catalytic Zn$^{2+}$ instead of water, with the latter being more typical of other Zn-dependent MDR’s including horse liver ADH (8ADH)$^{22, 23}$. By contrast, the structural Zn$^{2+}$ is located in a short α-helix containing loop (residues 98-116), which protrudes somewhat from the main catalytic domain. It is also in a tetrahedral coordination with cysteine residues, Cys100, Cys103, Cys106 and Cys114, respectively. Thus, taken together the observed structural features establish that the crystal structure of AtCAD5 belongs to the Zn$^{2+}$-dependent MDR family, as previously predicted.$^{14}$
| ATACD45 | 1 | MG1MSA-ER-K scrolls around BGL | 347 |
| ATACD44 | 1 | MG1SVEAGK-K scrolls around BGL | 347 |
| ATACD43 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| NTACD | 1 | MG1S1MV9-ER-K scrolls around BGL | 347 |
| ATACD42 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD41 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD40 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD39 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD38 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD37 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD36 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD35 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD34 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD33 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD32 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD31 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD30 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD29 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD28 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD27 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD26 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD25 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD24 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD23 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD22 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD21 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD20 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD19 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD18 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD17 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD16 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD15 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD14 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD13 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD12 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD11 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD10 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD9 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD8 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD7 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD6 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD5 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD4 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |
| ATACD3 | 1 | MG1S1M0V9-ER-K scrolls around BGL | 347 |
| ATACD2 | 1 | MG1S1M1G1V9-ER-K scrolls around BGL | 347 |
| ATACD1 | 1 | MG1S1M1V9-ER-K scrolls around BGL | 347 |

**Nucleotide binding domain**

- **b1** (346-349)
- **b2** (360-364)
- **b3** (331-345)

![Diagram of nucleotide binding domain](image-url)
Figure 2.4. Amino acid sequence comparisons of AtCAD5, AtCAD4, Nicotiana tabacum CAD (NtCAD, Genbank accession number X62344), aspen (Populus tremuloides) CAD (AsCAD, Genbank accession number AF217957), Eucalyptus gunnii CAD (EgCAD2, Genbank accession number X65631), Pinus taeda CAD (PtCAD, Genbank accession number Z37992), putative sinapyl alcohol dehydrogenase (SAD) from aspen (AsSAD, 1YQX, Genbank accession number AF273256), yahK from Escherichia coli (1UUF), a putative CAD from Saccharomyces cerevisiae (1Q1N), ADH from Bacillus stearothermophilus (1RJW), ADH from Pseudomonas aeruginosa (1LLU), ADH from Sulfolobus solfataricus (1R37) and horse liver ADH from Equus caballus (8ADH). Secondary structural elements of AtCAD5 are highlighted in colored bars on top of the corresponding sequence and the nucleotide-binding domain is boxed by a thin dotted line. The conserved glycine residues in GX(X)GXXG motif are marked with red dots. The conserved residues constituting a substrate-binding site are highlighted by green and gray colors indicating their belonging to different subunits. Eukaryotic 1YQX, 1Q1N and 8ADH are dimeric, whereas prokaryotic 1RJW, 1LLU and 1R37 are tetrameric Zn-dependent MDR’s. (The corresponding data for 1UUF has not been published.)
In the binary complex, NADP$^{+}$ is located at the active site clefts located between the catalytic and nucleotide-binding domains centered on the catalytic Zn$^{2+}$ as described in detail below. Upon cofactor binding, however, the backbone conformation of AtCAD5 did not change significantly (Figure 2.5a).

**Comparison to other MDR’s**

In terms of quaternary structure, the homodimeric MDR’s tend to occur in higher eukaryotes, whereas tetrameric MDR’s are found in prokaryotes and lower eukaryotes. Thus, in order to gain a preliminary insight into the catalytic mechanism and other functionally significant issues operative, structural comparisons with other MDR’s were next performed, through comparison of amino acid sequences (BLASTP), as well as 3D structures (Dali search), respectively.

First, amino acid sequence comparisons, through a BLASTP search in the NCBI database revealed that both AtCAD5 and AtCAD4 have the highest similarity (82.9 and 81.5%) and identity (76.5 and 75.1%) to a *bona fide* tobacco (*Nicotiana tabacum*) CAD.$^{14, 24}$ It also indicated that the similarity and identity of aspen SAD to AtCAD5 and 4 were rather low (i.e., 62.6/62.5% and 53.1/53.3%,$^{14}$ respectively), as were those for the putative yeast CAD (69.2/67.9% and 35.7/35.4%, respectively).

On the other hand, a Dali search$^{25}$ indicated that the most similar 3D structure currently in the PDB to AtCAD5/4 was the aspen SAD (AsSAD, 1YQX),$^{20}$ with a Z-score of 49.3 followed by the putative yeast CAD (1Q1N)$^{19}$ with a Z-score of 45.6.
Figure 2.5.  (a) Superimposed views of AtCAD5 in its apo- and binary complex forms. (b) Structure of the substrate-binding pocket of the NADP⁺ binary form of AtCAD5: The catalytic Zn²⁺ ion (red sphere) is tetrahedrally coordinated by Cys47, His69, Cys163 and Glu70 (blue) and the NADP⁺ molecule (orange) is held by Val192, Ser211, Ser212, Ser213, Lys216 and Gly275 (green). (c) Ternary complex model of AtCAD5 with p-coumaryl aldehyde (1) showing the structure of the substrate-binding pocket. Participating residues from one subunit are marked in green and from the other subunit in gray. The catalytic Zn²⁺ (red sphere) is tetrahedrally coordinated by Cys47, His69, Cys163 (blue) and the aldehyde oxygen of p-coumaryl aldehyde.
(1) (purple). (d) Surface representation of AtCAD5 active site cavity. Twelve residues (black) that constitute the substrate-binding pocket are indicated together with the one Zn-coordinating residue, Cys163, (yellow). In figures b and c, possible hydrogen bonds are shown as black dotted lines.

Additional PDB entries included an ADH-like protein yahK from *E. coli* (1UUF), an ADH from *Bacillus stearothermophilus* (1RJW)\(^{26}\) with Z scores of 45.4 and 45.3, respectively, as well as several other MDR’s, including 1R37 from *Sulfolobus solfataricus*\(^{27}\) and 1LLU from *Pseudomonas aeruginosa*\(^{28}\) (Figure 2.4). The amino acid sequence identities of the latter to AtCAD5/4 were much lower, i.e. 1UUF (42.9/41.2 %), 1RJW (33.2/30.9 %), 1R37 (26.5/27.8 %) and 1LLU (29.2/29.4 %), respectively.

Detailed sequence comparisons of the above-mentioned eukaryotic dimeric (8ADH, 1YQX, 1Q1N) and prokaryotic tetrameric (1RJW, 1LLU, 1R37) Zn-dependent MDR’s were also carried out. These enzymes have high sequential and structural similarities to AtCAD5, with the longest region of similarity located between Cys163 and Asp250, which in turn covers most of the nucleotide binding domain (Figure 2.4). Additionally, there is a substantial amount of sequential heterogeneity among the MDR’s in both the N and C-terminus regions, which are structurally disordered in most cases. There are also several areas of deletions and insertions in the amino acid sequences of the MDR’s compared. In particular, all the sequences shown in Figure 2.4 have a significant deletion in the loop area between β8 and β9 relative to horse liver ADH (8ADH), even though its deletion size is less severe than for tetrameric MDR’s. In horse liver ADH (8ADH), the flexible loop between β8 and β9 is located at the entrance to the active site contributing to substrate-binding, while at the same time restricting entry. The same deletion
occurs in a putative yeast CAD, this being considered to explain its broad range of substrate specificity.¹⁹

**Cofactor binding site**

The initial *Fo-Fc* map using diffraction data from the binary complex crystal, and the coordinates of the apo AtCAD5, clearly show the electron density corresponding to NADP⁺ (Figure 2.6). The nicotinamide ring is close to the catalytic Zn²⁺ at the bottom of the substrate-binding pocket (see Figure 2.5b) and both adenine and nicotinamide rings of NADP are in the syn-conformation. In the apo-form, the cofactor-binding pocket contains several water molecules, thereby forming a hydrogen-bonding network with the side-chains of the lined residues, especially in areas where phosphate groups reside (figure not shown). As predicted earlier, the nicotinamide ring is in the proper orientation for the A-face specific hydride transfer from C4 to the corresponding substrate.¹⁴ The binary complex with NADP⁺ maintains the same tetrahedral geometry of Zn²⁺ being coordinated by Cys47, His69, Glu70 and Cys163, as for the apo-form, which is different from two reported NADP⁺-binary complexes: ADH’s from *Thermoanaerobacter brockii* (TbADH) and *Clostridium beijerinckii* (CbADH).²⁹ In the binary complex of CbADH and TbADH with NADP⁺, however, the carboxyl side chain of Glu is no longer able to coordinate with Zn²⁺ indicative of a perturbation of Zn²⁺ coordination upon cofactor-binding.²⁹, ³⁰

Cofactor binding by AtCAD5 occurs through three flexible loops. As for other typical NAD(P)(H) dependent enzymes,⁹, ¹⁰ AtCAD5 has a glycine-rich motif at the first β-α-β unit of the cofactor-binding domain (¹⁸⁸GLGGVG¹⁹³), which participates in binding the pyrophosphate group of NADP⁺ through a helical dipole of αA. This typical GxGxxG spacing is shared by all Zn-dependent MDR’s of high similarity in Figure 2.4, except for ADH from *S. solfataricus*
In particular, the pyrophosphate group of the NADP$^+$ is within hydrogen bonding distance to the backbone amide nitrogen of residue Val192 (Figure 2.5b), thereby N-capping and compensating for the helix macro-dipole.$^{31}$ All the amino acids in this tight turn located between $\beta$A and $\alpha$B thus show relatively high temperature factors which probably facilitates their interaction with NADP$^+$ via conformational flexibility.

Figure 2.6. Difference Fourier maps for NADP$^+$ binary complexes of AtCAD5. A difference $(F_o - F_c)$ electron density map of the active site area is contoured at 2.0 $\sigma$.

Another loop region composed of residues 211-216 contains many conserved residues that also interact with NADP$^+$.$^{32}$ The side-chain of the highly conserved Ser213 in this area enables a preference for NADP(H) over NAD(H). The Ser213 is located at the carboxy end of the $\beta$B strand forming a hydrogen bond with the 2'-phosphate group of NADP(H), and this position is normally occupied by aspartic acid in enzymes that preferentially bind NAD(H), i.e. to form a hydrogen bond to both hydroxyl groups of the adenine ribose of NAD(H).$^{31, 32}$ Moreover, site directed mutagenesis of the Ser212 amino acid in $E.\text{gunnii}$ CAD2 to Asp212, followed by kinetic studies of both WT and mutant CAD2, had previously demonstrated the involvement of this residue in determining cofactor specificity.$^{33}$
In addition to Ser213 in the AtCAD5, the side-chains of neighboring residues, Ser211, Ser212 and Lys216 are located in the loop between βB and αC and are highly conserved among bona fide CADs (Figure 2.4). They form hydrogen bonds and electrostatic interactions with the 2′-phosphate of the adenine ribose (Figure 2.5b), and the side-chain of Lys216 is also within hydrogen bond distance with the O3′ of the same ribose ring. The third region involves cofactor binding from 275-286 (Figure 2.4), where the amide group of NADP⁺ interacts with the backbone of Gly275 located in the loop between βE and βF connecting two domains of the enzyme (Figure 2.5b).

Interestingly, the direct interaction of the above-mentioned three flexible loops, 188-191, 211-216 and 275-286, with the cofactor has been suggested to facilitate dissociation of products, a frequent rate-limiting step of many MDR enzymes.¹⁹

### Substrate binding

In spite of our extensive efforts to crystallize the ternary complex for AtCAD5, the initial Fo-Fc maps for all complex data gave no significant electron density suitable for positional refinement of the corresponding substrate p-coumaryl aldehyde (1). This was also the case for both the putative yeast CAD¹⁹ and the putative aspen SAD.²⁰ On the other hand, the unique and properly sized pocket for substrate-binding was easily deduced near the catalytic Zn²⁺. Indeed, after constraining the relevant aldehydic oxygen atom of the substrate via coordination to catalytic Zn²⁺ (with concomitant displacement of Glu70 from Zn²⁺), the resulting A-face of the nicotinamide ring was thus positioned for pro-R hydride transfer to and from its C4 atom. For example, Figure 2.5c shows the pro-S hydrogen of the C9 atom of the modeled p-coumaryl aldehyde (1) facing the C4 atom of the nicotinamide ring at a proper distance (~2.0 Å).
The putative substrate-binding site of AtCAD5 is lined with twelve residues, mostly hydrophobic amino acids. Nine residues from one subunit, Thr49, Gln53, Leu58, Met60, Cys95, Trp119, Val276, Phe299, Ile300 and three residues from the other subunit, Pro286, Met289 and Leu290, constitute the substrate-binding pocket (Figures 2.5c, 2.5d). The Cα carbon position of those residues thus superimposes well with those of the horse liver ADH\textsuperscript{23} and Eucalyptus gunnii CAD2 model.\textsuperscript{34} Notably, out of those 12 residues, nine amino acids are completely conserved among all bona fide CAD’s (Figure 2.4, residues in green). Two out of the remaining three also show a conservative heterogeneity with either serine or threonine for residue 49, and methionine or isoleucine for residue 289. Only the residue at 95 shows a higher polarity heterogeneity among bona fide CAD’s, that is, Leu, Val, Ile and Cys (Figure 2.4).

Interestingly, the entry region of the binding pocket of AtCAD5 is larger than that of horse liver ADH due to the major deletion between β8 and β9, although the precise physiological significance of this deletion cannot be gauged fully at this time.

As observed in some other Zn\textsuperscript{2+}-dependent MDR’s, the modeled substrate through its aldehydic group ligation to the catalytic Zn\textsuperscript{2+}, results in an extensive hydrogen-bonded network, thereby allowing coupling between hydride abstraction, transfer and protonation (Figure 2.5c). Specifically, the aldehydic oxygen of the substrate is within hydrogen bonding distance to the hydroxyl group of Thr49. In turn, the hydroxyl group of Thr49, which is conserved as either Thr or Ser in all ADH’s compared, is within hydrogen bonding distance to the O2’ of the nicotinamide ribose (not shown). As shown in Figure 2.5c and Figure 2.7, this residue is also in close proximity to the catalytic Zn\textsuperscript{2+}. Another highly conserved residue, His52, is also within hydrogen bonding distance to O3’ of the nicotinamide ribose (Figures 2.5b and 2.7). Therefore, both Thr49 and His52 not only fix the position of the nicotinamide ring during catalysis, but also
permit hydride transfer from cofactor to substrate. As for several other Zn$^{2+}$-dependent ADH’s, the Thr49 (or Ser) and His52 in AtCAD5, together with the O2’ and O3’ hydroxyl groups of the ribose ring, enable the proton relay mechanism through the hydrogen bond network among those functional groups, i.e. by shuttling a proton from the bulk solvent to the carbonyl oxygen of the aldehyde ligated to the catalytic Zn$^{2+}$ (Figure 2.7).

![Figure 2.7. Proposed proton shuttle mechanism during the reduction process in the active site of the AtCAD5. Solid arrows indicate the movement of two electrons among the functional groups during substrate reduction. The possible hydrogen bonds involved are shown with dotted lines.](image)

Significantly, there are no notable differences in the local conformations between AtCAD4 and AtCAD5. Indeed, most of the residues, within either contact or potential interacting distances to the substrates, were conserved between both. Only two of the 12 residues in the substrate-binding pocket differ; i.e. Cys95 and Met289 in AtCAD5 were replaced by Val96 and Ile290, respectively, in AtCAD4. This difference makes the binding pocket of the latter slightly more hydrophobic. Importantly, Cys95 that resides underneath Trp119 (Figures 2.5c, and 2.5d), and which corresponds to Phe93 in horse liver ADH (8ADH), has also been
proposed as affecting the substrate specificity in mammalian ADH’s and *S. solfataricus* ADH (1R37) by steric hindrance.\textsuperscript{27,35}

In an analogous manner to AtCAD5/4, provisional substrates, sinapyl (5), coniferyl (3) and cinnamyl (6) aldehydes were also modeled into the coordinates of the binary NADP\(^+\) complexes with the putative aspen SAD (AsSAD, 1YQX)\textsuperscript{20} and yeast CAD (1Q1N),\textsuperscript{19} respectively.

In this regard, the presumed active site of aspen SAD substantially differed from that of AtCAD5/4. Among the above-mentioned 12 residues constituting the proposed substrate-binding site, only two residues were identical to AtCAD5. Most of these substitutions were non-conservative, although there was no significant polarity change among them. As a result, the proposed tethering mechanism for the phenolic hydroxyl group of the corresponding substrates through the backbone carbonyl oxygen of Ala293 in the putative aspen SAD\textsuperscript{20} cannot occur in the binding-pocket of AtCAD5, due to both the changed conformation and resulting steric hindrance. Specifically, there were several other somewhat symmetrical substitutions, including the two pairs of amino acid residues previously noticed, namely Leu58/Pro286 and Trp119/Phe299 in AtCAD5 to Trp61/Phe289 and Leu122/Gly302 in SAD, respectively\textsuperscript{20} (Figure 2.4). The bulky side-chains of the hydrophobic residues, Trp119/Phe299, substantially shrink the size of the binding pocket and its entrance in AtCAD4/5 compared to that in the putative aspen SAD. Moreover, the Val276 residue in AtCAD4/5, that is in close contact with the modeled substrate, is substituted to Ala in aspen SAD, thus increasing the size of the pocket further. Taken together, these differences in binding pocket (size and shape) would appear to enable the putative aspen SAD to display even higher substrate versatility than either the AtCAD5 or AtCAD4.
The substrate-binding site of the putative yeast CAD (1Q1N) also substantially differs from that of AtCAD5. This is reflected in the observation that there is only one amino acid conserved among the 12 residues, with two of the others having substitutions of different polarity. Consequently, the funnel-shaped cavity of the putative yeast CAD is also relatively larger than that of AtCAD5. This perhaps helps to explain further its very broad substrate specificity range, as well as its being more polar than that of either AtCAD4/5 or the putative aspen SAD.

One of the two residues participating in the proton shuttle system, His52, is exposed to the bulk solvent, and is completely conserved among all of the compared MDR’s. This occurs even though Thr49 is substituted to Ser in some MDR’s, probably indicating that all of these reductases adopt an identical proton transfer system. By contrast, the alternative proton pathway that was proposed in horse liver ADH, from the 3’OH of a nicotinamide ring to the backbone carbonyl oxygen of Ile269 (Pro253 in AtCAD5), is not possible due to the relatively long distance between the two acceptor and donor groups. However, while this might be possible in the ternary complex (following conformational changes) this was not investigated.

**Site-directed mutagenesis of the fourth Zn$^{2+}$ coordination (Glu70) ligand**

In most ADH’s including horse liver ADH, the fourth ligand of catalytic Zn$^{2+}$ is a water molecule hydrogen-bonded to the hydroxyl group of Thr49 (or Ser) and which can be replaced by substrate. However, in some prokaryotic ADH’s, the catalytic Zn$^{2+}$ is instead coordinated to a strictly conserved Glu residue located opposite to the substrate binding pocket e.g. as observed for *S. solfataricus* (1R37), *C. beijerinckii* (CbADH) and *T. brockii* (TbADH). Additionally, among eukaryotic ADH’s, Zn$^{2+}$ coordination to a Glu residue is very rare. Moreover, since AtCAD5 (coordinated to Glu70) is only the second example of such a coordination in the
eukaryotes, following that of rat sorbitol dehydrogenase,\textsuperscript{38} this potentially suggested a role in catalysis. Indeed, upon substrate-binding, the aldehydic oxygen of the substrate is envisaged to displace Glu70 via direct coordination to Zn\textsuperscript{2+}, so that it remains tetra-coordinated.\textsuperscript{27, 39, 40} The resulting free side-chain of the Glu70 might then form a salt bridge to the nearby Arg345 as observed in apo-horse liver ADH.\textsuperscript{23} Interestingly, this Glu residue is also one of the most highly conserved in class I ADH\textapos;s, whose replacement in yeast ADH resulted in a considerable reduction in overall catalytic efficiency, even though the residue was not directly bound to catalytic Zn\textsuperscript{2+} in that enzyme.\textsuperscript{41} It has thus been proposed that the intermittent role of this dynamic Glu residue is in both facilitating binding of the aldehydic substrate and in subsequent product release.

The Zn\textsuperscript{2+}-coordination of the Glu70 is, however, observed in the apo-form of AtCAD5 and is maintained even after NADP\textsuperscript{+} binding, even though the opposite occurred in the case of the NADP\textsuperscript{+} complexes of both CbADH and TbADH.\textsuperscript{27, 29} We, therefore, considered it instructive to replace Glu70 with Ala70 via site-directed mutagenesis (see Experimental) to ascertain what effect, if any, this had on overall catalytic efficacy. The resulting mutagenized AtCAD5 (Glu70Ala) was thus next heterologously expressed, purified to apparent homogeneity, and assayed for its capacity to convert the corresponding aldehydes 1-5. The resulting mutagenized protein was catalytically inactive, in agreement with our hypothesis that perturbation of the Zn\textsuperscript{2+} catalytic center coordination adversely affected catalysis.

2.3 CONCLUSIONS

Herein we report the crystal structures of AtCAD5 and the modeled structure of AtCAD4, the two catalytically most active CAD\textapos;s from \textit{A. thaliana}. The observed structures underscore previous observations reported in various MDR-related ADH and provide us with a
useful tool for a comparative study of other enzymes of putative CAD and SAD-like character. In particular, we identified 12 residues apparently constituting the substrate-binding site, which are well conserved among the *bona fide* CAD’s from various species known thus far. By contrast, only 2 or 1 of these residues are conserved in the putative aspen SAD and in the putative yeast CAD, which in turn both result in significantly larger binding pockets. The true significance of such binding pocket differences will, however, only be objectively determined when the actual physiological roles of each are determined *in vivo* as previously noted by Bomati and Noel.\(^{20}\) In terms of the CAD’s whose roles are unambiguously established, it is of interest that their overall structures, including the location and number of the secondary structural elements are somewhat different to horse liver ADH, i.e. only four of the 12 amino acids are conserved. This is also a reflection of the similarities and differences between CAD and horse liver ADH with respect to their substrate-binding pocket, i.e. in terms of preferentially utilizing aromatic and aliphatic substrates, respectively. Nevertheless, the horse liver ADH coordinates were such as to enable provisional modelling of the tobacco CAD\(^{34}\) (discussed in Lewis *et al.*\(^{3}\)) as were the coordinates of the *S. solfataricus* ADH\(^{30}\) for AtCAD5 and AtCAD4.\(^{14}\) Finally, the relatively tight substrate binding pocket observed for the various hydroxycinnamyl aldehydes 1-5 examined using AtCAD5 and AtCAD4 now offers the opportunity to biotechnologically engineer the pocket to more readily utilize one substrate over another, e.g. to more efficiently channel metabolic flux, for example, for formation of health-related lignans.

### 2.4 EXPERIMENTAL

**Materials**
Coniferyl (3), sinapyl (5) and cinnamyl (6) aldehydes were purchased from Aldrich whereas p-coumaryl (1), caffeyl (2), 5-hydroxyconiferyl (4) aldehydes and p-coumaryl (7), caffeyl (8), coniferyl (9), 5-hydroxyconiferyl (10) and sinapyl (11) alcohols were synthesized as described in Kim et al.14

Expression and purification of AtCAD5 and AtCAD4

Plasmids (pET151D-TOPO® TA vector, Invitrogen) harboring AtCAD5 (GenBank accession number AY302082) and AtCAD4 (GenBank accession number AY302081) were used to individually transform ER2566 Escherichia coli cells (New England BioLabs Inc.), with the cells grown at 37 °C with shaking (250 rpm.) in Luria-Bertani (LB) medium supplemented with ampicillin (100 µg cm⁻³). Over-expression of AtCAD5 (or AtCAD4) was induced by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM at mid-log phase (A₆₀₀ = 0.5). After agitation (250 rpm) for 20 h at 20 °C, the cells were individually harvested by centrifugation (3,000 × g for 20 min). The cell pellets were individually resuspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0, 10% glycerol, 2 mM β-mercaptoethanol), and lyzed by sonication (5 × 10 s, model 450 sonifier®, Branson Ultrasonics Co.), with the lysates cleared by centrifugation (20,000 × g for 40 min). To each supernatant, Ni-nitrilotriacetic acid agarose (Qiagen, Hilden, Germany) was then added, with the slurry placed on a rocking shaker at 4 °C for 1 h. After eluting the unbound proteins and washing the resin ten times with washing buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0), the AtCAD5 (or AtCAD4) was then eluted stepwise with elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 100–300 mM imidazole pH 8.0) at a 50mM increment of imidazole concentration.
Thereafter, AtCAD5 (or AtCAD4)-enriched fractions (150-250 mM imidazole) were pooled, concentrated and the buffer exchanged to 5 mM Na phosphate (pH 6.8) using Amicon YM10 membrane (Millipore). This concentrated sample was passed over a CHT-10 hydroxyapatite column (BioRad, 1 × 10 cm), which was pre-equilibrated in Na phosphate buffer (5 mM, pH 6.8), at a flow rate of 3.0 cm⁻³ min. Each column was eluted with a linear gradient of Na phosphate (from 5 to 500 mM in 200 cm³). AtCAD5 and AtCAD4 were individually eluted at 50 and 100 mM Na phosphate, respectively. Fractions containing AtCAD5 (or AtCAD4) were concentrated as described previously, with the buffer exchanged to 50 mM Tris-HCl (pH 8.0) containing EDTA (1 mM) and dithiothreitol (1 mM) (Buffer A). The resulting AtCAD5 (or AtCAD4)-enriched protein fraction was loaded onto a MonoQ™ GL10/100 anion exchange column (Amersham Biosciences) equilibrated in Buffer A at a flow rate of 2 cm⁻³ min, with proteins eluted with a NaCl step gradient (0.05, 0.1, 0.2, 0.4, and 2 M; 20 cm³ for each step); the catalytically active AtCAD5 (or AtCAD4) fractions were eluted at 0.1 M NaCl. Confirmation of the presence and purity of the AtCAD5 (or AtCAD4) was made by SDS-PAGE.

**Size exclusion chromatography and multangle laser light scattering**

Determination of the molecular mass of AtCAD5 (and AtCAD4) in solution, using a static light scattering device, was performed as described previously in Youn et al.¹⁰ by loading a solution of AtCAD5 or AtCAD4 (100 µl, 2 mg cm⁻³ in PBS) onto a KW-803 column (8 × 300 mm, Shodex, Japan) pre-equilibrated in PBS buffer.

**Dynamic light scattering**

The radius and molecular weight of AtCAD5 (and AtCAD4) were estimated using a DynaPro-Titan (Wyatt Technology Corp.) instrument at 22 °C. Purified AtCAD5 or AtCAD4 (2 mg cm⁻³) in a freshly prepared Tris-HCl buffer (20 mM, pH 8.0), containing 1 mM EDTA and 1
mM dithiothreitol, were filtered through polyvinylidene difluoride filter (0.1 μm, Millipore). Scattering data were acquired through accumulation (5 times) of 10 scans with 10 s/scan, with the laser intensity set to a range of 50–60% (30-36 mW). The corresponding molecular weight and radius were calculated using the software package ‘DYNAMICS V6’ supplied with the instrument.

**Crystallization of AtCAD5**

AtCAD5 was crystallized in a range of enzyme concentrations from 4 to 10 mg cm⁻³ in 20 mM Tris-HCl buffer (pH 8.0) containing 1 mM EDTA and 1 mM dithiothreitol using the hanging drop vapor diffusion method at two different temperatures of 277 °K and 293 °K. For crystallization, apo-AtCAD5 was diluted two-fold with the reservoir solution [20% (w/v) PEG 3350 and 0.2 M tri-lithium citrate tetrahydrate, pH 8.1] for a final hanging drop volume of 3 mm³. Diamond-shaped crystals usually appeared after 3 days, and the best diffracting crystals with dimensions of ~ 0.2 × 0.2 × 0.8 mm were obtained in about 2 weeks. Apo-AtCAD5 crystallized in a tetragonal space group, P4₁2₁2 (a = b = 54.22, c = 312.29 Å), with one molecule in an asymmetric unit. For binary complex crystals, purified AtCAD5 (10mg cm⁻³) in 30 mM Tris-HCl buffer (pH 8.5) containing 100 mM NaCl, 2 mM dithiothreitol and 1.5 mM NADP⁺ was used. The binary AtCAD5 crystals were obtained by mixing the above solution of AtCAD5 (1.5 mm³) with an equal volume of a reservoir solution containing 20% (w/v) PEG 3350 and 0.2 M di-ammonium tartrate (pH 6.6). The binary complex was also crystallized in the same tetragonal space group, P4₁2₁2, with corresponding unit cells of a = b = 54.71, c = 303.93 Å. Diffraction data for the apo-form at 2.0 Å resolution, and the binary complex at 2.6 Å resolution, were collected using an ADSC Q210 CCD detector in the beam line 8.2.1 at Berkeley Advanced Light Source (ALS) and a Rigaku Saturn 92 CCD detector/MicroMax-007 X-ray generator,
respectively. All data were collected at a temperature of 100 °K. Before crystal freezing, the corresponding crystals were soaked for 5 min in cryoprotectant (25 % glycerol in the corresponding reservoir solution).

**Structural solution and refinement**

In order to obtain initial phase information of apo-AtCAD5 intensity data, the coordinates of the *E. coli* alcohol dehydrogenase-like protein yahK (1UUF) and the software package, AmoRe,\(^\text{42}\) were used for molecular replacement. Rigid body refinement was carried out with X-PLOR,\(^\text{43}\) beginning with the best solution of molecular replacement. The initial crystallographic R-value of a solution was 40.0%, with data resolution ranging from 15.0 to 3.0 Å. After several cycles of positional and temperature factor refinements, and a series of simulated annealing omit maps, a reasonable quality electron density map was achieved and all residues were fitted. The structure of the binary complex of AtCAD5 was solved by the molecular replacement method using the refined coordinates of apo-AtCAD5. The final R-factors for the apo-form and the binary complex were 19.5% (\(R_{\text{free}} = 23.5\%\) for the random 5% data) and 20.1 % (\(R_{\text{free}} = 23.0\%\) for the random 5% data), respectively (Table 2.1). The number of reflections above 2\(\sigma\) for the apo-form were 25,958 (98.2% completeness) between 10.0 and 2.0 Å resolution. The crystals of the binary complex did not diffract as well as the apo-form and gave reflection numbers of 9,631 (above 2\(\sigma\), 93.2% completeness) between 10.0 and 2.6 Å resolution.

The root mean square deviations (rmsd) (from ideal geometry) of the final coordinates corresponding to the apo-form and the binary complex are 0.01 and 0.02 Å for bonds and 2.6 and 2.9° for angles, respectively. All AtCAD5 coordinates have been deposited in the Protein Data Bank (apo-form: 2CF5, binary complex: 2CF6).
Table 2.1. Crystallographic data for AtCAD5

<table>
<thead>
<tr>
<th></th>
<th>Apo-from</th>
<th>Binary complex$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beam line</td>
<td>ALS 8.2.1</td>
<td>WSU (MM007)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.07812</td>
<td>1.5418</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50 to 2.0</td>
<td>50 to 2.6</td>
</tr>
<tr>
<td>Space group</td>
<td>P4$_1$2$_1$2</td>
<td>P4$_1$2$_1$2</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a = 54.22</td>
<td>a = 54.71</td>
</tr>
<tr>
<td></td>
<td>b = 54.22</td>
<td>b = 54.71</td>
</tr>
<tr>
<td></td>
<td>c = 312.99</td>
<td>c = 303.93</td>
</tr>
<tr>
<td>Asymmetric unit</td>
<td>1 molecule</td>
<td>1 molecule</td>
</tr>
<tr>
<td>Total observations</td>
<td>129,883</td>
<td>59,808</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>32,473</td>
<td>14,177</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.2</td>
<td>93.2</td>
</tr>
<tr>
<td>$R_{\text{sym}}$ (%)$^{b,c}$</td>
<td>5.8 (13.3)</td>
<td>6.5 (13.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>10 to 2.0</td>
<td>10 to 2.6</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>25,958</td>
<td>9,631</td>
</tr>
<tr>
<td>$R_{\text{cryst}}$ (%)$^d$</td>
<td>19.5</td>
<td>20.1</td>
</tr>
<tr>
<td>$R_{\text{free}}$ (%)$^e$</td>
<td>23.5</td>
<td>23.0</td>
</tr>
<tr>
<td>rmsd bonds (Å)</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td>rmsd angles (º)</td>
<td>3.019</td>
<td>3.463</td>
</tr>
<tr>
<td><strong>Number of atoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein &amp; ion &amp; ligand</td>
<td>2667</td>
<td>2715</td>
</tr>
<tr>
<td>Water</td>
<td>199</td>
<td>140</td>
</tr>
</tbody>
</table>

$^a$ AtCAD5 + NADP$^+$.  
$^b$ Numbers in parentheses refer to the highest shell.  
$^c$ $R_{\text{sym}} = \Sigma|I_h - <I_h>|/\Sigma I_h$, where $<I_h>$ is the average intensity over symmetry equivalent reflections.  
$^d$ $R_{\text{cryst}} = \Sigma|F_{\text{obs}} - F_{\text{calc}}|/\Sigma F_{\text{obs}}$, where summation is over the data used for refinement.  
$^e$ $R_{\text{free}}$ was calculated as for $R_{\text{cryst}}$ using 5% of the data that was excluded from refinement.  
$^f$ Root mean square deviations.
In turn, amino acid substitutions, insertions, and deletions for the AtCAD4 were next performed by using the graphics program O,\textsuperscript{44} starting from the refined coordinates of AtCAD5, followed by quick-energy minimization by using X-PLOR\textsuperscript{43} with potential function parameters of CHARMM19 as described previously.\textsuperscript{14} The initial position of substrate was obtained through the solid docking module on QUANTA (BioSYM/Micron Separations), which is based on conformational space, followed by a quick energy minimization by X-PLOR.\textsuperscript{43}

**Site-directed mutagenesis**

Forward (5′-CATGGTTTCTGGGCATGC*G*GTGGTAGGGG AAGTAG-3′) and reverse (5′-CTACTTCCCCTACCACC*G*C ATGCCCAGGAACCATG-3′) primers were designed and synthesized (Invitrogen, Carlsbad, USA) to convert AtCAD5 Glu70 into alanine (asterisks indicate mismatch for the Glu\(\rightarrow\)Ala substitution). Site specific mutagenesis was carried out with a QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, USA) following the manufacturer’s instructions with PCR conditions as follows: initial denaturation at 95 °C for 1 min, followed by 18 cycles of 95 °C for 50 s, 60 °C for 50 s and 68 °C for 6 min, with 7 min at 68 °C and an indefinite hold at 4 °C. After PCR completion, the PCR product was treated with \textit{Dpn} I for 1 h at 30 °C to digest the non-mutated parental DNA template, transformed into \textit{E. coli} TOP10 cells. After selection on LB plates containing 100 \(\mu\)g cm\(^{-3}\) carbenicillin, a positive clone, containing the mutation Glu70 \(\rightarrow\) Ala70, was confirmed by sequencing both strands (using pTrcHis2 forward and reverse primers (Invitrogen)) to ensure that there were no other mutation(s) in the open reading frame as a result of PCR. Heterologous expression of the Glu70 \(\rightarrow\) Ala70 mutant as well as recombinant AtCAD5 was carried out as described above. Both were purified and assayed using aldehydes 1-5 as described in Kim \textit{et al.}\textsuperscript{14}
2.5 ACKNOWLEDGMENTS

This research was supported in part by the National Institutes of Health (GM66173), the National Science Foundation (MCB-0417291), the United States Department of Energy (DE-FG-0397ER20259), the G. Thomas and Anita Hargrove Center for Plant Genomic Research, McIntire-Stennis, and the Murdock Charitable Trust. We thank D.J. Pouchnik for DNA sequencing.
2.6 REFERENCES


CHAPTER THREE

ASSESSMENT OF A PUTATIVE PROTON RELAY IN Arabidopsis CINNAMYL ALCOHOL DEHYDROGENASE CATALYSIS

Choonseok Lee, Diana L. Bedgar, Laurence B. Davin, and Norman G. Lewis

Abstract

Extended proton relay systems have been proposed for various alcohol dehydrogenases, including the Arabidopsis thaliana cinnamyl alcohol dehydrogenases (AtCADs). Following a previous structural biology investigation of AtCAD5, the potential roles of three amino acid residues in a putative proton relay system, namely Thr49, His52 and Asp57, in AtCAD5, were investigated herein. Using site-directed mutagenesis, kinetic and isothermal titration calorimetry (ITC) analyses, it was established that the Thr49 residue was essential for overall catalytic conversion, whereas His52 and Asp57 residues were not. Mutation of the Thr49 residue to Ala resulted in near abolition of catalysis, with thermodynamic data indicating a negative enthalpic change ($\Delta H$), as well as a significant decrease in binding affinity with NADPH, in contrast to wild type AtCAD5. Mutation of His52 and Asp57 residues by Ala did not significantly change either catalytic efficiency or thermodynamic parameters. Therefore, only the Thr49 residue is demonstrably essential for catalytic function. ITC analyses also suggested that for mechanism AtCAD5 catalysis, NADPH was bound first followed by $p$-coumaryl aldehyde.

3.1 INTRODUCTION
Cinnamyl alcohol dehydrogenases (CAD, EC 1.1.1.195) from vascular plants can catalyze *in vitro* the reversible conversion of *p*-coumaryl (1), caffeyl (2), coniferyl (3), 5-hydroxyconiferyl (4) and sinapyl (5) aldehydes into the corresponding alcohols (6–10) (Figure 3.1), with the forward reaction being favoured.\(^1\) In *Arabidopsis thaliana*, the first plant species whose genome was sequenced,\(^2\) 17 genes were provisionally annotated as CADs by The Arabidopsis Information Resource (TAIR).\(^3\) Of the 9 putative *Arabidopsis thaliana* CAD genes with highest homology to *bona fide* *Nicotiana tabacum* and *Pinus taeda* CADs,\(^1,3\) only AtCAD4 and AtCAD5 proteins were CADs proper based on both *in vitro* highest catalytic activity\(^1\) and *in vivo* studies using the *A. thaliana* double mutant cad4/cad5.\(^4,5\) The remaining 8 putative CADs had very low homology (0.9 ~ 1.6% similarity) to the *N. tabacum* and *P. taeda* CADs, and these proteins also lacked catalytic zinc.\(^1\)

**Figure 3.1.** Enzymatic reactions catalysed by AtCAD5 in presence of NADPH
Cinnamyl alcohol dehydrogenases belong to the medium-chain dehydrogenase/reductase family,\textsuperscript{6} and a structural biology study of the AtCAD5 apo-enzyme and its binary complex with NADP\textsuperscript{+} was reported by Youn \textit{et al.}\textsuperscript{7} In terms of the putative CAD mechanism, there are currently two possible mechanisms for related alcohol dehydrogenases, using either an extended proton relay (transfer)\textsuperscript{8} or a penta-coordinated zinc system.\textsuperscript{9} The extended proton relay mechanism assumes that a water molecule bound to tetrahedrally coordinated catalytic zinc is displaced by an incoming alcohol substrate and that this species is maintained during catalysis.\textsuperscript{8}

That proposed mechanism was based on the study of the ternary complex of horse liver alcohol dehydrogenase (HLADH), crystallized in the presence of NAD\textsuperscript{+} and \textit{p}-bromobenzyl alcohol, where it was reported that the catalytic zinc ion was ligated to the alcohol oxygen \textit{via} a tetrahedrally coordinated complex (Figure 3.2a).\textsuperscript{8, 10-13} In addition, an extended hydrogen bonding system was reported between the hydroxyl group of the substrate alcohol coordinated to zinc ion and Ser48, between Ser48 and nicotinamide ribose (O2'), and between nicotinamide ribose (O2') and His51 in the ternary complex of HLADH, whereas in the apoenzyme the hydrogen bond system was not observed.\textsuperscript{8} In contrast to HLADH, however, a ternary complex with substrate and cofactor is not available for AtCAD5 thus far.\textsuperscript{7}

In the alternative model (Figure 3.2b), binding of the alcohol substrate putatively results in formation of a penta-coordinated zinc complex,\textsuperscript{9} as deduced from analysis of results obtained with a \textit{Thermoanaerobacter brockii} alcohol dehydrogenase (TbADH) using extended X-ray absorption fine structure, pre-steady state kinetics, and density functional theory calculation approaches. In that study, two distinct penta-coordinate intermediates were postulated as being formed during the catalytic cycle. Specifically, the following was envisaged: When NAPD\textsuperscript{+} is
Figure 3.2. Two proposed catalytic mechanisms for alcohol dehydrogenases: (a) Proton relay system in horse liver alcohol dehydrogenase (HLADH); adapted and redrawn from Eklund et al.\textsuperscript{8} (b) Penta-coordinated zinc system in \textit{Thermoanaerobacter brockii} alcohol dehydrogenase (TbADH); adapted and redrawn from Kleifeld \textit{et al.}\textsuperscript{9} \textbf{TR1}: the first transient complex, \textbf{TR2}: the second transient complex.

Bound to TbADH, the tetra-coordinated zinc ion species is maintained, and the first penta-coordinated zinc ion transient complex is formed by addition of a water molecule. The penta-coordinated zinc ion species is then converted back to a tetra-coordinated zinc ion complex by dissociation of the ligated Glu60 residue. Binding of the incoming alcohol substrate regenerates a penta-coordinated zinc ion complex, the second transient species. The catalytic cycle is envisaged to be completed through dissociation of the water molecule and product, and re-
ligation of Glu60 to regenerate the original tetra-coordinated zinc ion species, i.e. returning the enzyme to its resting state. Site-directed mutagenesis however, gave a Glu60Ala mutant whose catalytic efficiency was only reduced circa four fold relative to the wild type enzyme. Overall, this mechanism (Figure 3.2b) does not require an extended proton relay transfer as for the one described for HLADH (Figure 3.2a).

In terms of developing an understanding of overall CAD catalysis, the first report was that describing molecular modelling of *Eucalyptus gunnii* CAD2 against HLADH. This study suggested that its putative substrate binding region in the active site included residues Ile95, Tyr113, Trp119 and Phe298 substituted from the corresponding residues Phe93, Phe110, Leu116 and Ile318 of HLADH as reviewed in Lewis et al. Both Trp119 and Phe298 residues were proposed as providing a “molecular sandwich”, whereby the phenolic ring of the substrate is stabilized, and in an orientation enabling abstraction of the pro- R hydride from NADPH during the reductive step. In the reverse reaction, the 9-pro- R hydrogen of coniferyl alcohol (8) is removed. In addition, Ser212/Arg217 residues in *E. gunnii* CAD2, and Asp223/Lys228 in HLADH, were predicted to be involved in cofactor binding in this model, with the Ser212 residue further confirmed to be closely related to NADPH binding through its site-directed mutant, Ser212Asp.

The putative three-dimensional molecular structure of *Eucalyptus gunnii* CAD2 also suggested that both residues, Ser49 and His52, served in a proton relay system in a comparable way as for Ser48 and His51 residues in HLADH. In a somewhat analogous manner, following analysis of the binary complex of AtCAD5, the corresponding residues, Thr49 and His52, were also putatively involved in a proton relay, together with Asp57 (Figure 3.3), with the latter being postulated via substrate modelling of the binary complex of AtCAD5.
Isothermal titration calorimetry (ITC) is now being widely used to investigate protein-protein, protein-nucleic acid and protein-lipid interactions, drug design and enzyme kinetics. In particular, this approach is very effective for the study of how proteins interact with either substrates, cofactors or ligands through analysis of binding stoichiometry ($N$), binding enthalpy ($\Delta H$), dissociation constant $K_d$, binding entropy ($\Delta S$) and Gibbs free energy of binding ($\Delta G$). Furthermore, for site-directed mutated proteins, ITC can effectively be utilized to compare the behaviour of native (recombinant) proteins and their mutated forms.

In the investigation herein, site-directed mutagenesis and ITC were used to further evaluate and assess the validity of a putative extended proton relay mechanism in AtCAD5, including study of the effects of mutating putative key amino acid residues on substrate and cofactor binding, as well as catalytic turnover. Additionally, this study examined the order of substrate/co-factor binding during AtCAD5 catalysis as an ordered bi-bi mechanism had been
suggested by Wyrambik and Grisebach.  

### 3.2 RESULTS AND DISCUSSION

Residues Thr49, His52 and Asp57 of AtCAD5 were individually mutated into Ala through PCR (see Table 3.1 for each primer set used). It was envisioned that these mutations would establish which of the three amino acid residues were potentially involved in a proton relay (Figure 3.3). After removal of parental methylated and hemi-methylated DNA with *Dpn I* restriction enzyme, each mutated DNA was individually transformed into One shot® Top 10 *Escherichia coli* cells. After selecting each mutated clone, mutations at each position were confirmed by sequencing, with each construct next transformed into *E. coli* cells for protein expression. Except for use of the Fast Protein Liquid Chromatography system, the purification method employed was as described in Kim *et al.* Each recombinant protein was obtained from an imidazole-containing eluent (from ~120 to 180 mM in 20 mM Tris-HCl, 500 mM NaCl, pH 7.9) followed by concentration/buffer exchange in 20 mM Tris buffer (pH 7.5). Each enzyme was purified to apparent homogeneity (data not shown) and investigated to obtain kinetic and thermodynamic parameters.

**Table 3.1.** Sequences of primers used for site-directed mutagenesis

<table>
<thead>
<tr>
<th>Primers</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCAD5 T49A F</td>
<td>GCTGTGGAATCTGCCACGCGGATCTTCATCAAAGCTA</td>
</tr>
<tr>
<td>AtCAD5 T49A R</td>
<td>TAGTTGATGAAGATCCGCGTGGCAGATTCTACAACTA</td>
</tr>
<tr>
<td>AtCAD5 H52A F</td>
<td>GAATCTGCCACACCAGATCTTCGCAAAGCTAATCGAGTTCTTA</td>
</tr>
<tr>
<td>AtCAD5 H52A R</td>
<td>CAAGATCCTTTTTAGTTGCGCAAGATCGGTGTTCCAGATCC</td>
</tr>
<tr>
<td>AtCAD5 D57A F</td>
<td>CAAGATCCTTTTTAGTTGCGCAAGATCGGTGTTCCAGATCC</td>
</tr>
<tr>
<td>AtCAD5 D57A R</td>
<td>AGACATGCGACGCCATTTTTAGTTG</td>
</tr>
</tbody>
</table>

**Kinetic analyses of AtCAD5 and its site-directed mutants**
Kinetic parameters for AtCAD5 and its site-directed mutants were initially obtained using the Michaelis-Menten equation\textsuperscript{31} with \( p \)-coumaryl aldehyde (1) employed as a substrate (Table 3.2).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>( K_m ) (( \mu )M)</th>
<th>( V_{\text{max}} ) (pkat ( \mu )g(^{-1}))</th>
<th>( k_{\text{cat}} ) (s(^{-1}))</th>
<th>( k_{\text{cat}}/K_m ) (M(^{-1}) s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCAD5</td>
<td>13 ± 1.9</td>
<td>88.1 ± 3.5</td>
<td>6.8 ± 0.3</td>
<td>523,000 (± 46,000)</td>
</tr>
<tr>
<td>T49A</td>
<td>114 ± 11.3</td>
<td>0.33 ± 0.01</td>
<td>0.026 ± 0.001</td>
<td>230 (± 18)</td>
</tr>
<tr>
<td>H52A</td>
<td>117 ± 6.9</td>
<td>518.9 ± 16.5</td>
<td>40.2 ± 1.3</td>
<td>345,000 (± 8,700)</td>
</tr>
<tr>
<td>D57A</td>
<td>36 ± 3.1</td>
<td>957.7 ± 31.0</td>
<td>74.2 ± 2.4</td>
<td>2,061,000 (± 102,000)</td>
</tr>
</tbody>
</table>

In our hands, the native (recombinant) AtCAD5 gave a \( K_m \) 13 \( \mu \)M, \( V_{\text{max}} \) 88.1 pkat \( \mu \)g\(^{-1}\), and catalytic turnover of 6.8 (s\(^{-1}\)). That is, the substrate, \( p \)-coumaryl aldehyde (1), was bound very tightly and was efficiently converted into the product, \( p \)-coumaryl alcohol (6). These data were thus in good agreement with our previous study of AtCADs.\textsuperscript{1}

As indicated earlier, in our previous study of the binary complex with AtCAD5, Thr49 was proposed to be hydrogen-bonded to the O2′ of the nicotinamide ribose moiety and thus was considered a potential residue in the putative proton transfer mechanism.\textsuperscript{7} Participation of this residue in the overall enzymatic reaction was established, \textit{via} analysis of the Thr49Ala mutant which was found to be essentially catalytically inactive. The affinity of the substrate, \( p \)-coumaryl aldehyde (1), was though still quite high (\( K_m \) 114 \( \mu \)M), whereas the \( V_{\text{max}} \) was 0.33 pkat \( \mu \)g\(^{-1}\) and the catalytic turnover number was 0.026 (s\(^{-1}\)). Thus, the Thr49Ala mutation specifically caused a near complete loss in catalytic efficiency.

On the other hand, kinetic analysis of the AtCAD5 H52A recombinant protein gave very
different results. The data obtained established that this residue was not essential for catalysis overall, even though it was provisionally envisioned earlier as being involved in the proton relay via hydrogen bonding to O3 of the nicotinamide ribose. In our hands, substrate affinity was still relatively high \((K_m = 117 \mu M)\) for \(p\)-coumaryl aldehyde (1), the \(V_{\text{max}}\) increased \(~6\) fold \((518.9 \text{ pkat} \mu \text{g}^{-1})\) with a catalytic turnover of \(40.2 (\text{s}^{-1})\). In an analogous manner, the D57A mutant gave a catalytically active enzyme with \(K_m = 36 \mu M\) for \(p\)-coumaryl aldehyde (1). However, the \(V_{\text{max}}\) also apparently increased \(~11\) fold with the overall catalytic turnover larger by \(~4\) fold over the native (recombinant) protein. Therefore, the kinetic analysis of AtCAD5 D57A demonstrated that the Asp57 residue was also not essential for overall catalysis in AtCAD5. Taken together, only the Thr49 residue can be considered essential for the overall catalytic mechanism of AtCAD5.

There is an earlier precedent for this observation with the \(\beta_1\beta_1\) isoenzyme of human liver alcohol dehydrogenase (HuLADH).\(^{32}\) Its site-directed mutant His51Gln (corresponding to His52 in AtCAD5) also had 2 (at pH 7, 8 and 9) or 3 (at pH 10) times higher \(V_{\text{max}}\) in the mutant, relative to wild type enzyme, this being rationalized as due to less tight binding of NADH thereby enabling it to be more readily dissociated.\(^{32-34}\) Accordingly, the mutation of His52 or Asp57 may facilitate release of NADP\(^+\) during catalysis.

**Isothermal titration calorimetry analyses of AtCAD5 and its site-directed mutants**

ITC measurements at constant temperature and pressure were carried out to further investigate the binding properties of AtCAD5 and its site-directed mutants using either \(p\)-coumaryl aldehyde (1) or NADPH. Thermodynamic parameters which included dissociation constant \(K_d (K_d = 1/K_a; \ K_a, \ \text{binding association constant})\) and enthalpy of binding \((\Delta H)\) were thus obtained. The Gibbs free energy of binding was calculated from \(\Delta G = -RT \ln K_a\), where \(R\) is the
gas constant and T the temperature. The binding entropy was obtained from \( \Delta S = (\Delta H - \Delta G)/T \).^{19}

Binding of \( p \)-coumaryl aldehyde (I) to AtCAD5 and its mutants, however, yielded inconclusive results due to very low heat generation for all enzymes tested, suggesting that \( p \)-coumaryl aldehyde (I) either binds very weakly or cannot bind to the free enzyme (data not shown).

In contrast, significant heat changes were observed upon binding of NADPH to AtCAD5 or its mutants (Figure 3.4). Specifically, the NADPH binding data obtained were fitted best to a model that specifies one set of sites, with \( N \) values from 0.7 to 0.8 obtained (Table 3.3), these being indicative of ~1:1 stoichiometry per enzyme dimer. These data thus indicate that one molecule of NADPH is bound to AtCAD5, a homodimer,\(^7\) or its site-directed mutants. Valencia \textit{et al.} also reported that a putative CAD-like homolog from \textit{Saccharomyces cerevisiae} binds one NADPH molecule per homodimer, this being rationalized to only one functional subunit probably being available for catalysis.\(^35\) Similarly, \textit{Pseudomonas putida} tartrate dehydrogenase (TDH), a homodimer, also reportedly showed a 1:1 stoichiometry of NADH binding per enzyme dimer.\(^36\)

<table>
<thead>
<tr>
<th>Protein</th>
<th>( N^a )</th>
<th>( K_d ) (( \mu )M)</th>
<th>( \Delta G ) (kJ mol(^{-1}))</th>
<th>( \Delta H ) (kJ mol(^{-1}))</th>
<th>( T\Delta S ) (kJ mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtCAD5</td>
<td>0.7 (± 0.02)</td>
<td>3.5 (± 0.1)</td>
<td>−31.7 (± 0.1)</td>
<td>26.0 (± 0.3)</td>
<td>57.7 (± 0.3)</td>
</tr>
<tr>
<td>AtCAD5 T49A</td>
<td>0.8 (± 0.002)</td>
<td>37.2 (± 11.6)</td>
<td>−25.8 (± 0.9)</td>
<td>−12.0 (± 5.7)</td>
<td>13.8 (± 6.3)</td>
</tr>
<tr>
<td>AtCAD5 H52A</td>
<td>0.8 (± 0.02)</td>
<td>7.1 (± 1.1)</td>
<td>−29.9 (± 0.4)</td>
<td>17.1 (± 0.7)</td>
<td>47.0 (± 0.3)</td>
</tr>
<tr>
<td>AtCAD5 D57A</td>
<td>0.8 (± 0.02)</td>
<td>8.2 (± 0.5)</td>
<td>−29.5 (± 0.1)</td>
<td>15.6 (± 0.2)</td>
<td>45.1 (± 0.1)</td>
</tr>
</tbody>
</table>

\( ^a \) \( N \) indicates binding stoichiometry of NADPH binding per enzyme homodimer.
Figure 3.4. Isothermal titration calorimetry analyses of AtCAD5 (a), AtCAD5 T49A (b), AtCAD5 H52A (c) and AtCAD5 D57A (d) with NADPH. The top panels show raw data of the heat pulses resulting from each titration, whereas the bottom panels show the integrated heat normalized per mole of NADPH as a function of the molar ratio (NADPH concentration/enzyme homodimer concentration).
The ITC data for binding of NADPH to AtCAD5 is summarized in Table 3.3, with a binding isotherm and the fitted data shown in Figure 3.4A. NADPH was found to bind to AtCAD5 with a dissociation constant of ~3.5 µM. Formation of the binary complex AtCAD5•NADPH is endothermic with a ΔH of 26.0 kJ mol\(^{-1}\) (Table 3.3, Figure 3.4a).

The essentially catalytically inactive T49A mutant was next examined (Table 3.3 and Figure 3.4b). The thermodynamics of NADPH binding were affected by the T49A mutation: the K\(_d\) of NADPH increased by >10 fold, indicating a significant decrease in binding affinity, with formation of the binary complex now being exothermic (ΔH of −12.0 kJ mol\(^{-1}\)) and driven by a less favourable −TΔS. As shown in the putative modelled ternary complex of AtCAD5 (Figure 3.3),\(^7\) the Thr49 is proposed to interact with the substrate and NAD(P)H and is also located near to the catalytic Zn\(^{2+}\). Replacement of Thr49 with Ala thus disrupts the interaction between this amino acid residue, catalytic Zn\(^{2+}\), and putatively NADPH and the substrate, the net effect being decreased NADPH binding affinity (Table 3.3) and a near complete abolition of catalytic activity (Table 3.2).

In agreement with the kinetic data, the thermodynamic properties of NADPH binding for H52A and D57A mutant proteins were very similar to that of WT (Table 3.3 and Figure 3.4c, d). Specifically, NADPH binding yielded a K\(_d\) of 7.1 µM, ΔH of 17.1 kJ mol\(^{-1}\) and TΔS of 47.0 kJ mol\(^{-1}\) for the H52A mutant, and K\(_d\) of 8.21 µM, ΔH of 15.6 kJ mol\(^{-1}\) and TΔS of 45.1 kJ mol\(^{-1}\) for the D57A mutant. Thus, this implies that mutation of His52 and Asp57 to alanine did not very significantly affect binding of NADPH to the corresponding protein in contrast to mutation of Thr49 to Ala.

In sum, NADPH binding to AtCAD5 and its mutants is a thermodynamically favourable processes as indicated by the negative ΔG values (between −25.8 and −31.7 kJ mol\(^{-1}\)).
Titration of the binary complex AtCAD5•NADPH with \( p \)-coumaryl aldehyde (1), followed by the enzymatic conversion into the corresponding alcohol (6), was next investigated at 25˚C. The overall process was found to be exothermic (Figure 3.5a), in contrast to formation of the binary complex AtCAD5•NADPH which was endothermic (Figure 3.4a). However, when NADPH was injected to the cell containing both AtCAD5 and \( p \)-coumaryl aldehyde (1), a very different profile was observed (Figure 3.5b). First, an endothermic reaction occurred corresponding to binding of NADPH to AtCAD followed by an exothermic response due to the catalytic conversion of 1 into 6. Together, these results further confirm that the cofactor binds first to the enzyme followed by the aldehyde substrate (1). Similar results were observed with the mutant enzymes (data not shown).

![Figure 3.5. Isothermal titration calorimetry (ITC) analysis of AtCAD5. (a) Titration at 25˚C of AtCAD5 and NADPH with \( p \)-coumaryl aldehyde (1). (b) Titration at 25˚C of AtCAD5 and \( p \)-coumaryl aldehyde (1) with NADPH. The top panel show raw data of the heat pulses resulting for each titration, whereas the bottom panels show the integrated heat normalized per mole of](image-url)
injectant as a function of the molar ratio (injectant concentration/enzyme homodimer concentration).

3.3 CONCLUSIONS

In the putative proton relay system for AtCAD5, three residues, Thr49, His52 and Asp57, were provisionally implicated in its catalysis as well as for other dehydrogenases. However, from the kinetic and ITC data of these site-directed mutations in AtCAD5, all three residues were apparently not essential for catalysis. Notably, only mutation of Thr49 residue caused a loss in catalytic activity for AtCAD5, this being further verified by a significant decrease in binding affinity, a negative $\Delta H$ from a positive $\Delta H$ in AtCAD5 and a decrease of $T\Delta S$ in NADPH binding.

However, mutation of His52 and Asp57 residues did not affect catalytic efficiency, but, in contrast, the turnover numbers and maximum velocities of the His52Ala and Asp57Ala mutants increased. Moreover, binding properties of AtCAD5 H52A and D57A remained largely unchanged.

From the ITC analyses, it was also concluded that NADPH binds first to AtCAD5 followed by the substrate.

Taken together, the data provide no evidence for an extended proton relay system.

3.4 EXPERIMENTAL

Materials

$p$-Coumaryl (1) aldehyde and $p$-coumaryl alcohol (6) were synthesized as in Kim et al. NADPH (tetrasodium salt) was obtained from Sigma. Water (Optima® LC/MS) and acetonitrile
(Optima® LC/MS) were obtained from Fisher Scientific, and glacial AcOH (HPLC grade) was procured from Mallinckrodt Baker.

The QuikChange® XL Site-Directed Mutagenesis Kit was purchased from Stratagene, whereas pTrcHis2-TOPO vector and One Shot TOP 10 competent E. coli cells were from Invitrogen. BugBuster® protein extraction reagent, Benzonase® nuclease and rlysozyme™ solution, were obtained from Novagen. The POROS 20 metal chelate resin was from Applied BioSystems.

**Instrumentation**

Polymerase chain reactions (PCR) were carried out using a PTC-0220 DNA engine dyad Peltier thermal cycler (MJ Research). Recombinant protein purification was carried out on a Fast Protein Liquid Chromatography (FPLC, Amersham Pharmacia Biotech) system using a column packed with POROS 20 metal chelate resin.

Reversed-phase chromatography employed an Ultra Performance Liquid Chromatography® system (UPLC, Waters) equipped with a BEH shield RP 18 column (Waters); flow rate of 0.3 ml min⁻¹; detection at 280 nm. The solvent system consisted of a concave gradient (Waters curve #8) of CH₃CN:3% AcOH (v/v) in H₂O from 5:95 to 40:60 between 0 and 8.30 min.

Isothermal titration calorimetry measurements used a VP-ITC microcalorimeter (MicroCal Inc.).

**Site-directed Mutagenesis**

AtCAD5, cloned into the pTrcHis2-TOPO vector, was used as previously described.¹ Primers were designed and synthesized (Invitrogen) to individually change Thr49, His52 and
Asp57 into Ala (Table 3.1). Site-directed mutagenesis was carried out using a QuikChange® XL Site-Directed Mutagenesis Kit following the manufacturer’s instructions with PCR performed as follows: 95 °C for 1 min, 18 cycles at 95 °C for 50 s, 60 °C for 50 s, 68 °C for 4 min, and 68 °C for 7 min. Parental plasmid DNA was next digested with Dpn I restriction enzyme (10 units), QuikChange® XL Site-Directed Mutagenesis Kit) at 37 °C for 1 h, with the resulting mutated plasmid DNA individually transformed into One Shot TOP 10 competent E. coli cells. Transformants were selected on LB medium containing carbenicillin (100 μg ml⁻¹). Positive clones were next subjected to sequencing to confirm single mutations.

**Expression and purification of AtCAD5 and its mutants**

AtCAD5 and its mutants were individually inoculated in 5 ml of LB medium containing carbenicillin (100 μg ml⁻¹) and incubated at 37 °C overnight. Next, 250 ml of LB medium contain carbenicillin (100 μg ml⁻¹) was inoculated with 2.5 ml of the corresponding 5 ml culture. The E. coli cells were incubated at 37 °C at 230 rpm until an OD₆₀₀ ~0.8 was reached, with induction then initiated by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.2 mM final concentration). After incubation at 20 °C for 24 h, the cells were harvested by centrifugation, and stored at −20 °C until needed.

Frozen pellets were individually thawed, resuspended in 10 ml of BugBuster protein extraction reagent containing Benzonase® nuclease (250 units) and rLysozyme™ solution (10,000 units) at room temperature for 20 min. Each lysed solution was centrifuged (4,300 × g, 25 min), with the supernatant filtered through a 0.2 μm syringe filter (Pall).

Recombinant protein purification was carried out at 4 °C on a POROS 20 metal chelate column (60 × 10 mm) pre-equilibrated in binding buffer (20 mM Tris-HCl, pH 7.9, 500 mM
NaCl, 5 mM imidazole) at a flow rate of 4 ml min\(^{-1}\). Aliquots (~10 ml) of each recombinant protein preparations were next loaded onto the affinity column, with the latter first washed with 10% elution buffer (20 mM Tris-HCl, pH 7.9, 500 mM NaCl, 500 mM imidazole) in binding buffer, after which a gradient from 10 to 100% was applied in 25 min. Fractions eluting between ~120 and ~180 mM imidazole were combined and concentrated with Centricon Plus-70 (Amicon), diluted with 20 mM Tris-HCl buffer (pH 7.5), and concentrated to 10 – 20 mg ml\(^{-1}\) (Centricon Plus-70, Amicon). Final purity was confirmed by SDS-PAGE (4 – 15% acrylamide, BioRad) with silver staining (Amersham Bioscience Application Note). Protein concentrations were determined by the Bradford method\(^{39}\) using \(\gamma\)-globulin as standard.

**Assay of AtCAD5 and its mutants**

Standard assays were as described in Kim et al.\(^1\) with initial velocity kinetics data carried out in presence of 50 ng ml\(^{-1}\) AtCAD5, 105 ng ml\(^{-1}\) AtCAD5 H52A, 50 ng ml\(^{-1}\) AtCAD5 D57A, and 14.3 \(\mu\)g ml\(^{-1}\) AtCAD5 T49A. Final \(p\)-coumaryl aldehyde (I) concentrations were: 0.76 – 80 \(\mu\)M for AtCAD5, 0.76 – 200 \(\mu\)M for AtCAD5 H52A and D57A, and finally, 7.6 – 480 \(\mu\)M for AtCAD5 T49A, respectively. Assays, carried out in quadruplicate, were initiated by enzyme addition, incubated at 30 °C for 1 min (2 min for AtCAD5 T49A) and stopped by addition of glacial AcOH (10 \(\mu\)l). Aliquots (10 \(\mu\)l) from each assay mixture were subjected to reversed phase UPLC as described in the Instrumentation section. For each enzyme preparation, kinetic data was calculated using Origin 7.5 (OriginLab).

**Isothermal titration calorimetry (ITC)**

Titrations were carried at constant pressure and at 30°C (for cofactor/substrate binding) or at 25°C (for order of binding study). Twenty nine injections (10 \(\mu\)l each) at 200 s intervals
were performed, with the first injection not used for data analysis. The stirring speed was 300 rpm.

For cofactor/substrate binding, the sample cell (1.4 ml) was filled with either AtCAD5 or its mutants at a final concentration of 50 µM (of homodimer) in HEPES buffer (20 mM, pH 6.8) containing 100 mM NaCl (Buffer A). The injection syringe (250 µl) contained NADPH or p-coumaryl aldehyde (1) (1 mM) in Buffer A. Titrations were carried out in triplicates. A control titration was carried out by injecting ligand in Buffer A in order to determine the dilution/mixing heat which was then subtracted for each experiment prior to data analysis.

To study the order of binding, p-coumaryl aldehyde (1, 1 mM) was titrated into a cell containing AtCAD5 (50 µM, homodimer) and NADPH (75 µM) in Buffer A. In another experiment, NADPH (1 mM) was titrated into a cell containing AtCAD5 (50 µM, homodimer) and p-coumaryl aldehyde (1, 75 µM). ITC analyses were then carried out as above.

The Origin 5.0 software (MicroCal) was utilized to obtain dissociation constant $K_d$, and binding enthalpy ($\Delta H$) for each.

3.5 ACKNOWLEDGEMENTS

This research was supported in part by a grant from the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences (DE-FG-0397ER20259), and the G. Thomas and Anita Hargrove Center for Plant Genomic Research. The latter support partially funded the graduate student stipend of C. Lee. The authors thanks Dr. Verna Frasca (G.E. Healthcare Life Sciences) for valuable discussions on ITC analyses.
3.6 REFERENCES


27. X. Li, D.-C. Chow and S.-C. Tu, Biochemistry, 2006, 45, 14781–14787.
CHAPTER FOUR

MECHANISTIC AND STRUCTURAL STUDIES OF APO-FORM, BINARY AND TERNARY COMPLEXES OF THE ARABIDOPSIS ALKENAL (DOUBLE BOND) REDUCTASE (At5g16970)


Copyright 2006

Abstract

In this study, we determined the crystal structures of the apo form, binary and ternary complexes of the Arabidopsis alkenal (double bond) reductase, encoded by At5g16970. This protein, one of 11 homologues in A. thaliana, is most closely related to the Pinus taeda phenylpropanal (double bond) reductase, involved in, for example, heartwood formation. Both enzymes also have essential roles in plant defense, and can function by catalyzing the reduction of the 7–8 double bond of phenylpropanal substrates, such as p-coumaryl and coniferyl aldehydes in vitro; At5g16970 is also capable of reducing toxic substrates with the same alkenal functionality, such as 4-hydroxy-2E-nonenal. The overall fold of At5g16970 is similar to that of the Zn-independent medium-chain dehydrogenase/ reductase superfamily, the members of which have two domains and are dimeric in nature, i.e. in contrast to their original classification as being Zn-containing oxidoreductases. As provisionally anticipated from the kinetic data, the
shape of the binding pocket can readily accommodate \( p \)-coumaryl aldehyde, coniferyl aldehyde, 4-hydroxy-2\(E\)-nonenal and 2-alkenals. However, the enzyme kinetic data among these potential substrates differ, favoring \( p \)-coumaryl aldehyde. Tyr260 is proposed to function as a general acid/base for hydride transfer. A catalytic mechanism for this reduction, and its applicability to related important detoxification mammalian proteins, is proposed.

4.1 INTRODUCTION

The phenylpropanoid pathway, which is essentially restricted to vascular plants, is of increasing significance and general importance due to the extensive medicinal/health protecting properties of many of its derivatives. For example, this pathway results in formation of the potent antiviral agent, podophyllotoxin\(^1\), which also serves as a semi-synthetic source of the highly successful cancer chemotherapeutic treatments, teniposide, etoposide and Etopophos\(^\text{®}\)\(^2-4\). Other examples include that of: matairesinol and secoisolariciresinol, which are some of the dietary sources of the “mammalian” lignans, enterolactone/enterodiol, these being protective against the onset of various malignancies\(^5\); the nordihydroguaiaretic acid derivatives, which show considerable promise against refractory cancers of the neck and the head\(^6\); the potent antioxidant, chlorogenic acid, which has well-documented anticancer properties\(^7,8\), as well as in reducing the risk of cardiovascular disease\(^9\); the monomeric allyl/prophenyl phenols, such as chavicol and eugenol, which have well-known antibacterial and analgesic properties\(^10\).

Other medicinally important phenylpropanoid (acetate) pathway metabolites include dihydroconiferyl alcohol (1; Figure 4.1), a potentially useful anti-inflammatory agent\(^11\), and phlorizin (5) which shows considerable promise for treatment of diabetes mellitus, obesity and stress hyperglycemia\(^12\).
Our recent studies have been directed towards establishing the various biochemical pathways associated with formation of such metabolites, including defining the catalytic mechanisms and high-resolution structures of the participating pathway enzymes. Examples of these include pinoresinol/lariciresinol reductases\textsuperscript{13, 14}, secoisolariciresinol dehydrogenase\textsuperscript{15-17}, phenylcoumaran benzylic ether reductase\textsuperscript{14, 18}, isoflavone reductase\textsuperscript{14}, cinnamyl alcohol dehydrogenases\textsuperscript{19, 20}, chavicol/p-anol and eugenol/isoeugenol synthases\textsuperscript{21, 22}, as well as dirigent proteins (in the presence of auxiliary oxidative capacity)\textsuperscript{23-25}. These studies are part of broader goals aimed towards (i) systematically engineering selected enzyme substrate binding pockets in terms of potentially modifying them to be more specific for a particular metabolite/metabolic pathway, and (ii) better understanding how these pathways in plants have evolved.

The objective of the study herein is to determine the mechanism and structures of the enzyme involved in formation of medicinally promising dihydrophenylpropanoid derivatives, such as dihydroconiferyl alcohol (\textbf{1}). In the Pinaceae, e.g. loblolly pine (\textit{Pinus taeda}), various dihydrophenylpropanoids accumulate as heartwood-forming constituents which contribute to the color, quality, and durability of its woody tissues; these can have either propanol, propionic acid or propanaldehyde side-chains [e.g. \textit{p}-dihydrocoumaric (\textbf{4})/dihydroferulic (\textbf{3}) acids and \textit{p}-dihydrocoumaryl (\textbf{2})/dihydro-coniferyl (\textbf{1}) alcohols (Figure 4.1) in \textit{Picea glauca}\textsuperscript{26}]. Interestingly, their amounts (e.g. \textbf{1} and \textbf{2}) are known to increase in the galls of \textit{P. glauca} upon aphid attack (e.g. by \textit{Adelges abietis})\textsuperscript{26}, in further support of roles in plant defense.
In 2001, we reported the discovery of a *P. taeda* phenylpropanal (α,β double bond) reductase (PtPPDBR)\(^1\) (see Figure 4.2A), whose encoding gene (see Figure 4.3) was cloned with the functionally recombinant protein obtained and preliminarily characterized\(^2\). This enzyme, which is a member of the Zn-independent medium chain dehydrogenase/reductase (MDR) superfamily, catalyzes the NADPH-dependent conversion of various monomeric and dimeric phenylpropanaldehyde (e.g. 6/7, Figure 4.2A) into the corresponding phenylpropanaldehydes (e.g. 8/9). In terms of its amino acid similarity/identity, PtPPDBR has the closest homology to the *Arabidopsis thaliana* At5g16970 (AtDBR1), as well as to a gene encoding (+)-pulegone reductase (PulR) from *Mentha piperita*\(^2\), i.e. with similarities/identities of 63/43 and 62/44\%, respectively (Fig. 4.3 and Table 4.1). In this study, we describe the characterization of the PtPPDBR homologue, At5g16970 (AtDBR1), which catalyzes the same conversions. By contrast, the (+)-pulegone reductase (PulR) from peppermint (*M. piperita*)\(^2\) was not investigated: it is specifically involved in a similar conversion of pulegone (10) to (+)-isomenthone (11) and (–)-menthone (12), with the latter predominating (Figure 4.2B).
Figure 4.2. Reactions catalyzed by NADPH-dependent alkenal/alkenone/\(\alpha,\beta\) double bond reductase in vitro. Reduction is shown of: A, \(p\)-coumaryl (6) and coniferyl (7) aldehydes by PtPPDBR and AtDBR1; B, pulegone (10) by PuR in Mentha piperita; C, various \(\alpha,\beta\) unsaturated ketones in prostaglandin biosynthesis (e.g. 13) by guinea pig leugotriene B\(_4\) 12-HD/PGR; D, 4-HNE (15) by AtDBR1.
In the meantime, in related mammalian systems, there also emerged other important aspects of similar biochemical alkenal/alkenone reductions. These were associated with $\alpha,\beta$ double bond reductions of various keto/aldehydic moieties in both prosta-glandin metabolism of guinea pig (*Cavia porcellus*) kidney tissue, e.g. where 13 is converted to 14 by 12-hydroxydehydrogenase/15-oxo-prostaglandin 13-reductase (12-HD/PGR, Figure 4.2C and Figure 4.3)\textsuperscript{29}, and in rat liver detoxification of 4-hydroxy-(2E)-nonenal (4-HNE, 15, Figure 4.2D and Figure 4.3)\textsuperscript{30}. With the latter, this occurs through action of a NAD(P)H-dependent alkenal/one oxidoreductase (AOR) to form 4-hydroxynonanal (4-HNA, 16)\textsuperscript{30}, in a manner analogous to the action of PtPPDBR. Both proteins, 12-HD/PGR and AOR, however, only have 51/34 and 51/33% similarity/identity to PtPPDBR, and 89/79% similarity/identity to each other (Table 4.1). Interestingly, in humans, at physiological concentrations, 4-HNE (15), a product of lipid peroxidation, can induce apoptosis, affect cell-signaling pathways, and also form 4-HNE-protein adducts, such as those found in Alzheimer and atherosclerotic plaques. Thus, there is a significant correlation in generation of these reactive intermediates to incidences of cancer, heart and Alzheimer diseases in humans\textsuperscript{30,31}.

In *A. thaliana*, the PtPPDBR homologue (At5g16970) was first described in 1995\textsuperscript{32}. At that point, it was reported as being induced by oxidative stress, and was considered then as a member of the $\zeta$-crystallin protein family, based on *circa* 25-41% identity to mammalian $\zeta$-crystallins of unknown biochemical function\textsuperscript{32,33}. While this *Arabidopsis* protein was later shown capable of reducing diamide/quinone linkages\textsuperscript{34}, Mano *et al.* have since demonstrated that it can also reduce 4-HNE (15) and related potential substrates\textsuperscript{35,36}. This has led to a consideration that the *Arabidopsis* protein may, therefore, be involved in lipid-peroxidation derived alkenal reductions in response to oxidative stress.
Figure 4.3. Amino acid sequence alignment of AtDBR1 with other oxidoreductases. PtPPDBR from *P. taeda*, PulR from *M. piperita*, AOR from rat (*Rattus norvegicus*), 12-HD/PGR from guinea pig (*C. porcellus*, PDB code 1V3V), and putative NADPH-dependent oxidoreductase from *M. musculus* (1VJ1). The nucleotide-binding domain is indicated by a thin dotted line; the conserved residues in the AXXGXXG motif and the conserved Tyr residues (Tyr-260 for AtDBR1) are highlighted in red and yellow, respectively. Secondary structural elements of AtDBR1 are highlighted in colored bars (orange for the β-strands and green for the α-helices).

### TABLE 4.1. Comparison of amino acid similarity and identity for various alkenal/alkenone reductases (%)

<table>
<thead>
<tr>
<th></th>
<th>AtDBR1</th>
<th>PulR</th>
<th>AOR</th>
<th>12-HD/PGR</th>
<th>1VJ1*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PtPPDBR</td>
<td>63/43</td>
<td>62/44</td>
<td>51/33</td>
<td>51/34</td>
<td>47/31</td>
</tr>
<tr>
<td>AtDBR1</td>
<td>79/63</td>
<td>56/39</td>
<td>56/41</td>
<td>45/30</td>
<td></td>
</tr>
<tr>
<td>PulR</td>
<td>56/39</td>
<td></td>
<td>58/37</td>
<td>48/28</td>
<td></td>
</tr>
<tr>
<td>AOR</td>
<td>89/79</td>
<td></td>
<td>50/37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-HD/PGR</td>
<td>51/37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* PDB code 1VJ1, the putative NADP-dependent oxidoreductase from *M. musculus*.
Taken together, all of the above reports give an indication of the broad range of substrate versatility, and thus of distinct potential biochemical functions, that the double bond reductases can have. Such broad substrate versatilities have, however, resulted in significant confusion as to the actual range of physiological roles in vivo. Based on the demonstrated identification of PtPPDBR physiological function\textsuperscript{27,37}, we can now describe the comparative detailed characterization of At5g16970. This is in terms of: (i) its substrate versatility (utilizing \(p\)-coumaryl (6)/coniferyl (7) aldehydes, and 4-HNE (15) as substrates in vitro), as well as the corresponding kinetic parameters; (ii) the crystal structures of the apo, binary and ternary complex forms for both \(p\)-coumaryl aldehyde (6) and 4-HNE (15), and the corresponding catalytic mechanism.

4.2 EXPERIMENTAL PROCEDURES

Material—All solvents and chemicals used were reagent or HPLC grade unless otherwise stated. Chemical reactions were carried out under anhydrous conditions in a \(N_2\) atmosphere using dry freshly distilled solvents. Silica gel thin-layer chromatography (TLC) and silica gel column chromatography employed Partisil\textsuperscript{®} PK5F (Silica Gel 150 Å, 1000 μm thickness), AL SIL G/UV\textsubscript{254} (Whatman, 20 × 20 cm, 0.25 mm) and silica gel 60 (EM Science), respectively, with UV and phosphomolybdic acid reagent (2.5% in \(H_2O\))–heat used for TLC plate visualization. NMR spectra were recorded on a Varian Mercury-Vx 300 MHz spectrometer operating at 300.1 (\(^1H\)) and at 75.5 (\(^13C\)), with chemical shifts given in \(\delta\) ppm relative to tetramethylsilane (TMS) and \(J\) values in Hz, respectively.

HPLC analyses employed an Alliance 2695 HPLC system equipped with a UV-Vis diode-array detection (Waters, Milford, MA), whereas GC-MS analyses utilized a HP 5973 MS detector (EI mode, 70 eV), a HP 6890 GC system, and a 7673 series injector equipped with a
RESTEK-5Sil-MS (30 m × 0.25 mm × 0.25 μm) column. The carrier gas was He with an initial flow of 1.4 ml min\(^{-1}\) at a pressure of 11.65 psi, with samples analyzed using the split injection mode and an injector temperature of 250 °C (split ratio: 10.1:1, split flow: 14.0 ml min\(^{-1}\)). The GC temperature program was initiated at 70 °C for 1 min, increasing to 170 °C at a rate of 8 °C per min, and held at 170 °C for 10 min. The mass range was scanned from \(m/z\) 50 to 800. HPLC electrospray ionization mass spectrometric analyses (LC-ESI-MS) were recorded on a Waters 2690 Alliance/Finnigan MAT LCQ, whereas electron impact mass spectra (EI-MS) were acquired on a Waters Integrity™ HPLC/MS system at an ionization voltage of 70 eV.

*Chemical Syntheses*—*E*-p-coumaryl aldehyde (6) was synthesized exactly as described in Kim *et al.*\(^{19}\), whereas coniferyl aldehyde (7) was from Aldrich.

Dihydroconiferyl aldehyde (9): To a solution of 4-*O*-tert-butyldimethylsilyl-*E*-coniferyl aldehyde\(^{19}\) (500 mg, 1.71 mmol) in dry MeOH (10 ml) was added 10% Pd/C (70 mg), with the resulting suspension stirred at room temperature under H\(_2\) for 6–8 h. The reaction mixture was then filtered, with the filtrate dried *in vacuo*. To the resulting 4-*O*-tert-butyldimethylsilyl dihydroconiferyl aldehyde derivative (450 mg, 1.53 mmol), dissolved in dry THF (10 ml) under N\(_2\) at 0 °C, was next added a solution of tetrabutylammonium fluoride (TBAF) (1.0 M in THF, 1.8 ml, 1.8 mmol) with the whole then allowed to stir for 45 min. The reaction mixture was quenched with saturated NH\(_4\)Cl solution (30 ml) and extracted with dry diethyl ether (50 ml × 2), with the resulting combined organic solubles washed successively with water (30 ml × 2). The organic solubles were then dried (Na\(_2\)SO\(_4\)) and evaporated to dryness *in vacuo*. The residue so obtained was subjected to silica gel column chromatography (eluent: CHCl\(_3\)–MeOH, 9:1) and then preparative silica gel TLC eluted with CHCl\(_3\)–MeOH (9:1) to yield dihydroconiferyl aldehyde (9, 190 mg, 1.05 mmol, 70% yield). \(^1\)H NMR \(\delta\) (300 MHz, CDCl\(_3\)): 2.75 (2H, m, H-8),
2.89 (2H, m, H-7), 3.87 (3H, s, OMe), 6.66 (1H, dd, J = 8.7, 2.3 Hz, H-6), 6.70 (1H, d, J = 2.3 Hz, H-2), 6.84 (1H, d, J = 8.7 Hz, H-5), 9.81 (1H, t, J = 1.67 Hz, CHO). EIMS (70 eV) m/z 180 [M^+] (50), 137 (100).

*p*-Dihydrocoumaryl aldehyde (8): This was synthesized exactly as described above for dihydroconiferyl aldehyde (9) by Pd/C-hydrogenation of 4-0-tert-butyl dimethylsilyl-**E-p**-coumaryl aldehyde^{19} (100 mg, 0.38 mmol) followed by deprotection with TBAF to afford *p*-dihydrocoumaryl aldehyde (8, 23.6 mg, 0.15 mmol, 52% yield). ESI-MS m/z: 149.2 [M – H].

NMR spectra of *p*-dihydrocoumaryl aldehyde (8) was in close agreement with reported data^{38,39}.

4-Hydroxy-2(***E**)-nonenal (4-HNE, 15) was synthesized exactly as described by Gardner et al^{40} except for the purification/isolation steps. 4-HNE (15) was purified by silica gel preparative TLC using hexanes–diethyl ether (6:4) as eluants, with detection by UV absorption and visualization of a bluish spot after staining and heating with phosphomolybdic acid reagent (2.5% (w/v) in H₂O). The band corresponding to 4-HNE (15) was excised and eluted with acetone–diethyl ether (1:2, 30 ml × 2), with the combined organic solubles filtered, and evaporated to dryness under N₂ atmosphere by adding MeOH (1 ml) to avoid loss of 4-HNE (15) due to its volatility and stored at –80 °C. 4-HNE (15) was obtained in *circa* 60-65% yield (46.7 mg, 0.3 mmol). GC-MS was next employed to determine its purity as well as its fragmentation pattern by preparing its trimethylsilyloxy derivative^{41} using bis(trimethylsilyl) trifluoroacetamide–chlorotri-methylsilane (99:1 Supelco, 50 μl) and pyridine (20 μl). The following fragments were observed at retention time 13.10 min (~90% pure) from the GC-MS analyses: m/z 228 [M**+]², 213 [M⁺ – CH₃], 199 [M⁺ – CHO], 157 [CHO-C₂H₂-CH-OSi(CH₃)₃⁺], 129 [CHO-C₂H₂-CH-OSi(CH₃)₃⁺ – CO], 73 [Si(CH₃)₃⁺]. NMR spectra for HNE (15) was in close agreement with reported data^{40,42}. 
4-Hydroxy nonanal (4-HNA, 16) was synthesized via the diisobutylaluminium hydride (DIBAL) reduction of γ-nonanoic lactone. The latter was synthesized by the Knoevenagel reaction of malonic acid with heptanal and lactonization with 85% sulfuric acid at 80 °C for 1 h. The lactone (500 mg, 3.2 mmol) so-formed was then reduced with a 1 M solution of DIBAL in toluene (4.8 ml, 4.8 mmol) at −78 °C for 2 h exactly as described by Bloch and Gilbert. 4-HNA (16) was purified by silica gel preparative TLC using hexanes–diethyl ether (1:1) as eluants, with the band corresponding to 4-HNA (16) excised and eluted with acetone–diethyl ether (1:1, 30 ml × 2). The organic solubles were then combined, filtered, dried under N₂ atmosphere to afford 4-HNA (16, 420 mg, 2.65 mmol, 80–85% yield). GC-MS analyses of 4-HNA (16) showed the following fragments at retention time = 11.57 and 11.64 min corresponding to both S- (16a) and R- (16b) isomers, ~95% pure: m/z 230 [M⁺], 215 [M⁺–CH₃], 159 [CHO–C₂H₄–CH–OSi(CH₃)₃⁺], 73 [Si(CH₃)₃⁺]. The NMR spectrum for 4-HNA (16) was in close agreement with reported data.

**Expression and Purification of AtDBR1**—AtDBR1 (At5g16970), cloned into an Invitrogen pTrcHis2-TOPO® TA vector, was transformed into TOP10 Escherichia coli cells. Expression of AtDBR1 was induced by addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a 0.5 mM final concentration at mid-log phase (A₆₀₀ = 0.5). The induced cell suspension cultures were grown for 12 h at 25 °C, with shaking at 250 rpm, with the cells subsequently harvested by centrifugation (3,000 × g for 20 min). The AtDBR1-derived pellet was suspended in lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8.0, 10% glycerol), sonicated (5 × 10 s, model 450 sonifier®, Branson Ultrasonics Co.), and centrifuged (20,000 × g for 40 min). After centrifugation, the supernatant was incubated with Ni-nitrilotriacetic acid agarose (Qiagen, Hilden, Germany) for 1 h in an overhead shaker at 4 °C. After washing with 10 column volumes
of wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 20 mM imidazole, pH 8.0), the fusion protein was eluted stepwise with 50 mM NaH₂PO₄, 300 mM NaCl, 100–300 mM imidazole pH 8.0. Thereafter, the AtDBR1-enriched fraction was subjected to anion exchange column chromatography (Self Pack™ POROS® 10HQ, Applied Biosystems) pre-equilibrated in Buffer A [50 mM Tris-HCl (pH 8.0) containing EDTA (1 mM) and dithiothreitol (1 mM)], at a flow rate of 3 ml min⁻¹. Recombinant AtDBR1 was eluted using a NaCl step gradient (to 0.1, 0.2, 0.5, and 2.0 M, 50 ml each), with the corresponding fractions of interest (eluting at 0.1 M NaCl) desalted and concentrated into Buffer B [20 mM Tris-HCl (pH 7.5)] by ultrafiltration in an Amicon 8050 cell with a 10-kDa cut-off membrane (Millipore). This fraction was applied to a MonoQ™ GL10/100 anion exchange column (Amersham Biosciences) equilibrated in Buffer B at a flow rate of 2 ml min⁻¹ and eluted with a NaCl step gradient (0.05, 0.1, 0.2, 0.4, and 2.0 M; 20 ml each); the catalytically active AtDBR1 fraction eluted at 0.05 M NaCl. The AtDBR1 so obtained was concentrated, with a final purity >99% as estimated by SDS-PAGE (Coomassie blue staining).

**Kinetic Parameter Determinations**—When p-coumaryl (6) and coniferyl (7) aldehydes were used as substrates, initial velocity kinetics were determined as follows: Assays consisted of MES buffer (100 mM, pH 6.25, 100 μl), 130 μl (3–8 μg) of AtDBR1 purified as described in Kim et al. 19 in Tris-HCl (20 mM, pH 7.5), aldehydes 6 or 7 (10 to 0.1 μM, 10 μl), NADPH (25 mM, 10 μl), in a total volume of 250 μl. Enzymatic reactions were initiated by addition of AtDBR1 and, after 10 min incubation at 30 ºC, were stopped by addition of glacial acetic acid (10 μl). An aliquot (80 μl) of each assay mixture was next subjected to reversed-phase HPLC analysis on a Symmetry Shield RP₈ column (Waters; 150 × 3.9 mm inner diameter, 5 μm particle size) with the following elution conditions at a flow rate of 1 ml min⁻¹ and detection at 280 nm:
the column was pre-equilibrated in a 5:95 ratio of CH$_3$CN (solvent A) and 3% AcOH in H$_2$O (solvent B). After introduction of the sample, this composition was held for 1 min, after which a linear gradient to A:B (40:60) over 39 min was carried out, followed by a linear gradient to A:B (5:95) in 5 min, this being held for 1 min. The amounts of products 8 and 9 formed were determined using pre-established calibration curves.

Assays with 4-HNE (15) as substrate were carried out as above but with 0.5 μg AtDBR1 and an incubation time of 2 min. After addition of glacial acetic acid (10 μl), 4-hydroxybenzaldehyde (10 mM, 5 μl) was added as an internal standard, with the mixture extracted with diethyl ether (1 ml × 3). After vortexing, the diethyl ether layers were removed and combined, with the ether solubles dried (over Na$_2$SO$_4$), and the volume reduced to ~100 μl under N$_2$ at which time 1,4-dioxane was added (100 μl). The resulting solution was transferred to a GC vial, with the volume reduced to ~100 μl under N$_2$ and the trimethylsilyloxy derivative prepared. The samples were next subjected to GC-MS analyses as described in the Chemical Synthesis section.

**Molecular Mass Determination**—Size exclusion chromatography/multiangle light scattering and dynamic light scattering were performed as previously described in Youn et al.$^{16,20}$, with light scattering data acquired through accumulation (3 times) of 10 scans.

**Crystallization of AtDBR1**—For crystallization, a solution of purified AtDBR1 (48 mg ml$^{-1}$) in 20 mM Tris-HCl (pH 8.0) containing 1 mM EDTA and 1 mM dithiothreitol was prepared. Crystallization trials were performed using the hanging drop vapor diffusion method at two temperatures (277 and 293 K). Apo-AtDBR1 crystals were obtained by mixing the above protein solution (1.5 μl) with an equal volume of reservoir solution containing 20% (w/v) PEG 3350 and 0.2 M potassium chloride. Crystals usually appeared after 5 days, and larger crystals
with dimensions of \(~0.3 \times 0.5 \times 0.8\) mm were obtained after 2 weeks. Although these crystals were fairly large, they were hollow and had a diffraction limit of \(3.5\) Å. Crystals were stabilized and the diffraction limit was increased (up to \(2.5\) Å) by slowly adding concentrated buffer solution to the drops in which crystals were grown. For most crystals, the final buffer composition was \(30\% (\text{w/v})\) PEG 3350, \(0.3\) M potassium chloride. The crystals of AtDBR1 belong to the orthorhombic space group, \(P2_12_12_1\) \((a = 49.46, b = 122.98, c = 148.00\) Å\), with two molecules in an asymmetric unit (Table 4.2). The binary complex (AtDBR1–NADP\(^{+}\)) and the ternary complex (AtDBR1–NADP\(^{+}\)–\(p\)-coumaryl aldehyde (6): ternary I) crystals were also produced under the same conditions, except for addition of \(5\) mM NADP\(^{+}\) and \(5\) mM \(p\)-coumaryl aldehyde (6) into the reservoir solution, respectively. Crystals of the other ternary complex (AtDBR1–NADP\(^{+}\)–4-HNE (15): ternary II) were produced by soaking the binary complex crystal in a solution of \(1\) mM 4-HNE (15) (Alexis Biochemicals Inc.). All binary and ternary complexes crystallized in an orthorhombic space group, \(P2_12_12_1\), with corresponding unit cells of \(a = 49.15, b = 122.66, c = 147.94\) (binary), \(a = 49.04, b = 122.54, c = 147.65\) Å (ternary I), and \(a = 49.08, b = 122.45, c = 147.84\) (ternary II), respectively. The apo-form (2.5 Å resolution), the binary complex (2.8 Å), and both of the ternary complex (2.8 Å) data were collected from the Berkeley Advanced Light Source (ALS, beam line 8.2.1 / apo-form and ternary II), Chicago Advanced Photon Source (APS, beam line NE-CAT/8-BMD / binary complex) and Rigaku Saturn 92/MicroMax-007 (Washington State University / ternary I) at a temperature of 100 K. Before freezing, the corresponding crystals were soaked for 5 min in cryoprotectant (25% glycerol in each reservoir solution).

**Structural Solution and Refinement**—The structure of apo-AtDBR1 was solved by the molecular replacement method using a coordinate of the guinea pig \((Cavia porcellus)\) leukotriene
B₄ 12-hydroxydehydrogenase/15-oxo-prostaglandin 13-reductase (12-HD/PGR, 1V3V)²⁹ and the software package AMoRe⁴⁸. The rigid body refinement of the initial position was carried out by using 15.0 to 3.0 Å resolution data and gave an R value of 35%. After several cycles of positional and temperature factor refinements using the program X-PLOR⁴⁹ and a series of simulated annealing omit maps, most residues were fitted against the electron density. The binary and ternary complexes of AtDBR1 were again solved by the molecular replacement method but now using the apo-AtDBR1 coordinates. The final R factors (Table 4.2) for the apo-form, as well as the binary and two ternary complexes of AtDBR1, were 19.7% (R_free = 24.3% for the random 5% data), 20.1% (R_free = 24.9% for the random 5% data), 18.9% (R_free = 23.1% for the random 5% data) and 19.9% (R_free = 24.1% for the random 5% data), respectively. The number of reflections above 2σ level for the apo-form were 30,000 (95% completeness) between 15.0 and 2.5 Å resolution. The crystals of the NADP⁺ binary complexes did not diffract as well as the apo-form and gave reflection numbers of 21,151 (above 2σ, 95% completeness) between 15.0 and 2.8 Å resolution. In addition, both ternary complex data of p-coumaryl aldehyde (6) and 4-HNE (15) were collected between 15.0 and 2.8 Å resolution, 20,191 and 20,187 (above 2σ), respectively. The root mean square deviations (r.m.s.d.) (from ideal geometry) of the final coordinates corresponding to the apo-form, and binary, ternary I and ternary II complexes are 0.014, 0.015, 0.024 and 0.025 Å for bonds and 3.2, 3.5, 4.5 and 4.6° for angles, respectively. All AtDBR1 coordinates have been deposited in the Protein Data Bank (apo-form: 2J3H, binary complex: 2J3I, ternary complex with p-coumaryl aldehyde (6): 2J3J, ternary complex with 4-HNE (15): 2J3K).
### TABLE 4.2. Crystallographic data for the AtDBR1 apo- and binary and ternary complex forms. Beam lines: ALS, Advanced Light Source; APS NE-CAT, Advanced Photon Source, Northeastern Collaborative Access Team; WSU, Washington State University.

<table>
<thead>
<tr>
<th>Data</th>
<th>Apoform</th>
<th>Binary&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ternary I&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Ternary II&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam line</td>
<td>ALS 8.2.1</td>
<td>APS NE-CAT</td>
<td>WSU (MM007)</td>
<td>ALS 8.2.1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.07812</td>
<td>0.97914</td>
<td>1.5418</td>
<td>1.07812</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50—2.5</td>
<td>50—2.8</td>
<td>50—2.8</td>
<td>50—2.8</td>
</tr>
<tr>
<td>Space group</td>
<td>$P_2_1\overline{2}1$</td>
<td>$P_2_1\overline{2}1$</td>
<td>$P_2_1\overline{2}1$</td>
<td>$P_2_1\overline{2}1$</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>$a = 49.46$</td>
<td>$a = 49.15$</td>
<td>$a = 49.04$</td>
<td>$a = 49.08$</td>
</tr>
<tr>
<td></td>
<td>$b = 122.98$</td>
<td>$b = 122.66$</td>
<td>$b = 122.54$</td>
<td>$b = 122.45$</td>
</tr>
<tr>
<td></td>
<td>$c = 148.00$</td>
<td>$c = 147.94$</td>
<td>$c = 147.65$</td>
<td>$c = 147.84$</td>
</tr>
<tr>
<td>Asymmetric unit</td>
<td>2 molecules</td>
<td>2 molecules</td>
<td>2 molecules</td>
<td>2 molecules</td>
</tr>
<tr>
<td>Total observations</td>
<td>314,652</td>
<td>97,838</td>
<td>152,240</td>
<td>134,464</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>61,038</td>
<td>22,577</td>
<td>22,542</td>
<td>23,815</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.5</td>
<td>98.9</td>
<td>99.3</td>
<td>82.6</td>
</tr>
<tr>
<td>$R_{sym}^{d,e}$</td>
<td>4.5 (15.5)</td>
<td>6.0 (14.2)</td>
<td>5.5 (12.6)</td>
<td>5.3 (14.5)</td>
</tr>
</tbody>
</table>

| Refinement            |                      |                    |                      |
| Resolution (Å)        | 15—2.5               | 15—2.8             | 15—2.8               | 15—2.8                 |
| No. of reflections    | 30,000               | 21,151             | 20,191               | 20,187                 |
| $R_{cryst}^{f}$ (%)   | 19.7                 | 20.1               | 18.9                 | 19.9                   |
| $R_{free}^{g}$ (%)    | 24.3                 | 24.9               | 23.1                 | 24.1                   |
| r.m.s.d. bonds (Å)    | 0.014                | 0.015              | 0.024                | 0.025                  |
| r.m.s.d. angles (°)   | 3.2                  | 3.5                | 4.5                  | 4.6                    |

| No. of atoms          |                      |                    |                      |
| Protein & ligand      | 5,304                | 5,400              | 5,420                | 5,422                  |
| Water                 | 138                  | 93                 | 115                  | 136                    |
| Water                 | 138                  | 93                 | 115                  | 136                    |

<sup>a</sup> AtDBR1 + NADP<sup>e</sup>

<sup>b</sup> AtDBR1 + NADP<sup>e</sup> + p-coumaryl aldehyde (6)

<sup>c</sup> AtDBR1 + NADP<sup>e</sup> + 4-hydroxy-2-nonenal (15)

<sup>d</sup>Numerals in parentheses refer to the highest shell

<sup>e</sup>$R_{sym} = \Sigma I_h - \langle I_h \rangle / \Sigma I_h$, where $\langle I_h \rangle$ is the average intensity over symmetry equivalent reflections

<sup>f</sup>$R_{cryst} = \Sigma |F_{obs} - F_{calc}| / \Sigma F_{obs}$, where summation is over the data used for refinement

<sup>g</sup>$R_{free}$ was calculated same as for $R_{cryst}$ using 5% of the data that was excluded from refinement

4.3 RESULTS

It was first instructive to obtain the needed kinetic data for the recombinant (His-tagged) AtDBR1 using as substrates, p-coumaryl aldehyde (6), coniferyl aldehyde (7) and 4-hydroxy-2E-
nonenal (4-HNE, 15), respectively. In this regard, all of the required substrates were synthesized as follows: \textit{E-\textit{p}}-coumaryl aldehyde (6) was obtained in 79\% overall yield by the lithium aluminium hydride (LAH) reduction of the 4-\textit{O}-tert-butyldimethylsilyl derivative of methyl-4-hydroxycinnamate, followed by oxidation with 2,3-dichloro-5,6-dicyanobenzo-quinone (DDQ) and finally by deprotection with tetra-butylammonium fluoride (TBAF)\textsuperscript{10}. \textit{E}-coniferyl aldehyde (7) was purchased from Sigma-Aldrich, whereas the (unstable) 4-HNE (15) was synthesized by epoxidation of 3(Z)-nonenal with 3-chloroperoxybenzoic acid followed by periodinane oxidation and treatment with 1.3 M NaOH\textsuperscript{40}. The reduced products, \textit{p}-dihydrocoumaryl (8) and dihydroconiferyl (9) aldehydes, were also individually synthesized via Pd/C hydrogenation of the corresponding 4-\textit{O}-tert-butyldimethylsilyl aldehyde derivatives with subsequent deprotection using TBAF in \textit{circa} 60-70\% yields (see Experimental Procedures). 4-HNA (16) was obtained via DIBAL reduction of \textit{\gamma}-nonanoic lactone to afford the desired product (\textit{R} and \textit{S} isomers) in an approximately 1:1 ratio, as evidenced by GC/MS analyses. Additionally, a variety of monomeric phenylpropanoid substrate analogues and their dihydro products were prepared, e.g. cinnamyl (17)/5-hydroxyconiferyl (18)/sinapyl (19) aldehydes, the corresponding dihydrocinnamyl (30)/5-hydroxydihydroconiferyl (31)/dihydrosinapyl (32) aldehydes, as well as \textit{p}-coumaric (20), caffeic (21), ferulic (22), 5-hydroxyferulic (23), and sinapic (24) acids and the potential dihydroproducts (3, 4, 33-35). In an analogous manner, \textit{p}-coumaryl (25), caffeyl (26), coniferyl (27), 5-hydroxyconiferyl (28), and sinapyl (29) alcohols, and the dihydro derivatives 1, 2, 36-38 were also synthesized (see Figs. 4.1 and 4.4 for structures; Kim \textit{et al.}, manuscript in preparation). Each of the purified potential substrates (6, 7, 17-29) and products (1-4, 30-38) were then used to establish standard curves using HPLC, whereas GC/MS was used for 4-HNE (15)/4-HNA (16) analyses, i.e. in order to directly quantify substrate turnover and product formation. By contrast,
all previous studies of 4-HNE (15) enzymatic reductions were carried out indirectly by measuring changes in UV absorption at 340 nm.

**Kinetic Parameters/Substrate Versatility**— With all of the corresponding potential substrates needed for the study in hand, the AtDBR1 cDNA was cloned into a pTrcHis-TOPO vector containing a N-terminal polyhistidine (6 × His) tag, with the plasmid used to transform *E. coli* TOP10 cells for gene expression. AtDBR1 was purified to apparent homogeneity (evaluated by SDS-PAGE with silver staining) following metal chelate affinity column chromatography as described in Kim *et al.*19.

![Phenylpropanoids (structures 17-29) and dihydrophenylpropanoids (structures 30-38).](image)

Table 4.3 summarizes the kinetic data so obtained. In our hands, both *p*-coumaryl (6) and coniferyl (7) aldehydes served as the most efficient substrates (*k*<sub>cat</sub>*K*<sub>m</sub> of 5360 and 1060 M<sup>−1</sup>s<sup>−1</sup>) relative to that of 4-HNE (15) which was less efficiently utilized (600 M<sup>−1</sup>s<sup>−1</sup>). Additionally, several other substrates were evaluated for their capacity to undergo comparable alkenal (α,β double bond) reductions, including cinnamyl (17), 5-hydroxyconiferyl (18) and sinapyl (19) aldehydes. However, none of these closely related substrate analogues were converted into the
corresponding dihydro derivatives (30-32), indicating that the substrate versatility of the AtDBR1 was fairly limited. This observation contrasts with reports that the alkenal oxidoreductase (AOR) from rat liver is capable of reducing the α,β double bond of cinnamyl aldehyde (17), along with a very broad range of potential substrates; however, in that study, the presumed conversion of 17 was evaluated indirectly by measuring changes in absorbance at 340 nm \(^{30}\), rather than by HPLC/MS of the resulting dihydro product \((30)\). In an analogous manner, none of the hydroxycinnamic acids (20-24) were converted into the dihydrocinnamates (3, 4, 33-35) by AtDBR1; nor were any of the corresponding alcohols (25-29) directly reduced to afford 1, 2, 36-38.

**TABLE 4.3.** Kinetic parameters for AtDBR1

<table>
<thead>
<tr>
<th>Substrate</th>
<th>(K_m)</th>
<th>(V_{\text{max}})</th>
<th>(k_{\text{cat}})</th>
<th>(k_{\text{cat}}/K_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Coumaryl aldehyde (6)</td>
<td>0.53</td>
<td>32.4</td>
<td>2.82</td>
<td>5360</td>
</tr>
<tr>
<td>Coniferyl aldehyde (7)</td>
<td>0.41</td>
<td>5.3</td>
<td>0.43</td>
<td>1060</td>
</tr>
<tr>
<td>4-HNE (15)</td>
<td>0.28</td>
<td>2.0</td>
<td>0.16</td>
<td>600</td>
</tr>
</tbody>
</table>

As anticipated, 4-HNE (15) was also readily reduced, as evidenced by direct GC/MS analyses of amounts of both 4-HNE (15) and the enzymatic product 4-HNA (16) in the assay mixtures. However, these assays revealed that while 4-HNE (15) had a slightly lower \(K_m\) than p-coumaryl (6)/coniferyl (7) aldehydes, the overall \(V_{\text{max}}\) was significantly lower leading to a decreased catalytic efficiency. Interestingly, Mano et al.\(^{35}\) had reported a much lower \(K_m\) (by indirect detection), as was also noted for the indirect measurement of cinnamyl aldehyde (17) reduction by Dick et al.\(^{30}\). In our hands, such an efficient conversion of 4-HNE (15) into 4-HNA (16) could not be duplicated; nor was cinnamyl aldehyde (17) reduced as indicated above. Based on the data herein, we, therefore, conclude that p-coumaryl aldehyde (6) serves as the best \textit{in vitro} substrate. The data so obtained also revealed that for effective catalytic activity, the double
bond must be in conjugation with an aldehydic moiety, although the substrate versatility was fairly limited.

Overall Structure of AtDBR1—In order to understand in greater detail the mechanistic-structural basis for the AtDBR1 kinetic properties, the recombinant AtDBR1 was next crystallized in its apo-form, NADP⁺ binary and ternary complexes using the two different substrates, p-coumaryl aldehyde (6) and 4-HNE (15), respectively. Crystals of both ternary complexes were obtained by mixing AtDBR1 with NADP⁺/p-coumaryl aldehyde (6) and by soaking the AtDBR1/NADP⁺ binary complex crystals in a solution containing 4-HNE (15), respectively. The apo form of AtDBR1 was determined at 2.5 Å resolution by the molecular replacement method using coordinates of the afore-mentioned 12-HD/PGR (1V3V) from guinea pig (C. porcellus)²⁹, which shows 56/41% similarity/identity to AtDBR1 (Table 4.1). Additionally, the binary and ternary complex structures of AtDBR1 were determined at 2.8 Å resolution, using the coordinates of the deduced structure of the AtDBR1 apo-form.

As shown in Figure 4.5, AtDBR1 is a homodimer with two subunits arranged through a non-crystallographic 2-fold axis. The two subunits are virtually superimposable, with a root mean square deviation (r.m.s.d.) of 0.67 Å between the corresponding Cα atoms of the two subunits without including the residues from 60 to 70 in the case of the apo-form. The loop region of the residues 60–70 connecting the α1 and β4 regions is disordered in one subunit, whereas the corresponding area of the other subunit is ordered due to crystal packing interactions. Both static and dynamic multiangle laser light-scattering data of the AtDBR1 molecules confirmed their dimeric nature (Figures 4.6A and 4.6B).

The subunit dimerization is achieved mainly through interactions between the two β-strands from each subunit (βF) in an anti-parallel manner, thereby forming an extended 12-stranded β-
sheet across the dimer interface (Figure 4.5). Another smaller dimer-forming interaction is at the opposite side of this βF strand interaction, involving two hydrogen bonds between the two pseudo two-fold related Val268 residues (not shown). There are additional salt bridges and hydrogen bonds between the side-chains, and between the side-chain and the main-chains of individual subunits.

**Figure 4.5.** Ternary complex of AtDBR1 homodimer with bound NADP⁺/p-coumaryl aldehyde (structure 6) (A) and NADP⁺/4-HNE (structure 15) (B). The substrate- and nucleotide-binding domains are depicted in green and violet in the lower subunit, and in brown and light blue in the upper subunit, respectively. NADP⁺ is shown in dark blue, with p-coumaryl aldehyde (6) or 4-HNE (15) in dark brown. The active site in the lower subunit is outlined by a dotted box, and the secondary structure elements of AtDBR1 are numbered sequentially as α1-α3/αA-αG and β1-β10/βA-βF.
Figure 4.6. Molecular mass of AtDBR1 in solution. A, elution profile of AtDBR1 monitored by multiangle laser light scattering (2 mg ml\(^{-1}\)). The elution profile is shown as molecular mass versus elution time. The thin solid lines represent changes in refractive index on an arbitrary scale that is proportional to protein concentration. The thick solid indicates the calculated molecular masses. B, dynamic light scattering data of AtDBR1 (2 mg ml\(^{-1}\)). The calculated molecular radius and molecular mass are 3.94 nm and 84 kDa, respectively, indicating its dimeric nature.

The overall fold of AtDBR1 belongs to that of the Zn-independent MDR superfamily. Like other MDR’s, each monomer is composed of two domains: a substrate-binding domain (residues 1–137 and 306–345) and a nucleotide-binding domain (Rossman fold, residues 138–305) (Figures 4.3 and 4.5). The substrate binding domain has three \(\alpha\)-helices and ten \(\beta\)-strands forming two \(\beta\)-sheets, one of which is a highly twisted, eight stranded, \(\beta\)-barrel-like structure as observed previously for the putative NADP-dependent oxidoreductase from \textit{Mus musculus} (Figure 4.3, 1VJ1)\(^50\). The nucleotide-binding domain of AtDBR1 also has seven \(\alpha\)-helices and six \(\beta\)-strands forming a typical six-stranded parallel \(\beta\)-sheet flanked by three helices on each side.
In the binary and ternary complexes, the corresponding NADP\(^+\) and substrates were located at the active site clefts between the substrate- and nucleotide-binding domains as described in detail below. Upon NADP\(^+\) and/or substrate binding, no significant changes were detected in either overall conformation or of the amino acids in the binding pockets as reflected in small r.m.s.d. values of 0.42–0.66 Å (not shown) among the C\(\alpha\) atoms of the apo form, binary and ternary complexes. A minor positional change in the backbone was observed in the area of residues 36–39, which connects \(\beta\)2 and \(\beta\)3.

**Structural Alignment**—In general, the MDR superfamily can be divided into two subgroups\(^{51}\). One subgroup contains no Zn\(^{2+}\), as for AtDBR1, PtPPDBR, 12-HD/PGR and AOR, whereas the other has catalytic and/or structural Zn\(^{2+}\), such as in the well-studied horse liver alcohol dehydrogenase\(^{52}\) and cinnamyl alcohol dehydrogenase\(^{20}\). As noted earlier (Figure 4.3), amino acid sequence alignments had also revealed that the plant enzymes, AtDBR1 and PulR, were the most closely related, albeit whose 3D structures had not been reported.

A Dali search\(^{53}\) of the PDB database indicated that the highest match was to the 12-HD/PGR from guinea pig *C. porcellus* (1V3V) with a Z-score of 43.0; this was followed by several quinone reductases, including the human \(\zeta\)-crystalline (1YB5) with a Z-score of 32.3 (not shown), and the putative NADP-dependent oxidoreductase from *M. musculus* (1VJ1) of 31.4. The latter two show only 30 and 19\% identity, however, to AtDBR1 based on amino acid sequence comparisons. Nevertheless, all of these high Z-scored proteins belong to the Zn-independent MDR superfamily.

The sequence alignment among the above-mentioned enzymes (Figure 4.3) also revealed that AtDBR1 has several areas of deletion and insertion when compared with the others, which include several loops and even some secondary structural elements. Especially, one loop region,
which is composed of residues 30-40 connecting β2 and β3, is quite different, particularly when compared with the mammalian enzymes (Figure 4.3). In addition, a highly disordered loop region (amino acids 60-89) in the structure of AtDBR1 is not aligned well among the other enzymes.

In general, however, all of the enzymes show a high degree of sequence similarity for most of the nucleotide-binding domain up to the conserved\textsuperscript{255}GxxS\textsuperscript{258} motif; this is known to stabilize both adenine and nicotinamide moieties of the cofactor in the NADPH-bound form of quinone oxidoreductase\textsuperscript{50}. On the other hand, the remaining part of the nucleotide-binding domain (residues 262-305) has a low level of sequence similarity, and mainly participates in subunit interactions.

\textit{NADP\textsuperscript{+} Binding Site}—As for other members of the MDR superfamily, the electron density corresponding to NADP\textsuperscript{+} in AtDBR1 was located in the cleft between the two domains formed by the carboxy ends of βA, βB, βD, βE, βF, and the loop connecting βF and αG. There was also a clear and continuous electron density in the initial \textit{Fo}-\textit{Fc} map (Figure 4.7A inset); all of the side-chains around this cofactor were thus in a well-defined electron density that is highly conserved among the MDR family. The corresponding binding pocket for NADP\textsuperscript{+}, however, is filled with water molecules in the apo form. The coordinates for the two NADP\textsuperscript{+} molecules were thus refined to the same relative position in each subunit, with the bound NADP\textsuperscript{+} molecules adopting the \textit{anti} configuration for both the ribose-nicotinamide and ribose-adenine glycosidic bonds (Figure 4.7A).

Like other typical MDR’s, AtDBR1 has a glycine-rich motif, AxxGxxG shared by several oxidoreductases instead of the typical GxxGxxG, (Figure 4.3), between βA and αB
\((^{163}\text{AASGAVG}^{169})\), which is known to participate in binding of the pyrophosphate group of \(\text{NAD(P)}^+\) or \(\text{NAD(P)}H\) through a micro-dipole of the helix \((\alpha B)^{54}\). Interestingly, PtPPDBR and PulR have a slightly different sequential motif, \(\text{AAAGSVG}\), the significance of which (if any) is not yet known.

The pyrophosphate group of \(\text{NADP}^+\) in AtDBR is within hydrogen bonding distance to both the backbone amide nitrogens of residues, Ala167 and Val168, which reside inside the AxxGxxG motif, as well as the side-chains of two residues, Asn334 and Lys337 (Figure 4.7A). In addition, the backbones of Cys254 in \(\beta E\) and Phe284/Val286 in the \(\beta F\) are within hydrogen bonding distance to the amide moiety of the nicotinamide ring (Figure 4.7A). Interestingly, the above-mentioned \(^{255}\text{GxxS}^{258}\) motif, which is also located near the nicotinamide ring, can be extended to “\(\text{CGxxSxX}\)” among the compared Zn-independent MDR enzymes. The net effect is that, the freedom of the nicotinamide ring is very restricted, presumably fixing the position of the C-4 atom during catalysis.

Notably, both the hydroxyl group of the Tyr260 and the backbone nitrogen of Tyr53, which are completely conserved among all of the oxidoreductases compared (see Figure 4.3), are within hydrogen bonding distance to the O2' of the nicotinamide ribose ring. In contrast to the well-anchored nicotinamide ring, however, the adenine ring does not have much interaction with the surrounding amino acids, and thus should be relatively free to move as indicated by its high B-value.

*The AtDBR1, \(\text{NADP}^+\) and \(\text{p-Coumaryl Aldehyde (6) Ternary Complex}*-In the ternary complex, the observed distances between the C-4 atom of the nicotinamide ring and the \(\beta\)-carbon of the \(\alpha,\beta\ (7,8)\)-unsaturated double bond of \(\text{p-coumaryl aldehyde (6)}\), in the two subunits are 3.8 and 4.0 \(\text{Å}\), respectively. Thus, \(\text{p-coumaryl aldehyde (6)}\), located at the active site cleft between
the substrate- and nucleotide-binding domains, is in the proper orientation to the nicotinamide ring for the well-established A-face specific hydride transfer from C-4 to the corresponding substrate reaction center (Figure 4.7B).

Figure 4.7. Stereo-view of the substrate-binding pocket of AtDBR1 (viewed from the bulk solvent). A, binary complex. B, ternary complex with NADP⁺ and p-coumaryl aldehyde (6). C,
ternary complex with NADP⁺ and 4-HNE (15). The glycosidic bond angles of the bound NADP⁺ for all cases adopt an \textit{anti} conformation. In the \textit{insets}, experimental difference Fourier maps (|$F_o$| - |$F_c$|) corresponding to NADP⁺, \textit{p}-coumaryl aldehyde (6) and 4-HNE (15) in binary and ternary complexes of AtDBR1 are shown at a contour level of 3.0 \textsigma. The participating residues in the binding substrate and cofactor are shown with their residue \textit{position numbers}. The residues are depicted in \textit{light blue} and \textit{orange} to represent their belonging to two different subunits.

Unlike the NADP⁺ molecule deeply buried inside the binding pocket with many interactions involving various amino acid residues, \textit{p}-coumaryl aldehyde (6) is relatively exposed to the solvent; thus a weaker binding constant can be expected when compared to that of the cofactor. The inner wall of this exposed substrate-binding site is also surrounded by the nicotinamide ring, as well as Tyr53, Tyr81, Met138, Tyr260, Ser287 and Tyr290 from one subunit, and Ile275 and Tyr276 from another, indicative of its quite polar nature (Figure 4.7B). In either the apo-form or the NADP⁺ binary complex, however, this substrate-binding site is filled with water molecules, reflecting its somewhat hydrophilic nature.

Of particular note is that the phenolic hydroxy group of Tyr260 is within hydrogen bonding distance to the aldehydic oxygen of the substrate (6), in addition to the 2’-hydroxyl group of the nicotinamide ribose previously described. Therefore, we considered that this conserved Tyr260 residue is hydrogen bonded to both. In addition, the Tyr81 hydroxyl group is also potentially within hydrogen bonding distance to the phenolic hydroxyl group (Figure 4.7B), whereas the Ser287 hydroxyl group is \textasciitilde3.9 Å away, and thus probably too far away for hydrogen bond formation. However, due to their location in the flexible (high temperature factor) loops, there remained a possibility that hydroxyl groups in Tyr81 and Ser287 are potentially involved in facilitating substrate binding.
Lastly, the phenol ring of the highly conserved Tyr53, which is located between a relatively conserved short loop and one turn α-helix, is in a potential stacking interaction with the corresponding phenol ring of the bound p-coumaryl aldehyde (6), thereby probably further facilitating orientation of the substrate within the specificity pocket (Figure 4.7B).

The AtDBR1, NADP+ and 4-HNE (15) Ternary Complex—A well-defined density for 4-HNE (15) was also located in the same position as for p-coumaryl aldehyde (6) (Figure 4.7C inset). As a result, the corresponding conformation of the individual side-chain and NADP+ did not change between the two ternary complexes. As observed in the p-coumaryl aldehyde (6) ternary complex, Tyr260 is also within hydrogen bonding distance to the aldehydic oxygen of 4-HNE (15) and to the 2’- hydroxyl group of the nicotinamide ribose (Figure 4.7C). The distances between the C4 atom of the nicotinamide ring and the β-carbon of the α,β-unsaturated double bond of 4-HNE (15), in the two subunits are 3.8 and 4.0 Å, respectively, which are approximately the same distances as for the p-coumaryl aldehyde (6) complex. Moreover, in the ternary complex with p-coumaryl aldehyde (6), the conserved Tyr53 is now in a hydrophobic interaction with the corresponding aliphatic chain of the bound 4-HNE (15) thereby again facilitating orientation of the substrate within the specificity pocket (Figure 4.7C).

4.4 DISCUSSION

In contrast to two other oxidoreductases involved in related aspects of phenylpropanoid metabolism, i.e. secoisolariciresinol dehydrogenase15-17 and cinnamyl alcohol dehydrogenase 519, 20, the crystal structure of AtDBR1 shows that it belongs to the Zn2+-independent medium-chain dehydrogenase/ reductase (MDR) family. This class of enzymes as indicated above includes PtPPDBR, PulR, 12-HD/PGR, AOR, 1VJ1 and the ζ-crystallins. The enzymatic reaction mechanism of this class of enzymes which lack Zn2+ is, however, still unclear, e.g. when
compared to the very well-established Zn$^{2+}$-containing MDR’s, such as ADH. Indeed, a majority of the enzymes provisionally belonging to this group are still depicted as putative gene products$^{55}$.

In terms of the overall distribution of secondary structural elements and the local active site, AtDBR1 displays a striking similarity to 12-HD/PGR. As discussed earlier, the latter catalyzes reduction of a very broad array of $\alpha,\beta$-unsaturated ketones and aldehydes, including toxic products of lipid peroxidation, in addition to that of the 13-14-double bond of 15-oxoprostaglandins$^{29,56}$. Accordingly, most of the residues constituting the NADP(H) binding sites are very similar between AtDBR1 and 12-HD/PGR. Indeed, the overall shapes of the binding pocket for the cofactor are also very similar, and the bound cofactor in both reductases adopts the same sugar conformation.

For the guinea pig 12-HD/PGR and the rat liver AOR, a ketoreductase reaction mechanism has been proposed with a conjugated enolate as the reaction intermediate$^{29,55}$. Apparently, the latter enzyme is also able to reduce one of the phenylpropanoids, cinnamyl aldehyde (17)$^{30}$, whereas AtDBR1 does not. It is tempting to speculate, therefore, that the hydrophobic nature of the phenyl moiety of the cinnamyl aldehyde (17) can be accommodated in the hydrophobic substrate-binding pocket of 12-HD/PGR, but not in the AtDBR binding pocket, which is more of a hydrophilic nature. In addition, the intrinsic difference in terms of rotational freedom around the C1-C7 bond between cinnamyl aldehyde (17) and $p$-coumaryl aldehyde (6) [resulting from an ability of the latter to more readily form resonance hybrids because of the phenol functionality at C-4], may also provisionally explain the substrate preferences in binding and catalysis between 12-HD/PGR and AtDBR1.
Yet, in contrast to the similarities between the 12-HD/PGR and AtDBR1, the previously proposed candidate catalytic residue (Tyr262) for the former could not be substantiated. In part, this is because this residue (Tyr276 in AtDBR1) is not conserved in the other sequences (see Fig. 4.3) and, from the analysis of the crystal structure, it is too far away (5.5–5.9 Å) in AtDBR1 to be able to hydrogen bond with the aldehydic group of the substrate.

On the other hand, the residue, Tyr 260, is conserved in all enzymes shown in Figure 4.3. Hence, from the observed hydrogen bonding pattern and the strong conservation of Tyr260 among related enzymes, we propose that this residue serves as a general acid/base by stabilizing the enol form of the transition state (Figure 4.8B); Tyr260 is also apparently hydrogen bonded to the 2'-hydroxyl group of the nicotinamide ribose, thereby potentially enabling the pKa of its hydroxyl group to be further modulated. A similar catalytic mechanism can also be envisaged for 12-HD/PGR, PtPPDBR, PulR, AO and the putative NADPH-dependent oxidoreductase from *M. musculus* (1VJ1).

From our studies using [4R-^3^H]- and [4S-^3^H]-NADPH (unpublished observations), AtDBR1 and PtPPDBR catalyze the transfer of the pro-R hydride from NADPH to that of the double bond between C-7 and C-8, with this bond being polarized due to its neighboring carbonyl group. In agreement with this, the observed interactions between the carbonyl oxygen of both substrates, *p*-coumaryl aldehyde (6) and 4-HNE (15), and the Tyr260 residue can also potentially stabilize the transient oxyanion enolate intermediate, thereby facilitating pro-R hydride transfer from the C-4 atom of the NADPH nicotinamide to the substrate (Figures 4.8A and 4.8B). As anticipated, however, AtDBR1 was not able to reduce the corresponding unpolarized double bond of the various phenylpropanols 25 – 29, as shown from our kinetic data.
Overall, we can propose a concerted catalytic mechanism for hydride transfer to carbon 7 of p-coumaryl aldehyde (6) (C-3 in 4-HNE (15)), this ultimately being followed by protonation (of carbon 8 in p-coumaryl aldehyde (6), and of C-2 in 4-HNE (15)) (Figure 4.8). Indeed, since there is no adjacent amino acid in AtDBR1 able to donate a proton to C-8, this proton is presumed to be from the solvent. As mentioned above and shown in Fig. 4.7, the substrate-binding site of AtDBR1 is well exposed to the solvent. Considering the fact that one side-wall of the substrate-binding site is made up of a nicotinamide ring of the tightly bound cofactor, the binding of either p-coumaryl aldehyde (6) or 4-HNE (15) can presumably be facilitated by cofactor binding. Therefore, it is reasonable to speculate that NADPH binds first and NADP$^+$ leaves last, indicating that the kinetic mechanism of AtDBR could be a ordered Bi-Bi or Theorell-Chance, as in the cases of 12-HD/PGR and AOR$^{55, 57}$; the dissociation of the NADP$^+$ could, however, be rate-limiting.

In addition, the C-7=C-8 unsaturated double bond of p-coumaryl aldehyde (6) is in conjugation with both the allylic aldehyde moiety and the aromatic ring. Accordingly, the highly conserved Tyr53 of AtDBR1 is thus possibly in a stacking position with the phenolic ring of the substrate. Therefore, $\pi-\pi$ interactions between the phenolic substrate (6) and the nearby Tyr53 residue could potentially stabilize further the propenal transition state, i.e. by withdrawing electron density away from C-8 (Figures 4.8B and 4.8D). This potential interaction may also explain why AtDBR1 can catalyze p-coumaryl aldehyde (6) reduction more efficiently than 4-HNE (15), as indicated from the kinetic data (Table 4.3).

4.5 Concluding Remarks

With the discovery of the physiological function of PtPPDBR, it was possible to study the overall substrate versatility and catalytic mechanism of the *Arabidopsis* At5g16970 (AtDBR1).
This first revealed that $p$-coumaryl (6)/coniferyl (7) aldehydes were the preferred substrates over that of 4-HNE (15). Additionally, it identified Tyr260 (i.e. in AtDBR1) as responsible for general acid/base catalysis, and thus enabled us to define the overall catalytic mechanism of the alkenal (double bond) reductases in vascular plants. These findings also provide a much fuller understanding of the catalytic mechanism of the more distantly related mammalian DBR homologs such as 12-HD/PGR and AOR.
Figure 4.8. Observed potential interactions in the ternary complex of AtDBR1 with NADP$^+$/p-coumaryl aldehyde (structure 6) (A), NADP$^+$/4-HNE (structure 15) (B), and their corresponding schematic reaction mechanisms (C and D). Arrows indicate the stacking interaction between phenol rings. The phenolic group of Tyr-81 is within a hydrogen bond distance of the p-coumaryl aldehyde (6) phenolic group, and the phenolic group of Tyr-260 is within hydrogen bonding distance of the carbonyl group of both bound p-coumaryl aldehyde (6) and 4-HNE (15).

Careful inspection of the AtDBR1 substrate binding site in both the apo-form and the ternary complex also revealed that it has several factors favoring p-coumaryl aldehyde (6) as a substrate, even though it can accommodate (and reduce) other phenylpropanoids and 2-alkenals of longer/shorter chains. For example, since the substrate binding pocket is polar in nature, the presence of hydroxyl groups in p-coumaryl (6)/coniferyl (7) aldehydes, as well as in 4-HNE (15), enhance their binding affinities. By contrast, similar sized molecules, such as cinnamyl aldehyde (17), which lack a phenolic/hydroxyl group, have a much lower affinity for the pocket as revealed by the kinetic data. Additionally, although not depicted in the proposed catalytic mechanism, the phenolic group of the p-coumaryl aldehyde (6) is fully conjugated to the allylic aldehydic moiety, and this (due to a potentially reduced level of rotational freedom in the C-1-C-7 bond of cinnamyl aldehyde (17) versus p-coumaryl aldehyde (6)) may provide a further rationale as to why the former does not serve as a substrate. Finally, because of the relatively tight fit of the overall binding pocket, this might also explain why more highly substituted potential substrates, such as 5-hydroxyconiferyl aldehyde (18) and sinapyl aldehyde (19), are unable to undergo reductive conversions (i.e., when the phenolic group at C-4 is flanked by $o,o'$ substituents). Future work will be directed toward more comprehensively establishing the range of physiological functions of AtDBR1 and PtPPDBR, respectively.
4.6 ACKNOWLEDGMENTS

We thank C. Ogata (Advanced Photon Source, Northeastern Collaborative Access Team beam line) and C. Ralston (Advanced Light Source, beam line 8.2.1) for assistance.
4.7 REFERENCES


FOOTNOTES

*This research was supported in part by the National Institute of General Medical Sciences (to CHK and NGL), the National Science Foundation (MCB-9976684, MCB-0417291), the United States Department of Agriculture (Agricultural Plant Biochemistry #2006-03339), McIntire-Stennis, the Murdock Charitable Trust and the G. Thomas and Anita Hargrove Center for Plant Genomic Research. We thank C. Ogata (APS, beam line NE-CAT) and C. Ralston (ALS, beam line 8.2.1).

The cost of publication of this article was defrayed in part by the payment of page charges. This article must therefore be hereby-marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The abbreviations used are: AOR; alkenal/one oxidoreductase, AtDBR1; Arabidopsis thaliana double bond reductase 1, 12-HD/PGR; 12-hydroxydehydrogenase/15-oxo-prostaglandin 13-reductase, 4-HNA; 4-hydroxynonanal, 4-HNE; 4-hydroxy-(2E)-nonenal, MDR; medium-chain dehydrogenases/ reductases, PDB; protein data bank, PtPPDBR; P. taeda phenylpropenal (α,β double bond) reductase, r.m.s.d.; root mean square deviation.
CHAPTER FIVE
NEXT GENERATION SEQUENCING IN PREDICTING GENE FUNCTION
IN PODOPHYLLOTOXIN BIOSYNTHESIS*


Capsule:

Background: Biosynthetic pathways to structurally complex plant medicinals are incomplete or unknown.

Results: Next-generation sequencing and metabolomic analyses led to regio-specific methylenedioxy bridge-forming CyP450s in podophyllotoxin biosynthesis.

Conclusion: Unknown podophyllotoxin biosynthesis genes were identified using next-generation sequencing/bioinformatics analyses.

Significance: New podophyllotoxin biosynthesis step identified and database of several medicinal plant transcriptome assemblies and metabolic profilings made available for scientific community.

Abstract

Podophyllum species are sources of (−)-podophyllotoxin, an aryltetralin lignan used for semi-synthesis of various powerful and extensively employed cancer-treating drugs. Its biosynthetic pathway though remains largely unknown, with the last unequivocally demonstrated intermediate being (−)-matairesinol. Herein, massively parallel sequencing of Podophyllum
hexandrum and P. peltatum transcriptomes and subsequent bioinformatics analyses of the corresponding assemblies were carried out. Validation of the assembly process was first achieved through confirmation of assembled sequences with those of various genes previously established as involved in podophyllotoxin biosynthesis, as well as other candidate biosynthetic pathway genes. This contribution describes characterization of two of the latter, namely the cytochrome P450s, CYP719A23 from P. hexandrum and CYP719A24 from P. peltatum. Both enzymes were capable of converting (−)-matairesinol into (−)-pluviatolide by catalyzing methylenedioxy bridge formation, and did not act on other possible substrates tested. Interestingly, the enzymes described herein were highly similar to methylenedioxy bridge-forming enzymes from alkaloid biosynthesis, whereas candidates more similar to lignan biosynthetic enzymes were catalytically inactive with the substrates employed. This overall strategy has thus enabled facile further identification of enzymes putatively involved in (−)-podophyllotoxin biosynthesis, and underscores the deductive power of next generation sequencing and bioinformatics to probe and deduce medicinal plant biosynthetic pathways.

5.1 INTRODUCTION

Massively parallel sequencing technologies\(^1\) are rapidly evolving, and increasingly provide unprecedented opportunities to significantly enhance the understanding of biosynthetic processes, including those in important and yet poorly understood (non-model) medicinal plants. One main advantage is that such technologies can potentially lower the time frame for discovery of new genes and thus more rapidly improve our understanding of metabolism, e.g., when compared with more traditional approaches including labeled precursor administration, potential intermediate identification, enzyme purification and characterization, gene cloning, expressed sequence tag (EST)\(^2\) libraries, etc. In this context, several recent investigations have used these
massive parallel sequencing technologies to study a variety of non-model plants, with transcriptome assemblies mainly being generated from data from 454 and Illumina sequencing. Among others, these include *Panax quinquefolius* L.\(^2\), *Panax ginseng*\(^3\), *Gynostemma pentaphyllum*\(^4\), *Phalaenopsis* orchids\(^5\)\(^-\)\(^6\), *Camellia sinensis*\(^7\), *Catharanthus roseus*\(^8\), *Papaver somniferum*\(^9\), *Acacia auriculiformis* and *Acacia mangium*\(^10\), *Cicer arietinum*\(^11\) and *Abies balsamea*\(^12\). Although massive amounts of information can be obtained in this way, careful and informed analysis is required in order to help select candidate genes and carefully determine if they have a specific biosynthetic function of interest.

*podophyllum* species produce the aryltetralin lignan, (−)-podophyllotoxin (1b), that is of great medicinal importance due to its extensive use in the semi-synthesis of the anticancer drugs, teniposide (2), *etopophos*\(^\text{®}\) (3), and etoposide (4) (Figure 5.1). The latter are topoisomerase II inhibitors that are widely used for treating several cancers, including lung and testicular cancers\(^13\). However, as the main source of (−)-podophyllotoxin (1b), *Podophyllum hexandrum* is intensively collected and some reports suggest it has become endangered due to over-harvesting\(^14\).

While various synthetic chemical approaches to (−)-podophyllotoxin (1b) have been described, its production is not economical through such routes\(^15\)-\(^17\). An alternative approach that may be more productive is to obtain it in higher amount via biotechnological manipulation, whether in cell culture or in whole plants. Yet, this is currently not possible as our knowledge of the (−)-podophyllotoxin (1b) biosynthetic pathway is still incomplete. Nevertheless, following monolignol formation, the entry point in its biosynthetic pathway occurs via stereoselective coupling of two *E*-coniferyl alcohol (5) derived free radicals, involving the participation of dirigent proteins\(^18\)-\(^20\), to afford (+)-pinoresinol (6a) (Figure 5.2). The latter then undergoes
enantiospecific reduction, via action of pinoresinol/lariciresinol reductase, to sequentially afford (+)-lariciresinol (7a) and (−)-secoisolariciresinol (8b). Stereospecific dehydrogenation next converts the latter into (−)-matairesinol (9b), the last unequivocal known step in (−)-podophyllotoxin (1b) biosynthesis. Several of these known steps have been subsequently confirmed using *Podophyllum* and *Linum* species. On the other hand, putative downstream steps converting (−)-matairesinol (9b) into (−)-podophyllotoxin (1b) have only been reported using crude enzymatic assays; no genes have yet been identified or the enzymes purified to homogeneity.

**Figure 5.1.** (−)-Podophyllotoxin (1b) and its derivatives teniposide (2), etopophos (3) and etoposide (4), used in cancer treatment.
Figure 5.2. Possible biosynthetic pathway and/or grid leading to (−)-podophyllotoxin (1b) and related lignans. Known biosynthetic steps are highlighted in blue and the last known reaction catalyzed by CYP719A23 and CYP719A24 described in this work is in green.
The investigation herein describes the use of transcriptome sequencing using Illumina technologies and bioinformatics, together with metabolomic analysis, as a strategy to facilitate rapid gene discovery in (−)-podophyllotoxin (1b) biosynthesis. Specifically, this led to discovery of two new genes in *P. hexandrum* and *P. peltatum*, which encode enzymes capable of catalyzing methylenedioxy bridge formation through conversion of (−)-matairesinol (9b) into (−)-pluviatolide (14b).

5.2 EXPERIMENTAL PROCEDURES

*Plant Material*—*Podophyllum hexandrum* and *Podophyllum peltatum* plants were obtained from Digging Dog Nursery (Albion, CA) and Companion Plants (Athens, OH), respectively, and maintained in Washington State University greenhouse facilities.

*Chemicals*—(−)-Matairesinol (9b)28, (−)-arctigenin (34b) and (+)-phillygenin (38a)29 were isolated from *Forsythia intermedia*. (±)-Pinoresinsols (6a/b)22, (±)-7′-hydroxymatairesinol (10a/b)18, (±)-7-hydroxymatairesinol (32a/b), (±)-isoarctigenins (33a/b)29 and (±)-piperitols (37a/b)30 were synthesized as described (see Figure 5.3 for structures). (−)-α-Conidendrin (36b) and (−)-5-methoxymatairesinol (35b) were a gift from Dr. Eric P. Swan (Forentek, Canada), whereas (−)-α- and (−)-β-peltatins (20b and 27b) were obtained from Dr. Paul M. Dewick (University of Nottingham, United Kingdom). (−)-Podophyllotoxin (1b) was purchased from Sigma-Aldrich (see Figure 5.3 for structures).

*Metabolite Extraction and Analysis*—Rhizomes (for *P. hexandrum*), roots (for *P. peltatum*), stem and leaves (2 g, fresh weight) were individually harvested, immediately frozen in liquid nitrogen, ground to a fine powder and subsequently lyophilized. Each tissue was then successively sized via passage through a 150 μm sieve and extracted with 10 μl/mg methanol—
water (7:3, v/v) with the corresponding extracts maintained at –80 °C until analysis. Samples were analyzed by liquid chromatography using a Waters Acquity ultra performance liquid chromatography system equipped with a Waters BEH C18 column (1.7 µm particles, 2.1 × 50 mm) with a binary mobile phase of acetic acid–water (3:97) (A) and acetonitrile (B), with detection at 280 nm and by electrospray ionization mass spectrometry in the positive mode. The gradient program was as follows: flow rate of 0.3 ml/min; linear gradient of water with 0.1% formic acid: acetonitrile with 0.1% formic acid from 95:5 to 75:25 in 11 min, to 60:40 in 5 min, to 0:100 in 4 min, followed by 1.5 min at 0:100. The column temperature was held at 25 °C and sample injection volume was 5 µl. Masses were determined using a Waters Xevo G2 Q-TOF mass spectrometer and using leucine–enkephalin as a lock-mass standard.


(–)-α-Peltatin (20b): m/z 423.1059 ([M + Na]^+, 60%), calc. 423.1050; 418.1503 ([M + NH₄]^+, 45%), calc. 418.1496; 401.1236 ([M + H]^+, 71%), calc. 401.1231; and 247.0608 (100%), calc. 247.0601.

(–)-β-Peltatin (27b): m/z 415.1397 ([M + H]^+, 44%), calc. 415.1387; 247.0605 (100%), calc. 247.0601; and 203.0708 (1%), calc. 203.0703.

α-Peltatin-glucoside (42): $m/z$ 580.2030 ([M + NH₄]⁺, 3%), calc. 580.2025; 563.1763 ([M + H]⁺, 3%), calc. 563.1759; $m/z$ 409.1134 (33%), calc. 409.1129; and, 247.0603 (100%), calc. 247.0601.

Figure 5.3. Lignans tested as putative substrates in assays for CYP719A23 and CYP719A24 methylenedioxy bridge formation and (−)-haplomyrfolin (40b).


For relative abundance assessment of metabolites, integration was performed using specific ion of each compound: m/z 397.128 for (−)-podophyllotoxin (1b), m/z 247.035 for (−)-α- and (−)-β-peltatin (20b and 27b), m/z 594.183 for podophyllotoxin-glucoside (41), m/z 409.083 for α-peltatin glucoside (42), m/z 577.153 for β-peltatin glucoside (43) and m/z 383.078 for 4′- desmethylpodophyllotoxin (16) (see Figure 5.4 for structures).

RNA Extraction and cDNA Preparation—Total RNA was individually isolated from 100 mg of flash frozen plant rhizome (for P. hexandrum), roots (for P. peltatum), stems and leaves using a Qiagen (Valencia, CA) RNeasy Mini kit according to the manufacturer’s instructions, including the additional Qiagen clean-up protocol. An aliquot (1 µg) of each was subsequently used for cDNA preparation. After DNAse I (Invitrogen) treatment, cDNA was prepared using SuperScript® III First-Strand Synthesis System according to the manufacturer’s instructions and used for candidate gene amplification.

Transcriptome Sequencing and Library Assembly—Total RNA samples (~25 µg) dissolved in water were evaluated for integrity using the Bioanalyzer 2100 (Agilent, CA, USA), with samples having an RNA integrity number 31 (RIN) >5.0 processed further. RNA sample concentrations were estimated using a RiboGreen assay in a Qubit fluorometer, and each was then processed using either Illumina RNA-Seq or Illumina TruSeq RNA sample preparation kits,
according to Illumina protocols. Briefly, poly-A+ RNA was isolated from total RNA samples using oligo-d(T)25 magnetic beads (Dynabeads: Invitrogen, CA USA), then fragmented with the supplied reagents. First strand cDNA synthesis followed using random hexamers and the enzyme master mix provided. After second strand cDNA synthesis, the cDNA fragments were end-repaired by treatment with T4 DNA polymerase and Klenow fragment of *E. coli* DNA polymerase I, followed by addition of a single deoxyadenosine to the 3′ end of blunt-ended phosphorylated fragments. Sequencing adapters were attached with bacteriophage T4 ligase, followed by agarose gel electrophoresis with excision of fragments circa 500 bp in length. Each library DNA was purified from agarose using Qiagen QiaQuick gel extraction reagents, and subjected to 15 cycles of PCR. Amplified libraries were evaluated for quality and quantity using the Bioanalyzer 2100 and Nanodrop ND-1000 (Thermo Scientific, DE USA), respectively. Based on Nanodrop concentrations, libraries were normalized to 10 nM and accurate concentrations of sequenceable molecules were determined by qPCR using a reference library of known concentration as a standard. Flow cells were prepared on the Illumina Cluster Station and paired-end 54 bp sequence reads were obtained using an Illumina Genome Analyzer IIx instrument.

Initial read sets were examined for gross anomalies (e.g., over-represented reads due to library preparation issues), and for the presence of known sequencing artifacts, specifically phi X and known Illumina adapters. Read sets were then partitioned into collections of roughly one million reads, with these being quality trimmed using the FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) at the level of phrap Q = 10.

Cleaned paired-end read data were next subjected to multiple assemblies using ABBySS\(^{32}\), \(^{33}\), run in parallel mode, over a range of kmers with \(24 \leq k \leq 54\). Contigs thus generated were
named synthetic ESTs which in turn were created by performing multiple assemblies using different kmer sizes, pooling the resulting synthetic ESTs and, as described below, performing a subsequent assembly with a standard EST assembler. Typically this led to the production of at least twenty sets of synthetic ESTs.

Contigs resulting from each ABYSS assembly were scaffolded using the ABYSS scaffolder taking advantage of read pairing constraints. The NNN gap spacers inserted through scaffolding were resolved using GapCloser from the SOAPdenovo suite\textsuperscript{34}, and synthetic EST sets, per kmer, were constructed from the resulting scaffolds of at least 80 nucleotides. All read data was incorporated in the assembly producing an overall transcript reference.

**Final Assembly**—Pooled synthetic EST sets were assembled using MIRA in EST assembly mode\textsuperscript{35}. To control redundancy explicitly (at 98 percent sequence identity), the assembly results were processed with cd-hit\textsuperscript{36}. Resulting contigs of at least 100 nucleotides were reported as the final contig set for the build. Manual assessments of the longest contigs, and contigs with anomalously low or high read counts, were performed by inspection of the pileup-view using Tablet\textsuperscript{37}. Final assembly for all tissues and species investigated can be accessed in http://medplants.ncgr.org/.

**Bioinformatic Analysis**—Amino acid sequences of reference genes for shikimate, phenylpropanoid and lignan biosynthetic pathways were obtained from the NCBI database. Most genes were from *Arabidopsis thaliana*, except *Petunia × hybrida* for prephenate aminotransferase, *Nicotiana tabacum* for hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase, *Forsythia × intermedia* for dirigent protein and pinoresinol/lariciresinol reductase and *P. peltatum* for secoisolariciresinol dehydrogenase and dirigent protein, respectively (supplemental Table S5.1). Amino acid sequences of each reference gene were applied to tblastn
in BioEdit (Ver. 7.0.5.3, OCT. 28, 2005)\textsuperscript{38} to search homologous genes against either \textit{P. hexandrum} or \textit{P. peltatum} contig databases. Some contigs with high homology (with identity \( \geq 30\% \) and \( E \leq 5 \times 10^{-23} \)) against each reference gene were selected and the respective ORFs for each contig determined using ORF Finder (NCBI). ORFs were individually translated into amino acid sequence using EMBOSS Transeq (European Bioinformatics Institute) and each amino acid sequence translation applied to tblastn (NCBI) to confirm whether it corresponded to each target gene. The procedures mentioned above were carried out to avoid ambiguity of chimeric contigs\textsuperscript{39,40}. In addition, \( E \)-values and identity (%) of each candidate gene were calculated using blastp (NCBI) against the amino acid sequence of the corresponding reference genes. Unknown candidate genes for (\(-\))-podophyllotoxin (Ib) biosynthesis were selected using the same approach, based on known sequences for cinnamate 4-hydroxylase (CYP73A1), \( p \)-coumaroyl CoA 3-hydroxylase (CYP98A44), ferulate 5-hydroxylase (CYP84A3), flavonoid 6-hydroxylase (CYP71D9), flavonoid 3′-hydroxylase (CYP75A1), corytuberine-synthase (CYP80G2) as well as the methylenedioxy bridge-forming enzymes piperitol/nesamin synthase CYP81Q1\textsuperscript{41} and (\( S \))-canadine synthase CYP719\textsuperscript{42}, respectively (supplemental Table S5.2).

\textbf{Gene Cloning and Yeast Expression}—Candidates for methylenedioxy bridge-forming genes, selected as described above, were amplified from \textit{P. hexandrum} cDNA using the following primers: CYP719A23 forward (5′-CACCATGTCTATGGAGATGAGTGTC-3′) and CYP719A23 reverse (5′-TCAAGGATTGCGAGGAATGAT-3′) for the candidate homolog to CYP719 coded CYP719A23; CYP81B57 forward (5′-CACCATGTCTCCTCCAATTTAGTATTTTCTTC-3′) and CYP81B57 reverse (5′-CTACACCTGGAGAAAGGAAATTCAACA-3′) for the homolog to CYP81Q1 coded CYP81B57; CYP73A107 forward (5′-CACCATGTCTCTCCTCTCATCTTAGA-3′) and
CYP73A107 reverse (5′-TCAAAACACTCTGGGCTTAACA-3′), for CYP73A107. From *P. peltatum* cDNA, CYP719A24 was amplified using the following primers: CYP719A24 forward (5′-CACCATGTCTATGGAGACGAGTG-3′) and CYP719A24 reverse (5′-TCAAGGATTGCGAGGAATGAT-3′). Primers used for the amplification of the remaining candidates are in supplemental Table S5.2. Amplification was performed using *PfuTurbo* DNA Polymerase (Agilent PCR) in a thermocycler with 35 cycles of 94 °C denaturing for 30 s, 55 °C annealing for 30 s, and 70 °C extension for 3 min and a final extension for 10 min. PCR products were resolved in 1 % agarose gels, where single bands of approximately 1,500 bp were obtained.

Products were cloned into pENTR/D-TOPO (Invitrogen) and subsequently transferred to a yeast expression vector pYES-DEST52 (Invitrogen) according to the manufacturer’s instructions. From *P. hexandrum*, yeast expression clones pYES-DEST52::CYP719A23, pYES-DEST52::CYP73A107 were obtained and from *P. peltatum*, pYES-DEST52::CYP719A24. Each was subsequently individually introduced in the *Saccharomyces cerevisiae* strain WAT11, using the lithium acetate procedure according to the vector manufacturer’s instructions. An empty vector pYES-DEST52 was also introduced into WAT11 and used as a negative control.

Transformed yeasts were selected using synthetic complete media lacking uracil (SC-U) plates with 2% agar and 2% glucose. Single colonies of transformed yeast were spiked in liquid SC-U (10 ml) containing 2% glucose, then grown overnight until O.D. \(_{600}\) reached 3-4 and subsequently inoculated in Erlenmeyer flasks (1 l) containing induction media (200 ml SC-U with 1% raffinose and 2% galactose) to a final O.D. \(_{600}\) of 0.05. Inductions were carried out for 24 h at 30 °C in an orbital shaker at 300 rpm until cells were harvested for immediate microsome preparation.
Microsomal Preparations—After induction, cells for each candidate recombinant enzyme were harvested by centrifugation at $3,900 \times g$ for 10 min, then re-suspended in 3 ml per gram of cell weight Tris-HCl buffer (50 mM, pH 7.4) containing EDTA (1 mM), sorbitol (600 mM), DTT (0.1 mM) and PMSF (0.4 mM). Cells were disrupted using 0.5 mm glass beads (Biospec products, Inc.), with approximately half of the total volume added to cell suspensions in Falcon tubes (50 ml), by vortexing for $10 \times 30$ s at full speed with 30 s intervals on ice. Cell lysates were individually separated from glass beads by decantation, and glass beads were washed twice with half the total volume of buffer used initially to re-suspend the cells. Cell debris was removed by centrifugation at $15,000 \times g$ for 10 min, and microsomal preparations were individually obtained as gelatinous pellets after ultra-centrifugation of the supernatant at $91,000 \times g$ for 75 min. Each microsomal fraction was re-suspended in Tris-HCl buffer (50 mM, pH 7.4) containing EDTA (1 mM) and 30% glycerol (500 µl per gram of fresh weight of cell harvested) and homogenized using a Dounce homogenizer. Microsomal preparations were kept at –80 °C for up to 8 weeks with no detectable loss in activity.

Enzymatic Assays—Assays were performed in sodium phosphate buffer (200 µl, 100 mM and pH 7.5), with addition of a methanol solution of the substrates (10 µl) at the desired concentration (from 0.1 mM to 25 mM), followed by NADPH (50 mM, 10 µl) in sodium phosphate buffer (100 mM, pH 7.5) and finally the microsomal preparation (30 µl) with a protein concentration of ~60 µg/µl. Upon addition of each microsomal preparation reaction, the mixtures were vortexed and incubated at 25 °C for 5 min with constant shaking. Reactions were individually terminated by addition of glacial acetic acid (10 µl) and then centrifuged at 16,000 x g for 30 min. Aliquots of supernatant were then directly analyzed by ultra performance liquid chromatography using the same methodology described for metabolite analysis. For all kinetic
data, assays were performed in three independent experiments. Kinetic parameters ($K_m$ and $k_{cat}$) were estimated by non-linear least-squares data fitting\textsuperscript{44} and cytochrome P450 (CyP450) content in microsomes was determined by the reduced CO difference spectrum\textsuperscript{45}.

*Isolation of Enzymatic Product*— One hundred and fifty enzymatic assays using microsome preparations of yeast expressing CYP719A23 and with (–)-matairesinol (9b) as substrate were pooled together and the whole (~37.5 ml) extracted three times with chloroform (40 ml). The combined organic solubles were evaporated to dryness *in vacuo*, re-suspended in methanol (1 ml) and the enzymatically formed (–)-pluviatolide (14b) was next purified by HPLC using a SymmetryShield RP\textsubscript{18} column (Waters, 5 μm particle size, 3.9 × 150 mm) eluted as follows: flow rate of 1 ml/min; linear gradient of water and acetonitrile from 9:1 to 4:6 in 25 min, to 1:0 in 2.5 min, followed by 4.5 minutes at 1:0. Fractions containing (–)-pluviatolide (14b) were pooled freeze-dried and subjected to $^1$H, $^{13}$C, and $^1$H-$^{13}$C heteronuclear single quantum coherence NMR spectroscopic analyses using deuterated chloroform as solvent and tetramethylsilane as internal standard in a Varian VNMRS 600 MHz spectrometer (supplemental Table S5.3 and Figures S5.1 – S5.3).

(–)-*Pluviatolide* (14b): $\delta_H$ (CDCl\textsubscript{3}): 2.45-2.62 (4H, m); 2.89 (1H, dd, $J = 7.0$ and 14.1); 2.96 (1H, dd, $J = 5.2$ and 14.0); 3.85 (3H, s); 3.86 (1H, dd, $J = 7.4$ and 9.1); 4.11 (1H, dd, $J = 7.1$ and 9.2); 5.93 (1H, d, $J = 1.4$); 5.94 (1H, d, $J = 1.4$); 6.44-6.47 (2H, m); 6.63 (1H, dd, $J = 1.8$ and 7.9); 6.67 (1H, d, $J = 1.8$); 6.69 (1H, d, $J = 7.7$); 6.84 (1H, d, $J = 8$). $\delta_C$ (CDCl\textsubscript{3}) 178.64, 147.85, 146.65, 146.32, 144.52, 131.59, 129.43, 122.07, 121.55, 114.22, 111.48, 108.79, 108.31, 101.04, 71.19, 55.87, 46.59, 41.00, 38.30, 34.62. MS: $m/z$ 379.1155 ([M + Na]$^+$, calc. 379.1157), 357.1337 ([M + H]$^+$, calc. 357.1338), 339.1230 ([M + H – H\textsubscript{2}O]$^+$, calc. 339.1232), 161.0604 (calc. 161.0603), 137.0604 (calc. 137.0603) and 135.0445 (calc. 135.0446).
5.3 RESULTS AND DISCUSSION

Metabolite Profiling—First, metabolite profiling was carried out to ensure that target and biochemically related metabolites were present in the various tissues of the *Podophyllum* species investigated. Thus, utilizing ultra performance liquid chromatography-electrospray ionization-mass spectrometry, the extracts of rhizome (*P. hexandrum*), roots (*P. peltatum*), stem and leaves were examined. Based on metabolite UV, retention time and mass spectra, it was readily possible to detect and confirm the presence of the target metabolite, (−)-podophyllotoxin (1b). In *P. hexandrum*, it accumulates in higher amounts in the rhizome, being barely detectable in leaves and stem (Figure 5.4). In *P. peltatum*, the same trend was observed, with a higher accumulation in the roots although the stem and leaves also had (−)-podophyllotoxin (1b) contents closer to that in the roots. Its identification was performed by comparison with an authentic standard, having the same retention time as well as mass spectrum (Table 5.1 and Experimental Procedures). Other two related lignans, (−)-α- and (−)-β-peltatins (20b and 27b), were detected in different tissues of both species and also identified using authentic standards (Table 5.1 and Experimental Procedures). Interestingly, in *P. hexandrum*, the accumulation pattern of these two lignans was quite different from that observed for (−)-podophyllotoxin (1b), with (−)-α-peltatin (20b) being detected throughout all the different tissues. Conversely, there was a higher accumulation of (−)-β-peltatin (27b) in the aerial tissues, especially in leaves (Fig. 5.4A). In *P. peltatum*, on the other hand, both lignans accumulated in a somewhat similar pattern to that of (−)-podophyllotoxin (1b), with increasing amounts from leaves to stems and with the highest abundance in the roots (Figure 5.4B).
**TABLE 5.1.** Characteristic ions from detected lignans in positive mode mass spectral analyses. Base peak in bold.

<table>
<thead>
<tr>
<th>Lignan</th>
<th>Molecular Mass</th>
<th>ESI-MS m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>(−)-Podophyllotoxin (1b)</td>
<td>414</td>
<td>437, 432, 415, 397, 247</td>
</tr>
<tr>
<td>(−)-α-Peltatin (20b)</td>
<td>400</td>
<td>423, 418, 401, 247</td>
</tr>
<tr>
<td>(−)-β-Peltatin (27b)</td>
<td>414</td>
<td>415, 247, 203</td>
</tr>
<tr>
<td>Podophyllotoxin-glucoside (41)</td>
<td>576</td>
<td>599, 594, 397</td>
</tr>
<tr>
<td>α-Peltatin-glucoside (42)</td>
<td>562</td>
<td>580, 563, 409, 247</td>
</tr>
<tr>
<td>β-Peltatin-glucoside (43)</td>
<td>576</td>
<td>594, 577, 415, 409, 247</td>
</tr>
<tr>
<td>4′-Desmethypodophyllotoxin (16)</td>
<td>400</td>
<td>423, 418, 401, 383, 247</td>
</tr>
</tbody>
</table>

![Graph A](image)

![Graph B](image)

![Chemical structures](image)
Figure 5.4. Relative lignan contents in different tissues of *Podophyllum hexandrum* and *P. peltatum*. Lignans identified included (−)-podophyllotoxin (1b), (−)-α-peltatin (20b), (−)-β-peltatin (27b), podophyllotoxin-glucoside (41), α-peltatin-glucoside (42), β-peltatin-glucoside (43), and 4′-desmethylpodophyllotoxin (16) in both *P. hexandrum* (A) and *P. peltatum* (B) tissues. The individual lignan amounts are presented as relative peak areas (280 nm) with each given compound having a relative peak area of 100% for the most abundant amount and the others are reported as a percentage of that value.

Based on the mass spectroscopic analyses (Table 5.1 and Experimental Procedures), it was also possible to identify the known glycosylated forms of the aforementioned lignans: podophyllotoxin-glucoside (41), α-peltatin-glucoside (42), and β-peltatin-glucoside (43). It was also possible to detect 4′-desmethyl-podophyllotoxin (16) with mass identical to the isobaric (−)-α-peltatin (20b) but with the additional base peak resulting from loss of water [M + H – H$_2$O]$^+$ of m/z 383.1134 (calc. 383.1125). The accumulation pattern of podophyllotoxin-glucoside (41) and 4′-desmethylpodophyllotoxin (16) were similar to that of podophyllotoxin (1), with higher amounts in the underground tissues in both species, especially in *P. hexandrum*. As for (−)-α- and (−)-β-peltatin glucosides (42b and 43b), they accumulated in all tissues, with slightly higher levels in the leaves (Figure 5.4).

Based on the metabolite profiles observed for (−)-podophyllotoxin (1b) and the other related lignans described above, it could be considered that the overall functional biosynthetic pathway leading to these lignans might be present in all tissues in both species; on the other hand, specific hydroxylation enzymes regiospecifically placing hydroxyl groups, leading either to (−)-podophyllotoxin (1b) or to (−)-α- and (−)-β-peltatin (20b and 27b), might be tissue specific.

*Transcriptome Assembly and Analysis*—Upon confirmation and analysis of metabolites as compared to authentic standards, RNA was extracted from tissues and transcriptome data was
generated, as described in the Experimental Procedure section. From the crude RNA, poly-A+RNA was isolated, fragmented and converted into cDNA using random hexamers and end-paired to increase data quality for later assembly. After addition of adapters and quality control of the products, sequencing was performed using the Illumina Genome Analyzer IIX generating the read sets. These first read sets were then examined for known anomalies like significant sequence similarity to either the phi X genome or the Illumina adapter reference set, and those anomalies were discarded; the initial data was thus slightly reduced, typically by 1-5 %. The reads were then partitioned in paired data blocks in order to take advantage of parallel processing in the workflow and subsequently assembled using ABySS, which provides the ability to associate input files of read data with their source library. In this step, kmer size tended to be the most sensitive factor in the construction and analysis of the de Bruijn graph, and different choices often led to similar, yet different, assemblies. Contigs from these assemblies tended to be 100-500 bp long, and were treated as “synthetic ESTs”, and these were then assembled into the final contig set. It is important to stress that working with large quantities of data in this way presents a number of challenges, and for this reason each of the stages above includes a number of integrity checks (e.g., incomplete processing caused by system failures), basic data quality measurements (e.g., abundance of sequencing contaminants), and biological significance. The transcriptome database assembled for *P. hexandrum* and *P. peltatum* produced final databases of 227,885 and 147,960 contigs, respectively, including several complete and incomplete ORFs. The transcriptome data obtained is available alongside that of other important medicinal plants and can be accessed in the following website: http://medplants.ncgr.org.

Next, comparative analysis of metabolite profiles and transcriptome assembly data from each of the various tissues [rhizomes (for *P. hexandrum*), roots (for *P. peltatum*), stems and
leaves from both species] was carried out in order to: (a) verify whether known genes previously cloned from *Podophyllum peltatum*\(^\text{18, 22}\) were correctly assembled; (b) identify candidate genes in the shikimate-chorismate pathway to phenylalanine, the entry point into the phenylpropanoid pathway and in the core phenylpropanoid pathway leading to monolignols, such as coniferyl alcohol (5). The latter is also the entry point metabolite into the (−)-podophyllotoxin (1b) biosynthetic pathway; (c) conduct a bioinformatics analysis to identify potential candidate genes encoding steps beyond (−)-matairesinol (9b) and leading to the target compound, (−)-podophyllotoxin (1b).

Thus, assemblies were first interrogated in order to assess the validity of the presence of contigs corresponding to previously described genes and to obtain a first measure of the assembly quality. In our earlier studies of (−)-podophyllotoxin (1b) biosynthesis in *P. peltatum*, we had cloned and characterized a dirigent protein responsible for mediating the stereoselective coupling of *E*-coniferyl alcohol (5) leading to (+)-pinoresinol (6a)\(^\text{18}\) and a secoisolariciresinol dehydrogenase, responsible for the conversion of (−)-secoisolariciresinol (8b) into (−)-matairesinol (9b)\(^\text{22}\). When the assemblies obtained for *P. peltatum* were interrogated using these known sequences, contigs with very high identity (>98%) and low E value (<10\(^{-10}\)) to the previously described genes were observed (Table 5.2). This suggested that the assembly process was providing high quality data related to known genes; however, whether the sequences were fully correct or not (e.g., if they had point mutations, etc.) was not explored further, i.e., by confirmation through cloning, protein expression, etc.
TABLE 5.2. Blast results for known protein sequences from *P. peltatum* and *P. hexandrum*.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Query accession number/species of origin</th>
<th>Species/transcript</th>
<th>Identity (%)</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirigent protein</td>
<td>AAK38666.1/<em>P. peltatum</em></td>
<td>PpDir1_Pp27246$^a$</td>
<td>99.0</td>
<td>2 E$^{-115}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhDir1_Ph08051</td>
<td>92.0</td>
<td>4 E$^{-93}$</td>
</tr>
<tr>
<td>Pinoresinol/lariciresinol reductase</td>
<td>ACF71492.1/<em>P. hexandrum</em></td>
<td>PhPLR2_Ph140193</td>
<td>99.4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpPLR1_Pp37193</td>
<td>95.5</td>
<td>0</td>
</tr>
<tr>
<td>Secoisolariciresinol dehydrogenase</td>
<td>AAK38664.1/<em>P. peltatum</em></td>
<td>PpSDH1_Pp12640</td>
<td>99.3</td>
<td>3 E$^{-163}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhSDH1_Ph12248</td>
<td>97.5</td>
<td>2 E$^{-152}$</td>
</tr>
</tbody>
</table>

$^a$Pp, *P. peltatum*, Ph, *P. hexandrum*

The search for homologs in the shikimate/phenylpropanoid and monolignol forming pathways was also successful, with several homologs to each gene in both species identified. This included: 10 enzymes in the initial shikimate/chorismate pathway and 9 enzymes from the core phenylpropanoid pathway (supplemental Table S5.1) with identity $\geq 30\%$ and $E \leq 5 \times 10^{-23}$, as stated in the Experimental Procedure section. Overall, this successful search for homologs for all known genes from the shikimate/phenylpropanoid pathway that lead to the target lignans and several other important compounds (e.g., flavonoids, lignin, etc.) indicated that the assemblies obtained had a satisfactory coverage of the species transcriptomes. On the other hand, cloning and confirmation of the absolute accuracy of the assembly and the actual function of the corresponding gene was not carried out for non CyP450’s as this was outside the scope of the current study.

Even considering the highly encouraging results from this initial assessment of the assembled datasets, the potential limitations of this overarching approach need to be considered. This is because the assembly protocol is a tradeoff of the stringency of the assembly and the length of the transcripts obtained. It is thus always possible to find chimeric contigs, especially in cases of highly similar transcripts (close homologs) or in those cases where repeating elements...
are present. On the other hand, if the stringency of the assembly procedure is increased to minimize the presence of chimeric contigs, the result can be an increase in the number of incomplete and redundant (more than one contig for the same transcript) contigs. During the assembly process, several parameters were thus optimized to reach a balance, but in the final assembly examples of both could still be observed.

_CyP450’s in Podophyllotoxin Biosynthesis:_ Once the quality of the assemblies was evaluated and validated _in silico_, a search for unknown putative genes involved in the last steps in (−)-podophyllotoxin (1b) biosynthesis was then undertaken. The possible biochemical modifications required for conversion of (−)-matairesinol (9b) into (−)-podophyllotoxin (1b) are readily deduced (Figure 5.2), but the order of these reactions, on the other hand, has several possible permutations. Indeed, these could possibly either occur in parallel, in a biosynthetic “grid”, or through a specific sequence of conversions (Figure 5.2). To verify the role of possible substrates, it would be necessary to test a variety of compounds, most of which are not commercially available. From (−)-matairesinol (9b), the biosynthetic pathway leading to (−)-podophyllotoxin (1b) can, however, be expected to include several CyP450’s, involving two hydroxylations, one carbon-carbon (C–C) coupling/cyclization and methylenedioxy bridge formation, respectively (Figure 5.2). Additionally, while the C–C coupling and hydroxylation on the 7 position could be putatively performed by enzymes other than CyP450, (e.g., by a laccase in the first case and/or an oxoglutarate dependent dioxygenase for the latter), the formation of a methylenedioxy bridge functionality should be catalyzed by CyP450s. Accordingly, the first genes chosen for investigation were CyP450’s.

The CyP450’s are a very large (e.g., more than 200 genes in _A. thaliana_) and diverse family of pivotal importance in plant secondary metabolism49–51. Many are involved in
phenylpropanoid metabolism and have been identified, cloned, and characterized, from several plant species. More specifically, they can be employed in monolignol biosynthesis, where cinnamate-4-hydroxylase, \( p \)-coumaroyl CoA 3-hydroxylase and ferulate 5-hydroxylase are responsible for the successive hydroxylation of the \( \text{C}_6\text{C}_3 \) core\(^{52,53} \), as well as in the biosynthetic pathway leading to many downstream products, e.g., in flavonoid\(^54 \) and lignan\(^{41,55} \) biosynthesis.

The prediction of CyP450 physiological function is, however, frequently a difficult endeavor. This is because members of distinct gene families can perform similar functions and, in some cases, members of the same families can catalyze different reactions\(^51 \). To increase the probability of obtaining the presumed CyP450’s responsible for methylenedioxy bridge formation, several different known CyP450’s were used as templates for the search of homologs that could be involved in (−)-podophyllotoxin \((1b)\) and related phenylpropanoid biosynthesis. Initially focusing on \( P. \) hexandrum, its transcriptome was mined as described in the Experimental Procedure section for homologs to cinnamate 4-hydroxylase (CYP73A1), \( p \)-coumaroyl CoA 3-hydroxylase (CYP98A), ferulate 5-hydroxylase (CYP84A3), flavonoid 6-hydroxylase (CYP71D9), flavonoid 3′-hydroxylase (CYP75A1), corytuberine synthase (CYP80G2), (+)-\( \delta \)-canadinene 8′-hydroxylase(CYP706B1), (S)-canadine synthase (CYP719) and piperitol/sesamin synthase (CYP81Q), respectively. The highest homology contigs in the assembled transcriptome were then selected for cloning and further analysis (supplemental Table S5.2). All genes cloned matched transcriptome contig sequences perfectly (100%), therefore again validating the sequencing and assembly procedures employed.

Each of the above cloned candidates was heterologously expressed as described in the Experimental Procedure section for methylenedioxy bridge-forming enzymes and assayed against a range of possible substrates. The first genes to be tested encoded homologs of well-
established enzymes CYP73A107 (cinnamate 4-hydroxylase homolog), CYP98A68 (p-coumaroyl CoA 3-hydroxylase homolog), CYP84A52 (ferulate 5-hydroxylase homolog) and CYP71BE30 (flavonoid 6-hydroxylase homolog); these CyP450’s were assayed and found to carry out the anticipated enzymatic conversion (data not shown). CYP73A107 showed highest activity and was thus used as a positive control in all assays. We then proceeded to investigate the methylenedioxy bridge-forming steps of (−)-podophyllotoxin (1b) and related lignans.

*Methylenedioxy Bridge Formation in Podophyllotoxin Biosynthesis*—Methylenedioxy bridge formation by CyP450’s has been described in isoflavonoid\(^5\), lignan\(^41,5\) and alkaloid\(^42,5\) biosynthesis, with encoding genes cloned and characterized. Transcriptome data was interrogated looking for sequences similar to either CyP450 families and we were able to identify putative methylenedioxy bridge-forming enzyme homologs to both lignan (CYP81Q1 and CYP81Q2)\(^4\) and alkaloid (CYP719A1 and AY610513)\(^4,5\) biosynthesis. In the assembled *P. hexandrum* transcriptome, there was one full length transcript with ~50% identity to the former, coded CYP81B57 (supplemental Table S5.2), and another full length candidate with ~68% identity to the latter, coded CYP719A23 (Figure 5.5). Later, the *P. peltatum* assembled transcriptome was also interrogated and the transcript coded CYP719A24 was selected showing ~68% identity to CYP719A1 and ~96% identity to CYP719A23 (Figure 5.5).

These three candidates were cloned and their sequences confirmed by traditional Sanger sequencing\(^6\). Subsequently, each was individually expressed in yeast and the corresponding recombinant proteins assayed for methylenedioxy bridge formation using a range of potential substrates and analogs thereof including (−)-matairesinol (9b), the last confirmed intermediate in (−)-podophyllotoxin (1b) biosynthesis in *Podophyllum* species, (±)-7′-hydroxymatairesinols (10a/b), (±)-7-hydroxymatairesinols (32a/b), (−)-5-methoxymatairesinol (35b), (±)-
isoarctigenins (33a/b), (−)-arctigenin (34b), (±)-pinoresinos (6a/b), (±)-piperitols (37a/b), (+)-phillygenin (38a) and (−)-α-conidendrin (36b) (Fig. 5.3). Of these, (−)-7′-hydroxymatairesinol (10b) has been shown to be an intermediate in (−)-5-methoxypodophyllotoxin (39b) formation in *Linum flavum* and isoarctigenin (33) could also be an intermediate in (−)-podophyllotoxin (1b) biosynthesis, if (−)-matairesinol (9b) underwent methylation in the 4′ position before other modifications. The other analogs, (±)-pinoresinos (6a/b), (−)-arctigenin (34b), (−)-α-conidendrin (36b), (±)-piperitols (37a/b), and (+)-phillygenin (38a) have similar overall structure to possible intermediates and were tested in order to assess the enzymes’ substrate versatility.

![Diagram of CYP719A1, CYP719A23, and CYP719A24 proteins with sequence alignments.](image-url)
**Figure 5.5.** Sequence alignment of methylenedioxy bridge-forming cytochrome P450s. CYP719A23 and CYP719A24, cloned from *Podophyllum hexandrum* and *P. peltatum*, respectively, show ~68% identity to *Coptis japonica* (S)-canadine synthase (CYP719A1).

The first candidate evaluated was the homolog to the piperitol/sesamin synthase, namely CYP81B57. In our hands no catalytic activity could be observed, including towards piperitol/sesamin synthase natural substrates, (+)-pinoresinol (6a) and (+)-piperitol (37a) (data not shown).

We then proceeded to evaluate the methylenedioxy bridge-forming candidate homologs to the enzymes putatively annotated as involved in alkaloid biosynthesis. These (S)-canadine synthase homologs were found to act on (−)-matairesinol (9b), in the presence of NADPH, leading to the formation of a product with a longer retention time in ultra performance liquid chromatography analysis (Figure 5.6A), whose protonated molecular ion was at *m/z* 357 (Figure 5.6C), corresponding to the loss of two hydrogens as compared to (−)-matairesinol (9b) (Figure 5.6B). It was not possible to detect any activity and/or product formation, however, with the other putative substrates tested (data not shown), indicating that the (S)-canadine synthase homologs had a considerable degree of specificity towards (−)-matairesinol (9b). From the three possible products formed by a methylenedioxy bridge-forming enzyme (either with formation of one methylenedioxy bridge in either one of the aromatic rings or in both), the expected biosynthetic pathway reaction product to (−)-podophyllotoxin (1b) would be (−)-pluviatolide (14b) rather than (−)-haplomyrfolin (40b). In this respect, the fragmentation pattern of the enzymatic product was consistent to that described in the literature for (−)-pluviatolide (14b) (Figure 5.7). The product had a base peak of *m/z* 355.1183, corresponding to [M–H]⁻ (calc. 355.
Figure 5.6. Ultra-performance liquid chromatography-mass spectrometry analysis of enzymatic assays. A. Ultra performance liquid chromatography chromatogram showing product formation in CYP719A24 and CYP719A23 assays in comparison to negative controls (empty vector and cinnamate-4-hydroxylase, CYP73A107). Positive ion mass spectra of substrate (−)-matairesinol (9b, B) and product (−)-pluviatolide (14b, C), showing loss of two mass units.
Figure 5.7. Fragmentation pattern of (−)-pluviatolide (14b) and (−)-haplomyrfolin (40b). The expected fragments generated by the two isobaric compounds 14b (A) and 40b (B) during LC-ESI-MS analyses are shown [adapted from Schmidt et al. (61)]. The fragment at m/z 161 and the absence of a fragment at m/z 163 points to (−)-pluviatolide (14b) as the products of CYP719A23 and CYP719A24.

1182) in the negative ion mode (data not shown). The positive ion mode spectrum was more informative (Figure 5.6C), showing a base peak of m/z 339.1230 corresponding to [M + H – H₂O]⁺ (calc. 339.1232), and further peaks of m/z 357.1337 and m/z 379.1155, corresponding to [M + H]⁺ (calc. 357.1338) and [M + Na]⁺ (calc. 379.1157), respectively. More importantly, it was possible to observe the predicted fragmentation that unequivocally indicated formation of the methylenedioxy bridge: the characteristic substituted tropylium cation corresponding to the methylenedioxy group with m/z 135.0445 (calc. 135.0446) as well as the hydroxy-methoxy substituted fragment with m/z 137.0604 (calc. 137.0603). Finally, a fragment of m/z 161.0604 was also observed that allowed the distinction between the two potential isobaric products, (−)-
pluviatolide (14b) and (−)-haplomyrfolin (40b) (Figure 5.7): this corresponds to the allylbenzodioxole fragment (calc. 161.0603) and shows that the methylenedioxy bridge was formed in the expected ring forming (−)-pluviatolide (14b), as the alternative product (−)- haplomyrfolin (40b) would produce a 4-allyl-2-methoxyphenol fragment of m/z 163.0754 (Figure 5.7). Therefore, the product obtained had the methylenedioxy bridge introduced into the correct aromatic ring. Analysis of the 1H, 13C, and HSQC NMR spectra of the enzymatically obtained product (see Supplementary material) also clearly supports the formation of the product, being in accordance with literature values63, e.g., with the methylenedioxy bridge characteristic peaks at 5.9 and 101 ppm in 1H- and 13C-NMR spectra, respectively (supplemental Table S5.3 and Figures S5.1 – S5.3. Kinetic Data of Putative Pluviatolide Synthase—Using (−)-matairesinol (9b) as a substrate, microsomes from S. cerevisiae expressing the P. hexandrum CYP719A23 showed a saturation curve consistent with Michaelis-Menten kinetics with a $K_m$ of 9.7 ± 2.2 µM (Figure 5.8 A) and $k_{cat}$ of 14.9 ± 1.0 min$^{-1}$. In addition, the homolog cloned from P. peltatum, CYP719A24, displayed an apparent $K_m$ of 5.8 ± 1.4 µM (Figure 5.8 B) and $k_{cat}$ of 7.8 ± 0.4 min$^{-1}$. This tight binding towards (−)-matairesinol (9b) indicated that these CyP450s from both species can be provisionally presumed to be those involved in the methylenedioxy bridge formation leading to (−)-podophyllotoxin (1b) and pathway related lignans. Still, in order to determine if (−)-pluviatolide (14b) is the true biosynthetic intermediate, this will ultimately require in vivo verification e.g., by down-regulation or knocking-out of this gene and identifying corresponding changes in metabolite profile. However, to date, there is no system in place to transform Podophyllum species.

It is very interesting that the candidates with higher homology to enzymes from alkaloid biosynthesis, the CYP719A1 (S)-canadine synthase (Figure 5.9), were found to be capable of
performing the same reaction in lignan metabolism, while the candidate CYP81B57, closely related to the lignan piperitol/sesamin synthase from the Pedaliaceae, *Sesamum indicum*, had no activity detected. *Podophyllum* species, however, belong to the Berberidaceae family, and Ranunculales order, the only family with CYP719 genes described thus far. The *Podophyllum* species also form a closely related phylogenetic group with other species known for biosynthesizing similar aryltetralin lignans but with no alkaloids in them reported so far, such as *Dysosma* and *Diphyleia* species. This clade forms a monophyletic group with many benzylisoquinoline alkaloid producing species, including those from which other CYP719s have been described. It can thus be tentatively postulated that this *Podophyllum* group has either no or strongly reduced alkaloid level of biosynthesis, while “recruiting” some of its genes for the podophyllotoxin pathway.
Figure 5.8. Kinetic parameters for CYP719A23 and CYP719A24. Steady state Michaelis-Menten kinetics derived from initial rates of CYP719A23 (A) and CYP719A24 (B) enriched microsomes with (−)-matairesinol (9b) as substrate. Assays were performed in triplicates.

Figure 5.9. Phylogenetic analysis of cloned and known cytochrome P450 enzymes illustrating the similarity of the cloned genes to known alkaloid biosynthetic enzymes. Phylogenetic tree generated by sequence alignment using ClustaW (version 1.4), constructed with neighboring joining clustering algorithm (version 3.5c) and visualized with Tree view (version 1.6.6). Amino acid sequences were obtained from UniProtKB, SwissProt or GenBank™ with the following accession numbers: Q948Y1, CYP719A1 [(S)-canadine synthase], Coptis japonica; EU882969.1, CYP719A2 and AB126256.1, CYP719A3 (stylophone synthase), Eschscholzia californica; EU883001.1, CYP719A4 (canadine synthase) Thalictrum flavum; EF451151, CYP719A13 (stylophone synthase), Argemone mexicana; EF451152, CYP719A14 (cheilanthifoline synthase), A. mexicana; AB374407, CYP719A18, C. japonica; AB374408, CYP719A19, C. japonica; EF451150, CYP719B1 (salutaridine synthase), P. somniferum; U09610,
CYP80A1 (berbamunine synthase), *Berberis stolonifera*; AF014801, CYP80B1V2 [(S)-N-methylcoclaurine 3′-hydroxylase], *E. californica*; AB025030, CYP80B2 [(S)-N-methylcoclaurine 3′-hydroxylase], *C. japonica*; AY610509, CYP80B4 [(S)-N-methylcoclaurine 3′-hydroxylase], *T. flavum*; AB288053, CYP80G2 [(S)-corytuberine synthase], *C. japonica*; P93147, CYP81E1 (isoflavone 2′-hydroxylase), *Glycyrrhiza echinata*; AY278229, CYP81E8 (isoflavone 2′-hydroxylase), *Medicago truncatula*; BAE48234, CYP81Q [(+)–piperitol/(+)–sesamin synthase], *Sesamum indicum*. A 10% change is indicated by the scale bar.

5.4 CONCLUSIONS

In this work, the utility of massively parallel sequencing was demonstrated for the study of non-model *Podophyllum* medicinal plants. The Illumina based technology produced abundant high quality data, with 100% agreement with sequences obtained from the cloned transcripts obtained using traditional Sanger sequencing. In addition to verify the validity of known sequences in the lignan pathway leading to (−)-podophyllotoxin (1b), the two CyP450s from *P. hexandrum* and *P. peltatum* studied herein are putative pluviatolide-synthases, i.e., capable of catalyzing methylenedioxy bridge formation for (−)-podophyllotoxin (1b) biosynthesis.

As an extension of this work and as a resource for the scientific community, the transcriptome data and metabolic profiling of the species investigated in herein and several other important medicinal plants are available for the community in the open websites [http://uic.edu/pharmacy/MedPlTranscriptome/index.html](http://uic.edu/pharmacy/MedPlTranscriptome/index.html) and [http://medplants.ncgr.org](http://medplants.ncgr.org).

5.5 ACKNOWLEDGEMENTS

We thank Prof. David Nelson (University of Tennessee) for assigning standard nomenclature to the CyP450 described herein and Amy Hetrick (Institute of Biological Chemistry) for growing the *Podophyllum* and *Linum* plants. We also gratefully acknowledge the assistance of Mr. Robin Kramer (NCCR) in the production of transcript assemblies used in this project.
5.6 REFERENCES


FOOTNOTES

* We thank the National Institutes of Health, National Institute of General Medical Sciences (1RC2GM092561), the National Science Foundation (MCB-1052557), and the G. Thomas and Anita Hargrove Center for Plant Genomic Research, for generous financial support.

(1) To whom correspondence should be addressed. Tel.: 509-335-2682; Fax: 509-335-8206; E-mail: lewisn@wsu.edu.

(2) The abbreviations used are: CyP450, cytochrome P450; EST, expressed sequence tag.
### TABLE S5.1. Shikimate, phenylpropanoid, and downstream lignan biosynthetic gene homologs search results for *P. peltatum* and *P. hexandrum*

<table>
<thead>
<tr>
<th>Shikimate/chorismate pathway</th>
<th>Species</th>
<th>Genes</th>
<th>Contig number</th>
<th>Identity (%)</th>
<th>E value</th>
<th>Origin of Reference genes</th>
<th>Accession Number of Reference genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Decoxy-D-arabinoheptulosonate-7-phosphate synthase (DAHPS)</td>
<td><em>Podophyllum hexandrum</em></td>
<td>PhDAHPS1</td>
<td>Ph000404</td>
<td>84.1</td>
<td>5E-31</td>
<td><em>Arabidopsis thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhDAHPS2</td>
<td>Ph18262</td>
<td>72.6</td>
<td>1E-122</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpDAHPS1</td>
<td>Pp02530</td>
<td>75.4</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDAHPS2</td>
<td>Pp29545</td>
<td>78.1</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDAHPS3</td>
<td>Pp05894</td>
<td>75.4</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDAHPS4</td>
<td>Pp10403</td>
<td>79.6</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDAHPS5</td>
<td>Pp06872</td>
<td>74.9</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDAHPS6</td>
<td>Pp22183</td>
<td>72.6</td>
<td>1E-36</td>
<td><em>A. thaliana</em></td>
<td>NM_120162.4</td>
</tr>
<tr>
<td>Dehydroquininate synthase (DQS)</td>
<td><em>P. hexandrum</em></td>
<td>PhDQS</td>
<td>P04590</td>
<td>85.4</td>
<td>1E-37</td>
<td><em>A. thaliana</em></td>
<td>NM_180948.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpDQS1</td>
<td>Pp06499</td>
<td>85.7</td>
<td>3E-28</td>
<td><em>A. thaliana</em></td>
<td>NM_180948.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDQS2</td>
<td>Pp13672</td>
<td>73.3</td>
<td>6E-42</td>
<td><em>A. thaliana</em></td>
<td>NM_180948.1</td>
</tr>
<tr>
<td>3-Decroxy-L-arabinoheptulosonate-7-phosphate synthase (DAHPS)</td>
<td><em>P. hexandrum</em></td>
<td>PhDHQDSD</td>
<td>Ph20177</td>
<td>67.6</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>BT012197.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpDHQDSD1</td>
<td>Pp06065</td>
<td>71.6</td>
<td>1E-148</td>
<td><em>A. thaliana</em></td>
<td>BT012197.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDHQDSD2</td>
<td>Pp23328</td>
<td>59.7</td>
<td>7E-73</td>
<td><em>A. thaliana</em></td>
<td>BT012197.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpDHQDSD3</td>
<td>Pp03544</td>
<td>60.6</td>
<td>1E-99</td>
<td><em>A. thaliana</em></td>
<td>BT012197.1</td>
</tr>
<tr>
<td>Shikimate kinase (SK)</td>
<td><em>P. hexandrum</em></td>
<td>PhSK1</td>
<td>P06031</td>
<td>62.2</td>
<td>7E-63</td>
<td><em>A. thaliana</em></td>
<td>BT032871.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhSK2</td>
<td>P07147</td>
<td>30.9</td>
<td>5E-21</td>
<td><em>A. thaliana</em></td>
<td>BT032871.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpSK1</td>
<td>Pp10907</td>
<td>66.7</td>
<td>7E-56</td>
<td><em>A. thaliana</em></td>
<td>BT032871.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpSK2</td>
<td>Pp11672</td>
<td>46.0</td>
<td>1E-59</td>
<td><em>A. thaliana</em></td>
<td>BT032871.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpSK3</td>
<td>Pp15169</td>
<td>50.0</td>
<td>1E-63</td>
<td><em>A. thaliana</em></td>
<td>BT032871.1</td>
</tr>
<tr>
<td>5-Enolpyruvylshikimate-3-phosphate synthase (EPSPS)</td>
<td><em>P. hexandrum</em></td>
<td>PhEPSPS1</td>
<td>Ph08207</td>
<td>75.6</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhEPSPS2</td>
<td>Ph13491</td>
<td>85.4</td>
<td>8E-125</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpEPSPS1</td>
<td>Pp11679</td>
<td>74.8</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpEPSPS2</td>
<td>Pp15664</td>
<td>84.6</td>
<td>0</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpEPSPS3</td>
<td>Pp02437</td>
<td>72.1</td>
<td>7E-134</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpEPSPS4</td>
<td>Pp78080</td>
<td>86.9</td>
<td>3E-130</td>
<td><em>A. thaliana</em></td>
<td>BT022026.1</td>
</tr>
<tr>
<td>Chorismate synthase (CS)</td>
<td><em>P. hexandrum</em></td>
<td>PhCS</td>
<td>P05866</td>
<td>77.5</td>
<td>6E-156</td>
<td><em>A. thaliana</em></td>
<td>NM_1036081.2</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpCS1</td>
<td>Pp02628</td>
<td>77.8</td>
<td>2E-156</td>
<td><em>A. thaliana</em></td>
<td>NM_1036081.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpCS2</td>
<td>Pp51178</td>
<td>74.0</td>
<td>8E-148</td>
<td><em>A. thaliana</em></td>
<td>NM_1036081.2</td>
</tr>
<tr>
<td>Chorismate mutase (CM)</td>
<td><em>P. hexandrum</em></td>
<td>PhCM1</td>
<td>Ph09196</td>
<td>67.5</td>
<td>8E-115</td>
<td><em>A. thaliana</em></td>
<td>BT024732.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpCM1</td>
<td>Pp05518</td>
<td>62.8</td>
<td>2E-96</td>
<td><em>A. thaliana</em></td>
<td>BT024732.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpCM2</td>
<td>Pp03866</td>
<td>57.9</td>
<td>6E-92</td>
<td><em>A. thaliana</em></td>
<td>BT024732.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpCM3</td>
<td>Pp13075</td>
<td>52.0</td>
<td>1E-43</td>
<td><em>A. thaliana</em></td>
<td>BT024732.1</td>
</tr>
<tr>
<td>Aromatic amino acid pathway</td>
<td><em>Phenolphthalein-3-phosphate synthase</em> (PAA)</td>
<td><em>P. hexandrum</em></td>
<td>PpPAAT</td>
<td>P03484</td>
<td>80.4</td>
<td><em>Petunia x hybrida</em></td>
<td>HM635905.1</td>
</tr>
<tr>
<td></td>
<td><em>P. peltatum</em></td>
<td>PpPAAT1</td>
<td>Pp04526</td>
<td>80.0</td>
<td>0</td>
<td><em>Petunia x hybrida</em></td>
<td>HM635905.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpPAAT2</td>
<td>Pp18831</td>
<td>71.7</td>
<td>0</td>
<td><em>Petunia x hybrida</em></td>
<td>HM635905.1</td>
</tr>
<tr>
<td>Arogenate dehydratases</td>
<td><em>P. hexandrum</em></td>
<td>PhADT1</td>
<td>Ph13631</td>
<td>66.2</td>
<td>3E-31</td>
<td><em>A. thaliana</em></td>
<td>NM_101051.2</td>
</tr>
<tr>
<td>Phenylpropanoid pathway</td>
<td>Phenylalanine ammonia lyase (PAL)</td>
<td>P. hexandrum</td>
<td>PhPAL1</td>
<td>Ph14570</td>
<td>82.8</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. peltatum</td>
<td>PpPAL1</td>
<td>Pp15454</td>
<td>83.4</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpPAL2</td>
<td>Pp00823</td>
<td>82.8</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td>Cinnamate 4-hydroxylase (C4H)</td>
<td>P. hexandrum</td>
<td>PhC4H1</td>
<td>Ph35992</td>
<td>91.4</td>
<td>5 E^-39</td>
<td>A. thaliana</td>
<td>AY065145.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. peltatum</td>
<td>PpC4H1</td>
<td>Pp07790</td>
<td>83.1</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpC4H2</td>
<td>Pp02595</td>
<td>71.1</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpC4H3</td>
<td>Pp138661</td>
<td>71.0</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td>4-Coumarate CoA ligase (4CL)</td>
<td>P. hexandrum</td>
<td>Ph4CL</td>
<td>Ph15155</td>
<td>61.5</td>
<td>0</td>
<td>A. thaliana</td>
<td>NM_104046.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. peltatum</td>
<td>Pp4CL1</td>
<td>Pp05144</td>
<td>72.8</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pp4CL2</td>
<td>Pp00905</td>
<td>60.6</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pp4CL3</td>
<td>Pp05741</td>
<td>56.7</td>
<td>9 E^-15</td>
<td>A. thaliana</td>
</tr>
<tr>
<td>Hydroxycinnamoyl-CoA shikimate hydroxycinnamoyltransferase (HCT)</td>
<td>P. hexandrum</td>
<td>PhHCT1</td>
<td>Ph06039</td>
<td>80.3</td>
<td>0</td>
<td>Nicotiana tabacum</td>
<td>AJ507825.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. peltatum</td>
<td>PpHCT1</td>
<td>Pp07896</td>
<td>80.3</td>
<td>0</td>
<td>N. tabacum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpHCT2</td>
<td>Pp04822</td>
<td>58.9</td>
<td>5 E^-72</td>
<td>N. tabacum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpHCT3</td>
<td>Pp04364</td>
<td>63.5</td>
<td>8 E^-51</td>
<td>N. tabacum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpHCT4</td>
<td>Pp05997</td>
<td>51.7</td>
<td>1 E^-40</td>
<td>N. tabacum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpHCT5</td>
<td>Pp06413</td>
<td>56.5</td>
<td>1 E^-40</td>
<td>N. tabacum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PpHCT6</td>
<td>Pp42575</td>
<td>50.2</td>
<td>4 E^-94</td>
<td>N. tabacum</td>
</tr>
<tr>
<td>p-Coumarate 3-hydroxylase (pC3H)</td>
<td>P. hexandrum</td>
<td>PpC3H</td>
<td>Pp08223</td>
<td>74.1</td>
<td>0</td>
<td>A. thaliana</td>
<td>NM_180006.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P. peltatum</td>
<td>PppC3H1</td>
<td>Pp07398</td>
<td>78.2</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PppC3H2</td>
<td>Pp01710</td>
<td>74.7</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PppC3H3</td>
<td>Pp28051</td>
<td>66.1</td>
<td>0</td>
<td>A. thaliana</td>
</tr>
<tr>
<td>Hydroxycinnamoyl CoA O-methyl- transferase (CCoAOMT)</td>
<td>P. hexandrum</td>
<td>PhCCoAOMT1</td>
<td>Ph10963</td>
<td>71.8</td>
<td>8 E^-90</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhCCoAOMT2</td>
<td>Ph26362</td>
<td>73.2</td>
<td>1 E^-66</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhCCoAOMT3</td>
<td>Ph39179</td>
<td>57.3</td>
<td>2 E^-78</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
</tr>
<tr>
<td></td>
<td>Accession</td>
<td>E-Value</td>
<td>Species</td>
<td>Genbank ID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------</td>
<td>---------</td>
<td>---------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCCoAOMT4</strong></td>
<td>Ph18352</td>
<td>55.6</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCCoAOMT5</strong></td>
<td>Ph02950</td>
<td>53.9</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P. peltatum</strong></td>
<td>PpCCoAOMT1</td>
<td>86.1</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PpCCoAOMT2</td>
<td>67.2</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PpCCoAOMT3</td>
<td>70.0</td>
<td>A. thaliana</td>
<td>AY143979.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCCR1</strong></td>
<td>Ph13780</td>
<td>52.5</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCCR2</strong></td>
<td>Ph224439</td>
<td>52.1</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCCR3</strong></td>
<td>Ph06201</td>
<td>48.7</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCCR1</strong></td>
<td>Pp05054</td>
<td>76.8</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCCR2</strong></td>
<td>Pp08870</td>
<td>60.4</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCCR3</strong></td>
<td>Pp06256</td>
<td>51.2</td>
<td>A. thaliana</td>
<td>AF321114.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCOMT1</strong></td>
<td>Ph04497</td>
<td>54.3</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCOMT2</strong></td>
<td>Ph01260</td>
<td>55.1</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCOMT3</strong></td>
<td>Ph13451</td>
<td>50.8</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCOMT4</strong></td>
<td>Ph07881</td>
<td>50.9</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCOMT1</strong></td>
<td>Pp08491</td>
<td>68.2</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCOMT2</strong></td>
<td>Pp05762</td>
<td>53.6</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCOMT3</strong></td>
<td>Pp09983</td>
<td>52.6</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCOMT4</strong></td>
<td>Pp02858</td>
<td>52.0</td>
<td>A. thaliana</td>
<td>AY081565.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhF5H1</strong></td>
<td>Ph03439</td>
<td>65.9</td>
<td>A. thaliana</td>
<td>NM_119790</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhF5H2</strong></td>
<td>Ph06572</td>
<td>36.8</td>
<td>A. thaliana</td>
<td>NM_119790</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpF5H1</strong></td>
<td>Pp02250-1</td>
<td>61.7</td>
<td>A. thaliana</td>
<td>NM_119790</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpF5H2</strong></td>
<td>Pp02250-2</td>
<td>72.7</td>
<td>A. thaliana</td>
<td>NM_119790</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD1</strong></td>
<td>Ph04622</td>
<td>61.9</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD2</strong></td>
<td>Ph05963</td>
<td>52.3</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD3</strong></td>
<td>Ph07601</td>
<td>50.6</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD4</strong></td>
<td>Ph08319</td>
<td>47.2</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD5</strong></td>
<td>Ph07288</td>
<td>46.4</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PhCAD6</strong></td>
<td>Ph07772</td>
<td>46.5</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCAD1</strong></td>
<td>Pp04201</td>
<td>74.8</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCAD2</strong></td>
<td>Pp07432</td>
<td>58.9</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCAD3</strong></td>
<td>Pp28826</td>
<td>54.4</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCAD4</strong></td>
<td>Pp03284</td>
<td>52.0</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PpCAD5</strong></td>
<td>Pp11902</td>
<td>49.3</td>
<td>A. thaliana</td>
<td>AY302082.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table S5.2. CyP450 genes selected for cloning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP73A1</strong> (cinnamate 4-hydroxylase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP73A107</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-AACACCCACACCACCCAGTG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-TGTCCTTACTCAAACACTCTGGG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP98A44</strong> (p-coumarate 3-hydroxylase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP98A68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-CCATGGCTCTCTCTGTTCT-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-TCATACGGCCCATG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP84A3</strong> (ferulate 5-hydroxylase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP84A52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-AGATGGAGTTTGCATACCAAG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-TCACTCAGTGCAAAAGCCG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP71D9</strong> (flavonoid-6-hydroxylase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP71BE30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-CGATGGATCTTCGTCATTC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-CTAACAGGACCAGTGACTTATAA-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP75A1</strong> (flavonoid-3′-hydroxylase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP75B65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-CCATGGTCCTCTAGAGCTCCTCTCTC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-TTCAGATGACGGAAGGCAACC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP80G2</strong> (corybubrine-synthase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP80G3 (fragment)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-AGATGGATCCTGTCAGGAGCA-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-CTACACCCCTGGATTTCGGAATG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP81Q1</strong> (piperitol/sesamin synthase) homolog (fragment):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP81B56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-ATGAGTATCGGACGGTGAG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-GGAGCATATCGAAATTACTAACC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP81Q1</strong> (piperitol/sesamin synthase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP82D61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-CCTCCCCACACTTTGTC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-GCATAATCAAAGGATAGTATCGAAAAC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP81Q1</strong> (piperitol/sesamin synthase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP81B57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-ATGGGAACAAATATTAGTATTTCC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-CTACACCTGAGAAAGGAAATTCAAC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP719A1</strong> ((S)-canadine synthase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP719A23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-GTGGTATCTAGTAGGGAGATG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-CAAGAAAAGGGCAGGAAAAC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CYP719A1</strong> ((S)-canadine synthase) homolog:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP719A24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Primer: 5'-GCATACTGAGTGAGGAGTG-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reverse Primer: 5'-TCAAAAGGATTAC-3'</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>$^1$H (ppm)</td>
<td>$^{13}$C (ppm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>129.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6.67 (d, J = 1.8)</td>
<td>111.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>146.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>144.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.84 (d, J = 8)</td>
<td>114.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6.63 (dd, J = 1.8, J = 7.9)</td>
<td>121.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>2.89 (dd, J = 7.0, J = 14.1)</td>
<td>34.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b</td>
<td>2.96 (dd, J = 5.2, J = 14.0)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.45-2.62 (m)</td>
<td>46.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>178.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1'</td>
<td>-</td>
<td>131.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2'</td>
<td>6.44-6.47 (m)</td>
<td>108.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td>-</td>
<td>147.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4'</td>
<td>-</td>
<td>146.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>6.69 (d, J = 7.7)</td>
<td>108.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>6.44-6.47 (m)</td>
<td>122.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7'</td>
<td>2.45-2.62 (m)</td>
<td>38.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8'</td>
<td>2.45-2.62 (m)</td>
<td>41.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9’a</td>
<td>3.86 (dd, J = 7.4, J = 9.1)</td>
<td>71.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9'b</td>
<td>4.11 (dd, J = 7.1, J = 9.2)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCH$_2$O a</td>
<td>5.93 (1d, J = 1.4)</td>
<td>101.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCH$_2$O b</td>
<td>5.94 (d, J = 1.4)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMe</td>
<td>3.85 (s)</td>
<td>55.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE S5.1. (−)-Pluviatolide (14b) (CDCl₃) 600MHz-¹H-NMR
FIGURE S5.2. (−)-Pluviatolide (14b) (CDCl₃) 600 MHz-¹³C-NMR

ppm (t1)

178.64 147.85 146.65 146.32 144.82 131.98 129.43 122.07 121.55 114.22 111.48 108.31 101.04

71.19 55.87 46.59 41.00 38.30 34.62
FIGURE S5.3. (−)-Pluviatolide (14b) (CDCl$_3$) 600 MHz $^1$H-$^{13}$C-HSQC-NMR