

**Functional characterization of 4-Coumarate-CoA
ligases involved in phenylpropanoid biosynthesis
from *Ocimum kilimandscharicum***

By

Lavhale Santosh Govind

Registration Number : 10BB15A26044

A thesis submitted to the
Academy of Scientific & Innovative Research
for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Sciences

Under the guidance of

Dr. Ashok P. Giri



CSIR-National Chemical Laboratory, Pune



Academy of Scientific and Innovative Research

AcSIR Headquarters, CSIR-HRDC campus

Sector 19, Kamla Nehru Nagar,

Ghaziabad, U.P. – 201 002, India

March 2022

*This thesis is dedicated
to my
beloved parents*

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, **“Functional characterization of 4-Coumarate-CoA ligases involved in phenylpropanoid biosynthesis from *Ocimum kilimandscharicum*”**, submitted by **Lavhale Santosh Govind** to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of Doctor of Philosophy in sciences embodies original research work carried-out by the student. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.



(Signature of Student)
Lavhale Santosh Govind
Date: 14/03/2022



(Signature of Supervisor)
Dr. Ashok P. Giri
14/04/2022

STATEMENTS OF ACADEMIC INTEGRITY

I Lavhale Santosh Govind, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10BB15A26044 hereby undertake that, the thesis entitled “**Functional characterization of 4-Coumarate-CoA ligases involved in phenylpropanoid biosynthesis from *Ocimum kilimandscharicum***” has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on “*Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)*” and the CSIR Guidelines for “*Ethics in Research and in Governance (2020)*”.



Signature of the Student

Date : 14/03/2022

Place : CSIR-NCL, Pune

It is hereby certified that the work done by the student, under my/our supervision, is plagiarism-free in accordance with the UGC Regulations on “*Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)*” and the CSIR Guidelines for “*Ethics in Research and in Governance (2020)*”.



Signature of the Supervisor

Name : Dr. Ashok P. Giri

Date : 14/03/2022

Place : CSIR NCL, Pune

Contents

Acknowledgements.....	v
List of tables.....	vii
List of figures.....	viii
Abbreviations.....	xii
Synopsis.....	xiv
Thesis organization.....	xxi

Chapter 1

Introduction and review of literature

1.1 Plant phenylpropanoids	2
1.2 Unique catalytic features of plant 4CL enzymes.....	3
1.3 Evolution and diversity of plant 4CLs	5
1.4 Gene structure and genomic location of plant 4CLs.....	8
1.5 Contribution of 4CL in lignification, flavonoid biosynthesis and stress response.....	9
1.5.1 Lignification.....	9
1.5.2 Flavonoid biosynthesis.....	10
1.5.3 Protection from biotic and abiotic stresses.....	11
1.6 Activity and expression of 4CL is tightly regulated at various levels.....	12
1.6.1 Regulation at transcriptional level.....	12
1.6.2 Regulation by plant hormones.....	14
1.6.3 Regulation in response to biotic and abiotic stresses	16
1.7 Functional specificity of 4CL isoforms: a case study.....	17
1.8 Scope of 4CLs in metabolic pathway engineering for various products.....	18
1.8.1 Fuel and paper industry	18
1.8.2 Natural product industry.....	19
1.8.3 Food technology	20
1.9 Conclusions and future perspectives.....	20
1.10 Objectives of the work.....	21

Chapter 2

Diversity of 4-coumarate-CoA ligases in *Ocimum kilimandscharicum* and functional characterization of two candidate Ok4CLs

2.1 Introduction.....	24
2.2 Materials and methods.....	25
2.2.1 Plant material and chemicals.....	25
2.2.2 Identification of 4-CLs from <i>O. kilimandscharicum</i> transcriptome datasets.....	25
2.2.3. Phylogenetic, Sequence Similarity Network and domain analysis.....	27
2.2.4. Gene expression analysis.....	28
2.2.5. Cloning, recombinant protein expression and purification of Ok4CL7 and -15.....	29
2.2.6. Enzymatic assays of recombinant Ok4CL7 and -15.....	31
2.2.7. Characterization of recombinant Ok4CL products on HPLC and LC-MS/MS.....	32
2.2.9. Statistical analysis.....	32
2.3 Results and discussion.....	33
2.3.1. Ok4CLs diverged depending on their biological roles.....	33
2.3.2. Ok4CL7 and -15 exhibits differential expression in various tissues.....	38
2.3.3. Recombinant Ok4CL7 and -15 proteins differ in biochemical properties and substrate specificity.....	40
2.4. Conclusions.....	44

Chapter 3

Molecular dissection of substrate specificity mechanism in Ok4CLs

3.1 Introduction.....	46
3.2 Materials and methods.....	47
3.2.1 Comparison of 4CL protein sequences from <i>Ocimum kilimandscharicum</i> and reported 4CLs from other plants.....	47
3.2.2 Molecular docking analysis for determination of residues interacting with the substrate in the active site.....	47

3.2.3	Development of 4CL mutant models to alter substrate specificity.....	48
3.2.4	Cloning and expression and purification of mutant Ok4CL proteins.....	49
3.3.	Results and discussion.....	50
3.3.1	Adenylate forming domain and conserved residues present in Ok4CLs.....	50
3.3.2	Differential interactions of Ok4CL7 and -15 with substrates contributes to variation in substrate preferences.....	52
3.3.3	Mutant Ok4CL7 and -15 models shows variation in affinity towards sinapic acid substrate compared to wild.....	54
3.4	Summary.....	59

Chapter 4

In planta characterization of three 4CL isoforms from *Ocimum kilimandscharicum*

4.1	Introduction.....	61
4.2	Materials and methods.....	62
4.2.1.	Generation of <i>Ok4CL7</i> , <i>-11</i> and <i>-15</i> overexpression gene constructs.....	62
4.2.2	Development of <i>Ok4CL</i> overexpression lines of <i>N. benthamiana</i>	63
4.2.3	Analysis of root growth in <i>Ok4CL</i> overexpression lines and wild-type <i>N. benthamiana</i> plants.....	64
4.2.4	Metabolite profiling and transcriptome of <i>Ok4CL</i> overexpression lines and wild-type <i>N. benthamiana</i> plants.....	65
4.2.5	Root growth phenotype rescue by gene silencing and media supplements.....	66
4.2.6	Quantification of GUS activity from the <i>Ok4CL11</i> overexpression lines and wild-type <i>N. benthamiana</i> plants transformed with the <i>ProDR5-GUS-pBI101</i> construct.....	68
4.2.7	Grafting (homo- and hetero-grafts) to study its effect on root growth	69
4.3.	Results and discussion.....	69
4.3.1	Overexpression lines were confirmed by PCR amplification of <i>Ok4CLs</i> using genomic DNA and cDNA from putative transformants of <i>N. benthamiana</i>	69

4.3.2	Root growth was affected in <i>N. benthamiana</i> plants overexpressing <i>Ok4CL11</i> gene	70
4.3.3	Flavonoid and lignin content was altered in leaf and stem tissues of <i>Ok4CL11</i> overexpressing <i>N. benthamiana</i> lines.....	72
4.3.4	‘Rootless phenotype’ in <i>Ok4CL11</i> overexpressing <i>N. benthamiana</i> plants was rescued by transforming <i>Ok4CL11</i> RNAi construct.....	74
4.3.5	RNA-seq and metabolite analysis to identify the cause behind ‘reduced root growth’ phenotype in <i>Ok4CL11-OE</i> lines.....	75
4.3.6	Preliminary analysis suggest that <i>Ok4CL7</i> and <i>OK4CL15</i> localize to peroxisomes.....	77
4.3.7	List of other ongoing experiments to validate the function of <i>Ok4CL11</i> in producing rootless plants.....	78
4.4	Summary.....	80

Chapter 5

Summary and future prospects

5.1	Summary.....	83
5.2	Future prospects.....	85
5.3	Open questions.....	86
	Bibliography.....	87
	Abstract.....	102
	List of publications.....	103
	Published papers.....	104
	Curriculum vitae	

Acknowledgement

It gives me immense pleasure to look over my Ph.D. journey and remember all who have helped and supported me. This thesis is the outcome of several people's contributions. Without their support, this incredible journey, a blend of learning, patience, help, encouragement, aspiration, ecstasy and melancholy, would not have been possible. Hence, it becomes my innermost will to acknowledge all those who have extended their support to me scientifically, morally, or otherwise to make my way smooth. I am incredibly grateful to Dr. Ashok P. Giri, my research guide, for providing great support and having faith in me. This work would not have been possible without his guidance, critical inputs, and freedom to execute my ideas. He has always supported me in every sense throughout these years. I also appreciate his meticulous attention to the minute details. I appreciate all his effort, ideas and time to make this happen. He introduced me to Plant molecular bio

logy, specifically to the miraculous orchestra of plant specialized metabolites by sharing interesting papers, ideas, and all required to be a good researcher.

During my doctoral work, I am grateful to Dr. Rakesh Joshi, who has been like my co-guide (though not officially). I know him from IBB, Pune University. I remember his guidance and all the possible help from the start of my Ph.D. admission process to completing my thesis work. Specifically, he made the job of bioinformatics feasible for me, which was a completely new experience. He supported me not only in bioinformatics work but also in other works like manuscript writing, preparation for the presentation, group discussion, exploring new scientific ideas and group activities; this helped me to be a good researcher and colleague.

Another important person in the journey is Dr. Kirtikumar Kondhare. He introduced me to an entirely new field called plant tissue culture. He taught me the development of transgenic lines of model plants. I appreciate the constructive suggestions offered by him in all my Ph.D. work and also in my career. He is always there to help me and anyone who wants to discuss science. It's a fantastic experience to discuss any scientific concepts related to plant development and metabolic pathways with Ranjit and Kirtikumar at lunchtime, weekends dinner and all the time. For me, he is like a guide, colleague, and friend..... in short a fantastic personality.

I express my heartfelt gratitude to Dr. Vidya Gupta, who has been a constant source of inspiration during my stay at NCL. I also thank Dr. Narendra Kadoo, who has always been available for discussion and continuously supported my research work.

I am grateful to Dr. Atul Anand and Dr. Raviraj Kalunke for their guidance and help not just scientifically but as a friend also. During initial days of my Ph.D., both supported work from experiment design and data analysis to prepare for the presentation. I'm thankful to both of them for helping in molecular biology experiments.

I am also thankful to Dr. Vitthal Baravkar from Savitribai Phule Pune University, Pune for his help in analysis of metabolites from WT and *Ok4CLOE N. benthamina* lines. I am thankful to Dr. Yashwant Kumar from THSTI, Faridabad for the help in characterization of Ok4CL enzyme products. I am thankful to Dr. Amey Bhide, Prof. Anjan Banerjee and Tushar Dubey

for their help in confocal imaging of samples from localization experiment. I am incredibly thankful to my seniors, Dr. Pranjali Oak, Dr. Nidhi Saikhedkar, Dr. Sucheta Patil Dr. Rahul Tanpure, Dr. Neha Mahajan, Dr. Saleem Dar, Dr. Balkrishna Shinde, Dr. Amol Kasodekar, Dr. Tejas Chirmade and for their guidance in the handling of the instruments.

It is probably the best time to express my gratitude to my B.Sc. professors and faculty Dr. Anil Khanadagale, Dr. Birajdar, and Dr. Jamdade. My M.Sc. faculties, Dr. M. Kachole, Dr. V. Hivrale, Dr. M. Padul, Dr. V. Shende and Dr. M. Fawade, kindled my scientific interest and encouraged me to explore various dimensions.

I take this opportunity to sincerely acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi, for providing me with the Junior and Senior Research Fellowships, which facilitated me in carrying out my work. I would like to thank The Director, National Chemical Laboratory (NCL), Pune and HODs of Biochemical Sciences and AcSIR staff (Komal, Vijaya and Vaishali) at the NCL for the smooth processing of Ph.D. documents.

I would like to mention Gopal, Ranjit, Nidhi, Veena, Smriti, Bhakti, and Dr. Monika for their joyful company in all outside activities. I'm grateful for the generous help offered by all of them during my experiments. My heartfelt thanks to Ajit, Dr. Rashmi, Dr. Shivraj, Dr. Sancharini, Ravindra, Sonal, Sonali, Sagar, Amrita, Tanuja, Tejas V, Shounak, Ashwini, Sharada, Nivedita, Gayatri, Ravindra, Shamlal, Aruna, Vaishnavi, Meenakshi, Yogita, Bhagyashri, Vikram, Deepti, Preshita, Veenothini, Shweta, Nikita, Apurva, Akshay, Bela and all other lab mates for making my experience enjoyable. Thank you all for sparing liquid nitrogen, co-autoclaving, decontaminating and numerous merry tasks that we have done together.

I would like to mention my friends who helped directly or indirectly in my research and personal life Dr. Balaji, Dr. Shreya, Amol and Dr. Namdev. I'm grateful for their support throughout the journey. Your support and encouragement were worth more than I can express on paper.

Last but not least, I owe my parents (Aai-Shantabai Lavhale, Baba-Govind Lavhale and sisters- Anita and Savita) for their continuous encouragement and support. They have sacrificed a lot for my success. They have always been there for me through my ups and downs.

I apologise if I may have missed anyone's name unintentionally.

List of tables

Table 1.1 Number of annotated and characterized *4CL* genes in plants.

Table 1.2 List of reported factors/elements that regulate *4CL* in plant.

Table 2.1 Primers used for qRT-PCR and full-length cloning of *Ok4CL7* and *Ok4CL15*.

Table 2.2 Classified features of Ok4CL enzyme (clade and domain arrangement) along with selected other plant 4CLs.

Table 3.1 List of the substrates used for interaction study with Ok4CL mutants and their respective metabolic pathways.

Table 3.2 Predicted Stability Change ($\Delta\Delta G$) of Ok4CL7 variants. Replacement of amino acids at selected with all other 19 standard amino acids leads to change in folding energy. Replacements showing positive $\Delta\Delta G$ chosen for further study (Highlighted in red circle).

Table 3.3 Predicted stability change ($\Delta\Delta G$) of Ok4CL15 variants. Replacement of amino acids at selected with all other 19 standard amino acids leads to change in folding energy. Replacements showing positive $\Delta\Delta G$ chosen for further study (Highlighted in red circle).

Table 4.1 QC reports for RNA-sequencing samples. The total RNA was quantified on Qubit and RNA integrity was further checked using Bioanalyzer. OE1 and OE2 are overexpression lines. Rep= Biological replicate; RIN= RNA integrity number.

List of figures

- Figure 1.1** General phenylpropanoid pathway.
- Figure 1.2** Mechanism of reaction catalyzed by 4CL.
- Figure 1.3** Phylogenetic tree showing the relationship among different 4CL proteins.
- Figure 1.4** Gene structure of 4CLs from *Oryza sativa* ssp. japonica, *Arabidopsis thaliana*, and *Physcomitrella patens*.
- Figure 1.5** Structural features of promoter might be involved in regulation of 4CL gene expression.
- Figure 1.6** Co-expression analysis of *At4CL1*, *At4CL3* and *Gm4CL4*.
- Figure 2.1 A** Sequence Alignment of 4CLs from *O. kilimandscharicum* and a representative from *A. thaliana* and *G. max*. **B.** Flowchart of analysis performed to select putative 4CL sequences from *O. kilimandscharicum* transcriptome data.
- Figure 2.2** RNA isolation from different tissue of *O. kilimandscharicum* plants.
- Figure 2.3** Phylogenetic analysis of plant 4CLs. The phylogenetic tree was constructed using 946 4CLs from UniProt database and 15 4CLs from *O. kilimandscharicum*. Plants 4CLs are distributed in 9 clades (highlighted in different colors).
- Figure 2.4** Phylogenetic analysis of 4CLs. Class-I 4CLs are involved in the biosynthesis of lignin and Class-II 4CLs are involved in the biosynthesis of non-structural phenylpropanoids.
- Figure 2.5** The sequence similarity network analysis of various plant 4CLs, including that from *O. kilimandscharicum* showed five clusters at the E-value of $1e^{140}$.
- Figure 2.6** Domain analysis. OK4CL7, -15 and -8 from *O. kilimandscharicum* have several domains, including AMP BD- AMP binding domain; CaiC (Acyl CoA Synthetase); 4CL (4-Coumarate-CoA Ligase).
- Figure 2.7** *Ok4CL* gene expression analysis in different tissues of *O. kilimandscharicum*. **A.** Gene expression analysis of putative fifteen *Ok4CLs* in ten different tissues of *O. kilimandscharicum*, such as androecium, gynoecium, inflorescence, mature leaf, petal, root, sepal, stem, trichome, and young leaf. **B.** *Ok4CL7* and -15 expression analyses in all ten tissue types of *O. kilimandscharicum*.

Figure 2.8 Recombinantly expressed (A) Ok4CL7 and (B) Ok4CL15 showed the predominant band on 12% SDS PAGE. Ok4CL7 and -15 proteins were observed at expected size 90.4 kDa and 58.9 kDa, respectively.

Figure 2.9 Determination of optimal conditions for 4CL7 and -15 activities. B. Determination of optimum pH for Ok4CL7 and -15. C. Determination of optimal temperature for Ok4CL7 and -15.

Figure 2.10 The activity of 4CL7 and -15 with different substrates. A. Ok4CL7 showed activity with *p*-coumaric acid (*p*-coumaroyl CoA: 333nm), ferulic acid (feruloyl CoA: 345 nm), and caffeic acid (caffeoyl CoA: 345 nm). B. Ok4CL15 showed activity with *p*-coumaric acid, ferulic acid, and sinapic acid (sinapoyl CoA: 352nm). Analysis of enzyme assay reaction on HPLC. Three runs of each enzyme with each substrate. i) standard substrate; ii) heat-inactivated enzyme and all other components; iii) reaction mixture with the active enzyme; C. Enzyme activity with *p*-coumaric acid (COU), *p*-coumaroyl CoA product peak observed at RT 21.05; D. Enzyme activity with caffeic acid (CAF), caffeoyl CoA product peak observed at RT 19.02; E. Enzyme activity with ferulic acid (FER), feruloyl CoA product peak observed at RT 21.38; F. Enzyme activity with sinapic acid (SIN), sinapoyl CoA product peak observed at RT 20.7.

Figure 2.11 Analysis of 4CL enzyme assay products on LC-MS. A: Coumaroyl CoA; B: Feruloyl CoA; C: Caffeoyl CoA. Different peaks of *p*-coumaroyl CoA (912.5418 Da for at 1.73 retention time) and feruloyl CoA (942.1552 Da at 3.89 retention time) were observed.

Figure 3.1 Analysis of DNA sequence for the desired mutation in *OK4CL7* and -15 gene.

Figure 3.2 Alignment of Ok4CL7 and -15 amino acid sequences from *O. kilimandscharicum* reported 4CL from *A. thaliana*, *G. max* and *N. tabacum*.

Figure 3.3 Substrate preferences of Ok4CLs. Molecular interaction of substrate binding pocket of Ok4CL7 with substrate-specific functional groups.

Figure 3.4 Binding energy of Ok4CL7 variants with different substrates.

Figure 3.5 Binding energy of Ok4CL15 variants with different substrates.

Figure 4.1 *Ok4CLs* overexpression constructs in a binary vector pRI101AN. (A) Schematic representation of the overexpression constructs of *Ok4CL* genes; (B) Confirmation of *Ok4CL7*,

-11 and -15 gene constructs in pRI101AN vector by *EcoRI* and *NdeI/Sall* restriction enzyme digestions (Gel i, and ii).

Figure 4.2 Flow chart depicting the procedure for generation of transgenic *N. benthamiana* plants using *Agrobacterium*-mediated leaf transformation method (as described in the Methods section).

Figure 4.3 The total RNA separation on 1% agarose gel. Well numbers (1 to 6) represent RNA isolated from wild-type *N. benthamiana* plants, well numbers 7 to 12 indicate RNA from *OK4CL11-OE1 N. benthamiana* line, whereas well numbers 13 to 18 represent RNA for *OK4CL11OE2 N. benthamiana* line.

Figure 4.4 (A) Mechanism of siRNA mediated post-transcriptional gene silencing, and (B) Schematic representation of the *Ok4CL11-pRI101AN* RNAi construct.

Figure 4.5 Schematic representation of *DR5pro:GUS-pB1101*- construct. *GUS* gene (1.8 kb).

Figure 4.6 Screening of *Ok4CLs* overexpression lines using RT-PCR. RT-PCR was performed using cDNA isolated from leaves of (A) *Ok4CL11*, (B) *Ok4CL7*, and (C) *Ok4CL15* overexpression lines and WT *N. benthamiana* plants (negative control). Plasmid (35S:*Ok4CL11*-pRI101AN) construct was used as a template for PCR positive control. Binary construct plasmids (35S:*Ok4CL11*-pRI101AN, 35S: *Ok4CL7*-pRI101AN and 35S:*Ok4CL15*-pRI101AN) served as a template for PCR of respective *Ok4CL* gene (positive control). PCR master mix was used as a template for PCR negative control.

Figure 4.7 (A) Representative phenotypes of *N. benthamiana* plants overexpressing *Ok4CL7*, -11 (OE1 and OE2), and -15 compared to wild-type (WT) plants. Analysis of root phenotypes from WT and *Ok4CL*-OE *N. benthamiana* lines (*Ok4CL11*-OE1 and *Ok4CL11*-OE2). (B) Number of roots per plant, (C) number of lateral roots on the root; (D), and the overall root length (cm). WT is a non-transformed plant of *N. benthamiana*. N= number of biological replicates. At least fifteen plants per line or WT (i.e. n=15) were used for measurements of root growth parameters.

Figure 4.8 (A) Analysis of root apex growth. (B) root length, (C) lateral root number, and (D) lateral root length. N= number of biological replicates. For WT, n=13; for OE1 line -1, n= 9; and OE2 line, n=27) were used for measurements of root growth parameters.

Figure 4.9 (A) Relative abundance of kaempferol in *Ok4CL11*-OE lines (OE1 and OE2) and wild-type *N. benthamiana* plants. (B) Lignin accumulation in *Ok4CL11*-OE lines (OE1 and OE2) and wild-type (WT) *N. benthamiana* plants.

Figure 4.10 Grafting experiment using *Ok4CL11*-OE2 line and wild-type (WT) *N. benthamiana* plants. (A) Stock and scion from WT (WT homo-graft). (B) Stock from the OE2 line and scion from WT *N. benthamiana* plants (WT and OE2 hetero-graft). (C) Stock and scion from the OE2 line (OE2 homo-graft). (D) Stock from WT and scion from the OE2 line (OE2 and WT hetero-graft). Scale= 1 cm.

Figure 4.11 RNA-seq and gene expression analysis : (A) Table for DE genes and (B) Venn diagrams for DE genes (C) RT-qPCR for of phenylpropanoid pathway genes. *GT*, *Glycosyl transferase*; *4CL*, *4-Coumarate-CoA ligase*; *CHS*, *Chalcone synthase*; *CHI*, *Chalcone isomerase*; *F3H*, *Flavonoid 3'-monooxygenase*; *PAL*, *Phenylalanine ammonia-lyase*; *CAD*, *Cinnamyl alcohol dehydrogenase*; *CCR*, *Cinnamoyl-CoA reductase*; *C4H*, *Cinnamate 4-hydroxylase*; *HCT*, *Shikimate O-hydroxycinnamoyl transferase*

Figure 4.12 (A) Schematic representation of *mCherry*-pRI101AN and *mCherry:Ok4CL*-pRI101AN construct prepared for localization study; (B) Confirmation of *mCherry* and *Ok4CL7* gene in pRI101AN vector by *EcoRI* and *NdeI/Sall* restriction enzyme digestion.

Figure 4.13 *Ok4CL7* and -15 protein localization. Localization of *mCherry* to the nucleus and cell membrane (A-C); *Ok4CL7* (D-F); and *Ok4CL15* (G-I).

Figure 4.14 Phenotypes observed in ongoing experiments. (A) *Pr-DR5-GUS-pBI101* expressed in wild *N. benthamiana* and *Ok4CL11* overexpressing *N. benthamiana* (*Ok4CL11-OE1-pRI101AN*). (B) Root growth was observed in *Ok4CL11-OE2-pRI101AN* upon expression of *Ok4CL11-pRI101AN* RNAi construct. (C) Potato shoots regenerated from agro-infected leaf tissues represent putative transformants of *Ok4CL11-OE* in *Solanum tuberosum* cv. Désirée.

Abbreviations

4CL	4-coumarate:CoA ligase
AU	Absorbance Units
ACN	Acetonitrile
bp	Base pair
C3H	<i>p</i> -Coumarate 3-hydroxylase
C4H	Cinnamate 4-hydroxylase
CAD	Cinnamyl alcohol dehydrogenase
CAF	Caffeic acid
CHS	Chalcone synthase
CIN	Cinnamic acid
CoA	Coenzyme A
COMT	Caffeic acid O-methyltransferase
COU	<i>p</i> -Coumaric acid
CVOM	Chavicol O-methyltransferase
DFR	Dihydroflavonol 4-reductase
EOMT	Eugenol O-methyltransferase
F5H	Ferulate 5-hydroxylase
F5H	Ferulate 5-hydroxylase
h	Hour
HCT	<i>p</i> -Hhydroxycinnamoyl-CoA:quinic shikimate <i>p</i> -hydroxycinnamoyl transferase
HPLC	High performance liquid chromatography
Kb	Kilo base pair
kDa/kD	Kilo dalton
M	Molar
mg	Milligram
min	Minute
mL	Millilitre
mM	Millimolar
mmol	Millimoles
MS	Mass spectrometry

MW	Molecular weight
μL	Microliter
μM	Micromolar
nm	Nanometer
ng	Nanogram
PCR	Polymerase chain reaction
PDB	Protein data bank
PAL	Phenylalanine ammonia lyase
SDS PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SIN	Sinapic acid
RT-PCR	Reverse transcriptase-polymerase chain reaction

Synopsis

Introduction

The *Ocimum* genus belongs to the Lamiaceae family. According to World Flora Online, this genus comprises 66 species. In this study, we made an attempt to characterize 4CLs from *O. kilimandscharicum*. Functional characterization of 4CL includes gene expression analysis, *in vitro* and *in vivo* characterization using *Escherichia coli* and *Nicotiana benthamiana* model systems, respectively.

Statement of problem

Ocimum is known for several important medicinal properties, including antimicrobial activity, antiinflammation activity, antistress activity, antidiabetic, antioxidant and wound healing, etc. Exact compounds or groups of compounds associated with these activities are not known in *Ocimum* species. Metabolic analysis revealed that *Ocimum* species are rich in either phenylpropanoids (*O. gratissimum*) or terpenoid (*O. kilimandscharicum*) or they may have combination of both compounds in substantial amounts (*O. canum*). In most of the *Ocimum* species of 4CL enzymes are not characterized besides having its important role in the phenylpropanoid biosynthesis pathway (Gurav et al., 2021; Singh et al., 2015). Functional characterization of 4CL isoforms in *Ocimum* will help to understand the role of individual isoforms in phenylpropanoid diversity and plant physiology. Knowledge about the contribution of Ok4CLs in modulation of flux towards either lignin or non-structural phenylpropanoids will be useful to improve plant physiology or essential metabolite content in this important plant species. Further, this knowledge could be used for the development or screening of a variety with high essential oil production.

Objectives

On the basis of available information, the major objective of this study is the characterization of 4-Coumarate CoA-Ligase isoforms from *Ocimum kilimandscharicum*. The work of this thesis is divided into divided into three parts.

1. Functional characterization of recombinant Ok4CL isoforms.
2. Determination of active site residues of Ok4CLs responsible for substrate specificity.
3. Functional characterization of Ok4CLs in *Nicotiana benthamiana*.

Methodology**A) *In vitro* functional characterization of Ok4CL isoforms****i) Transcriptome analysis**

We have transcriptome data from four tissues of *O. kilimandscharicum*, including trichome, root, flower and leaf. All 4CL annotated sequences were selected for further analysis. A number of 4CL sequences annotated from each tissue, includes 25 from a flower, 24 from a trichome, 11 from a root and 18 from a leaf. All sequences were assembled using CAP3 Sequence Assembly Program (Huang, X. and Madan, A. 1999). Based on assembly results, we removed the overlapping portion of sequence and assembled the remaining part to produce longer sequences. With this, we got 48 sequences of putative 4CLs. Then ORF analysis was performed using the ORF finder web tool from the NCBI. Based on this, we selected all those sequences having start to end codon sequence and the presence of two conserved signature motifs in its sequence. Two signature motifs present in 4CL include box-I and box-II. All those enzymes having Box-I (SSGTTGLPKG ν) are grouped into a superfamily of adenylate-forming enzymes. Box-II (GEICIRG) is involved in catalysis. Sixteen such sequences were found having these two conserved motifs and we have selected these for further characterization.

ii) Gene expression analysis

Expression of *Ok4CL* isoforms analysed in young leaf, mature leaf, stem, root, trichome, inflorescence, androecium, gynoecium, sepal and petal from *O. kilimandscharicum*. 2 μ g RNA was used for cDNA synthesis using SuperScript III reverse transcriptase system (Invitrogen, Carlsbad, USA). cDNA synthesis was confirmed by PCR using a set of primers that amplifies a short sequence elongation factor 1 α (EF1 α). qRT-PCR primers were designed for all putative 15 *Ok4CL* genes using Primer 3.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and Oligo Analysis online tool (<http://www.operon.com/tools/oligo-analysis-tool.aspx>). qRT-PCR primers were synthesized from Eurofin India (Banglore, India). qRT-PCR was performed on an Applied Biosystems 7500 Fast Real-Time PCR System by using the SYBR green protocol. Three biological and technical replicates were used for each sample. The reaction mixture contains 5 μ l of SYBR green master-mix, 0.5 μ l of 10 μ M forward and reverse gene-specific primers and 1 μ l of diluted cDNA (1:2) with nuclease-free water added to make up a volume to 10 μ l. Standard plots of putative 4CLs were generated using a gene-specific set of primer-pairs. Different cDNA dilutions (1:2, 1:3, 1:8, 1:16 and 1:32) were used for generation of standard plots.

iii) Phylogenetic analysis

Phylogenetic tree constructed with amino acid sequences of functionally characterized 4CLs from other plants and putative Ok4CLs of *O. kilimandscharicum*. Deduced amino acid sequence generated from putative Ok4CL genes of *O. kilimandscharicum* using online ExPASy Translate tool. Amino acid sequences of functionally characterized 4CLs from other plants downloaded from the NCBI database. Selected sequences were aligned using the Clustal W tool. All parameters kept as a default for alignment. A phylogenetic tree was constructed using the MEGA 6.0 tool. The type of tree is neighbor joining and the model used for this is the Poisson model. Thousand bootstrap replicates are used while other parameters kept are as default settings. Scale shown at the bottom of the tree indicates that the length of the branch represents 0.02 amino acid changes.

iv) Gene cloning, expression and functional characterization of Ok4CLs

pGEX-4T and pET28a expression vectors used for expression of Ok4CL7 and -15, respectively in the bacterial system. *Ok4CL7* (*NdeI* and *EcoRI*) and *Ok4CL15* genes were cloned using restriction enzyme-based cloning. Various constructs were confirmed using restriction enzyme digestion and sequencing.

Vectors having *4CL* gene(s) were transformed into BL21 chemically competent cells and standard protocol was used for protein expression. Expression was induced by IPTG (500 μ l of 1M for 1 litre or 0.5 μ M). Ok4CL7 and -15 proteins were purified using GST and Ni-NTA slurry, respectively. Purified protein separated on 12% SDS-PAGE gel. The reaction mixture of 200 μ l contains substrate (0.2 mM), adenosine triphosphate (2.5 mM), coenzyme A (0.2 mM) and enzyme (5 μ g). All assays were performed in Tris-HCl buffer (200 mM; pH 8.0) having MgCl₂ (25 mM). Blanks used for all assays contain heat-inactivated enzymes and all other components. The change in absorbance was monitored with a UV-visible spectrophotometer (Labindia Pvt Ltd., India). The optimum pH and temperature required for activity of recombinant Ok4CL7 and -15 was determined using ferulic acid and sinapic acid, respectively. The range of pH 5 to pH 9 used for determining optimum pH, while the optimum temperature range was from 20°C to 80°C. The absorbance was measured at absorption maxima at 311, 333, 345, 346, and 352 nm for the corresponding cinnamoyl-CoA, 4-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, and sinapoyl-CoA products, respectively. The optimum pH required for the activity of Ok4CL7 and -15 was determined using range assay buffers from pH 4 to 9. Temperatures ranging from 20 to 80°C were used to determine the optimum enzyme activity.

B) Determination of active site residues**i) Molecular docking analysis for determination of residues interacting with substrate in active site**

Interaction and substrate preferences of recombinant Ok4CL7 and -15 with their substrate was studied using molecular docking. Three-dimensional structures of Ok4CL7 and -15 were predicted using *Populus tomentosa* 4CL (PDB ID: 3NI2) and *Nicotiana tabacum* 4CL (PDB ID: 5BST) crystal structures as templates, respectively. Structures were energy minimized using Maestro 10.1 Tools and substrate binding pocket residues were predicted by superimposing models with templates. Ok4CL structures were then prepared for docking by adding Kollman and Gasteiger charges in AutoDock Tools. Docking of substrates with Ok4CLs were performed using AutoDock Vina. After docking simulations, 10 docking poses were generated with each substrate. Binding pose with the lowest binding score were selected for further analysis. Substrate binding poses were analyzed using the BIOVIA Discovery Studio 4.5 software.

ii) Determination of specificity determining residues in Ok4CL isoforms

Based on molecular docking analysis, amino acids interacting with a substrate in the active site of Ok4CLs were selected for the development of mutants. Based on docking study analysis and previous reports, 27 amino acids of Ok4CL7 were selected for further study- S226, V276, Y277, S280, L281, H317, S345, C346, G347, A348, A349, P350, G371, Y372, G373, M374, E376, A379, L380, A381, V396, D456, V468, R471, K473, K477 and Q482. Similarly, 29 amino acids interacting with a substrate in the active site of Ok4CL15 were selected for further study- S189, H237, V238, Y239, V242, S243, P279, V306, C307, G308, A309, A310, P311, Q331, G332, Y333, G334, M335, T336, L339, V340, L341, M343, C359, T417, D419, I431, R434 and K525.

Using DUET tool models developed using replacement amino acid residue at all selected positions with all other standard amino acids. Based on predicted stability change ($\Delta\Delta G$), models were selected for molecular docking analysis. Molecular docking analysis of mutant models performed using AutoDock Vina with all possible substrates related to plant metabolism. Variants showing significant variation in binding energy with a substrate selected for the experimental validation. Mutant Ok4CLs were generated using the QuikChange lightning site-directed mutagenesis kit. Standard protocol provided by the manufacturer used for the mutant development. Mutation at the desired position was confirmed by DNA sequencing. Expression of the mutant proteins and functional characterization work is in progress.

C) *In vivo* characterization of Ok4CL isoforms

i) Cloning and development of transgenic *Nicotiana benthamiana* lines

Based on *in vitro* enzyme activity and substrate preferences, recombinant Ok4CLs are selected for *in vivo* characterization. Ok4CL7, -11 and -15 were further sub-cloned in the pRI101AN vector using restriction enzyme-based cloning. Clones were confirmed by restriction enzyme digestion and sequencing. Constructs were transformed into *Agrobacterium tumefaciens* GV2260 strain using heat-shock method. Positive colonies were inoculated and used for Agroinfection by co-cultivation method. Explant kept in liquid Murashige and Skoog (MS) medium having *A. tumefaciens* for 30 min on a shaker at 28°C. Then transferred to co-cultivation medium having MS medium and 2% sucrose. After 48 hours explants were transferred to regeneration medium having cefotaxime (500mg/L), kanamycin (50mg/L), benzylaminopurine (1mg/L) and naphthaleneacetic acid (0.1mg/L). Explants were sub-cultured after every two weeks on a fresh regeneration medium. Callus growth was induced from the explant and after 3 to 5 weeks, shoots were emerged on from the explant callus. Shoots were transferred to the root induction medium having cefotaxime (500mg/L), kanamycin (50mg/L) and indole-3-butyric acid (0.1mg/L). Positive lines were confirmed using PCR both from genomic DNA and cDNA (generated from mRNA). Positive lines were transferred to hardening and tissue from these lines will be used for gene expression and metabolite analysis.

ii) Subcellular localization of Ok4CLs isoforms

mCherry (red fluorescence protein) fused with Ok4CL isoforms were cloned in the pRI101AN vector for fusion protein expression in plants. Constructs were confirmed by restriction enzyme digestion and sequencing. Constructs transformed into *A. tumefaciens* GV2260 strain. Positive colonies were inoculated in the LB medium and used for the agroinfiltration experiments. Two month old *N. benthamiana* and *O. kilimandscharicum* used for the transient gene expression using agro-infiltration method. Tissue was harvested after 72 hours of post infiltration and observed under a confocal microscope and obtained data was analyzed.

iii) Qualitative analysis of lignin using phloroglucinol staining

Stem from the wild and the transgenic lines of *N. benthamiana* used for lignin analysis. Phloroglucinol specifically stains lignin and forms a bright pink-red colour. Handmade stem section stained using phloroglucinol stain (3% in ethanol) and observed under brightfield microscope.

Conclusions

Transcriptome data from four tissues of *O. kilimandscharicum* (including trichome, root, flower and leaf) were analysed. In total, 15 ORFs were selected on the basis of full length sequences and conserved motifs. The phylogenetic analysis with the reported 4CL isoforms clustered into two clades namely, class I and class II. Class I is involved in the biosynthesis of lignin and Class II is involved in the biosynthesis of phenylpropanoids other than lignin. Selected 4CLs were grouped into three major clades belonging to known three classes (CLs involved in lignin biosynthesis, non-structural phenolics biosynthesis and uncharacterized) as mentioned earlier. Ok4CL8, -9 and -16 grouped with other 4CLs in class-I, while all other Ok4CLs appeared in the uncharacterized clade (Class III) along with other 4CLs. Ok4CL3, -4 and -5 form a separate clade due to the presence of the fatty acyl CoA synthetase domain. Interestingly, Ok4CL15 is not clustered with any of the class. Comparatively, it is close to class I and class II than unclassified 4CLs. Comparative analysis of gene expression values indicates differential and tissue-specific expression of 4CLs in *O. kilimandscharicum*. Most of the 4CLs were expressed in the trichome, while the least expression was observed in floral organs. *Ok4CL11* was highly expressed in young and mature leaves, trichomes, androecium and petal. Whereas, *Ok4CL8* had high expression in stem tissue. *Ok4CL5* and -15 were abundant in inflorescence and root tissues. *Ok4CL7* is highly expressed in trichome, young leaf, mature leaf and root, while *Ok4CL15* shows expression in most of the *O. kilimandscharicum* tissues, with the highest levels in the root.

Recombinant Ok4CL7 and -15 were purified and separated on 12% SDS-PAGE. The activity of recombinant Ok4CL7 and -15 proteins were studied with cinnamic acid, p-coumaric acid, caffeic acid, ferulic acid and sinapic acid. Product formation was analysed by measuring the absorption of reaction mix at corresponding λ max of CoA product: p-coumaroyl CoA: 333nm), ferulic acid (Feruloyl CoA 345 nm), and caffeic acid (Caffeoyl CoA 345 nm). Out of the five substrates, Ok4CL7 utilized p-coumaric acid, caffeic acid and ferulic acid. Ok4CL15 showed activity with p-coumaric acid, ferulic acid and sinapic acid. Both recombinant 4CLs proteins utilized p-coumaric acid and ferulic acid and inactive against cinnamic acid.

Interaction analysis of Ok4CL7 and -15 docked complexes illustrated the functional groups of substrates and their intermediates form several non-covalent interactions with binding site residues. On the basis of docking results mutant models developed using DUET server. Mutants of Ok4CL7 and -15 expressed in bacteria and purified. Currently, efforts are underway to test the activity of 4CL mutants with different substrates to study the effect of mutation on enzyme specificity.

Coding sequences of *Ok4CL7*, *-11* and *-15* were cloned in pRI101AN binary vector for *in planta* expression. *N. benthamiana* transgenic lines were generated using the cocultivation method. *Ok4CL11*-Nb overexpression (OE) shows rootless phenotype. Preliminary analysis using phloroglucinol staining shows lignin content is affected in the stem of *Ok4CL11*-OE-Nb lines. Transgenic lines were transferred to glass house for hardening. Gene expression and variation in metabolite content will be studied using qRT-PCR and LCMS, respectively.

Localization of Ok4CLs is studied using the fusion constructs of Ok4CLs with mCherry. Fusion constructs were transformed in *Agrobacterium tumefaciens* (GV2260) and infiltrated in the leaves of 1 month old *N. benthamiana* plants. Confocal imaging showed that both *Ok4CL7* and *-15* localizes to peroxisomes. The presence of peroxisome targeting signal 1 (PTS1) in their protein sequences further confirms the observed localization findings. Further, experiment are required to validate this data using organelle specific markers or dyes.

Organization of thesis

The thesis is organized into five chapters; the contents of which are listed below.

Chapter 1: Introduction and review of literature

This chapter gives detailed description of the topics underlying the genesis of thesis. Literature about the importance of plant 4-coumarate-CoA ligase in plant specialized metabolite biosynthesis is covered in this chapter.

Chapter 2: Diversity of 4-coumarate-CoA ligases in *Ocimum kilimandscharicum* and functional characterization of two candidate Ok4CLs

Gene expression of fifteen *OK4CL* isoforms is studied in different tissues of *O. kilimandscharicum*. Moreover, two isoforms of Ok4CL (Ok4CL7 and -15) are functionally characterized. Optimum parameters required for enzyme activity are also analyzed in this chapter.

Chapter 3: Molecular dissection of substrate specificity mechanism in Ok4CLs

This chapter deals with the *in-silico* analysis of enzyme substrate interaction. Amino acid residues interacting with substrate in active site are predicted by molecular docking analysis. Mutant models are also developed to identify specificity determining residues.

Chapter 4: *In planta* characterization of three 4CL isoforms from *Ocimum kilimandscharicum*

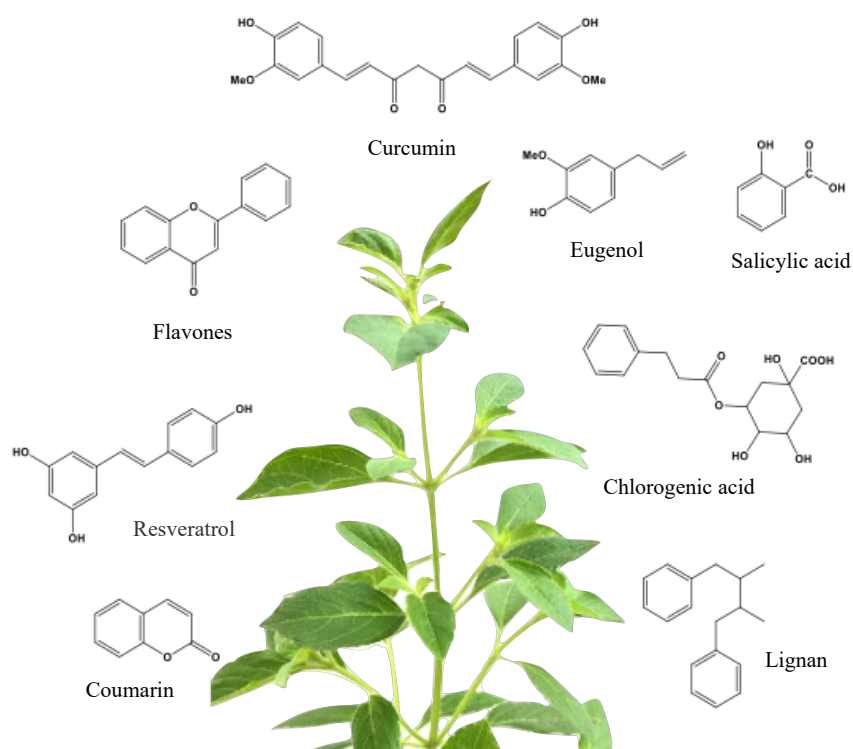
Three *Ok4CL* isoforms (*Ok4CL7*, -11 and -15) were overexpressed in *Nicotiana benthamiana* to study their roles in plant growth and development. Overexpression of *Ok4CL11* in *N. benthamiana* affected root growth. Different approaches, such as metabolite profiling, transcriptomics, gene expression analysis, RNAi, grafting, chemical applications, etc. are performed to find the reason behind root growth affected phenotype.

Chapter 5: Summary and future prospects

This chapter gives the overall summary of the thesis and future directions of this work.

Chapter 1

Introduction and review of literature



Contents of Chapter 2 have been published in the review article listed below.

Lavhale, S.G., Kalunke, R.M. & Giri, A.P., Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta* (2018), 248: 1063-1078 (*Cited 46 times*)

1.1 Plant phenylpropanoids

Plants have diverse range of metabolites that are obligatory for normal growth, development, and reproduction. Functions played by metabolites have very crucial role in plant life cycle which includes mechanical support, attraction of pollinators and frugivores, protection from biotic and abiotic stresses, interaction with environment, allelopathy effect, etc. (Bennett and Wallsgrove 1994; Ferrieri et al. 2015; Goufo et al. 2017). Plant secondary metabolites are categorized based on chemical structure, composition, solubility and pathway in which they are synthesized. For example, on the basis of chemical structure three groups (i) terpenoids, (ii) phenylpropanoids and (iii) alkaloids are known. The recent advances in metabolite analysis technologies empower the study of metabolites on large-scale derivatives of individual metabolite and metabolic pathways at cellular and organelle level (Fukushima et al. 2009; Kueger et al. 2012; Freund and Hegeman 2017; Floros et al. 2017). Number of secondary metabolites have been characterized (Supplementary material, Table S1), containing numerous valuable properties/applications such as colour, flavour, medicine, etc. (Korkina 2007; Tatsis and O'Connor 2016; Citti et al. 2017).

Phenylpropanoids are synthesized from aromatic amino acids such as phenylalanine (Phe) and tyrosine (Tyr) via phenylpropanoid pathway (Herrmann and Weaver 1999; Tzin and Galili 2010; Anand et al. 2016). Along with these aromatic precursor amino acids, tryptophan is also synthesized in plastid via shikimic acid pathway in plants (Schmid and Amrhein 1995; Herrmann and Weaver 1999; Tzin and Galili 2010). The shikimic acid pathway utilizes phosphoenolpyruvate from glycolysis and erythrose 4-phosphate from hexose monophosphate pathway to generate aromatic amino acids (Tohge et al. 2013). Further, general phenylpropanoid pathway begins with oxidative deamination of Phe and Tyr by phenylalanine ammonia lyase (PAL) and by tyrosine ammonia lyase (TAL) which leads to formation of cinnamic acid and *p*-coumaric acid, respectively. In subsequent steps, cinnamic acid is converted to various methoxy and hydroxy derivatives such as *p*-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid and sinapic acid using hydroxylases and methyltransferases enzymes (Fraser and Chapple 2011) (**Figure 1.1**). These methoxy and hydroxy derivatives utilized to corresponding CoA esters by 4CL. Products of 4CL are subsequently used by various oxygenase, reductases and transferases for biosynthesis of lignin, flavonoids, anthocyanins, aurones, stilbenes, coumarins, suberin, cutin, sporopollenin, etc. (Vogt 2010). Thus, 4CL is one of the key branch point enzyme in phenylpropanoid pathway. The gene family encoding these enzymes constitutes multiple isoforms of 4CL, which potentially perform diverse functions, which are yet to be fully characterized (Costa et al. 2005; Gui et al. 2011) .

However, literature reviews are available on phenylpropanoid pathway that comprehends little information on 4CL (Vogt 2010; Fraser and Chapple 2011).

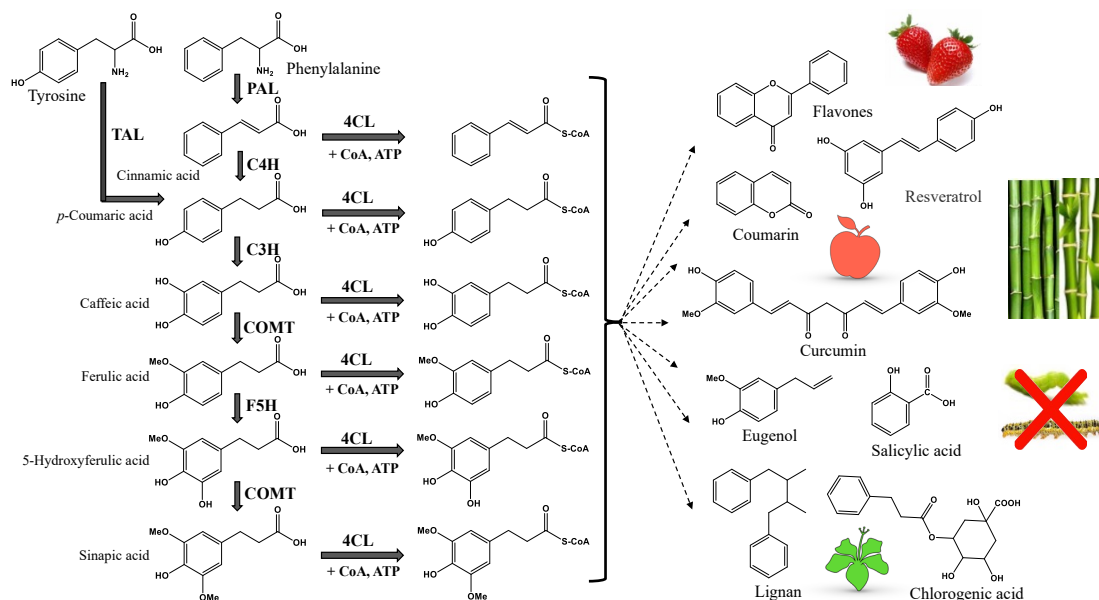


Figure 1.1 General phenylpropanoid pathway.

1.2 Unique catalytic features of plant 4CL enzymes

The 4CL catalyzes the ligation of Coenzyme-A (CoA) with cinnamic acid and its methoxy/hydroxy derivatives, like caffeic acid, ferulic acid etc. (Figure 1.1). This enzyme belongs to the family of adenylate-forming enzymes, which has two conserved peptide motifs: box I (SSGTTGLPKGTV) and box II (GEICIRG) (Uhlmann and Ebel 1993; Allina et al. 1998; Hu et al. 1998; Ehlting et al. 1999; Schmelz and Naismith 2009). Box I is the adenosine monophosphate (AMP) binding domain and is conserved in all the members of adenylate-forming enzyme family. This AMP binding domain widely exists in proteins of all living organisms and are involved in wide range of functions with fruit fly luciferase, gramicidin S synthetase, all type of CoA ligases, etc (Fulda et al. 1994). While the function of Box II is unclear and not directly involved in catalysis (Stuible et al. 2000). Further, the substrate-binding domain (SBD) is variable in different isoforms of 4CL. In *Arabidopsis thaliana*, domain responsible for the substrate specificity determination was studied using domain-swapping approach for two isoforms, At4CL1 and At4CL2. Both enzymes utilize 4-coumarate as a substrate but only At4CL1 is capable to utilize ferulate. Two adjacent domains known as SBD I and SBD II were identified. Out of these two, either one SBD of At4CL1 is sufficient to recognize ferulate (Ehlting et al. 2001). At4CL2 contain 12 amino acid residues in substrate

binding pocket (SBP), which were determined using crystal structure with the help of gramicidin S synthase homology model. Amino acids in SBP include Ile-252, Tyr-253, Asn-256, Met-293, Lys-320, Gly-322, Ala-323, Gly-346, Gly-348, Pro-354, Val-355, and Leu-356. These 12 amino acids are flanked by conserved box I and box II motifs (Schneider et al, 2003). The members of acyl-activating enzyme superfamily share a little sequence identity, but all they have conserved box I. In *Arabidopsis*, this superfamily has 77 4CL genes and their phylogenetic analysis shows that they formed seven groups and all have unique conserved box I (Shockey et al. 2003).

The probable function of any enzyme can be deciphered by analysis of its primary sequence and three-dimensional structure. Various features of 4CL structure were examined with the help of site-directed mutagenesis, domain exchange, mathematical modeling, crystallography, etc. Crystal structure of Pt4CL1 from *Populus tomentosa* was studied by employing anomalous dispersion together with molecular replacement method using luciferase from firefly as a model. Pt4CL1 contains 536 amino acids and has two globular domains, one at N-terminal and other at C-terminal having 434 and 102 amino acids, respectively. N-terminal domain has three sub-domains: N1, N2 and N3. N1 and N2 are similar in structure and have 6 parallel along with 2 antiparallel β -sheets in center. The eight central β -sheets are flanked by 4 and 2 α -helices at each end. Three residues are crucial for Pt4CL1 catalytic activity (Lys-438, Gln-443, and Lys-523) while five residues for substrate binding (Tyr-236, Gly-306, Gly-331, Pro-337, and Val- 338) (Hu et al. 2010). Based on crystal structure and reported sinapic acid converting 4CLs from other plants, At4CL2 has been modified using domain exchange and site-directed mutagenesis techniques to achieve an improved sinapic acid conversion rate. This has been achieved in two steps 1) replacing substrate- binding pocket region of At4CL2 from Box I to Box II with substrate- binding pocket from *At3g21230* gene encoding 4CL like protein and 2) deletion of either V355 or L356 but not both. They replace three amino acid residues (N256A, M293P and K320L) in At4CL2; this modified 4CL shows 30 folds improvement in conversion rate of sinapic acid (Schneider et al. 2003).

Some 4CLs have inter-protein interactions for its regulatory function. For example, Ptr4CL3 and Ptr4CL5 isoforms from *Populus trichocarpa* interact with each other and form tetramer having three Ptr4CL3 and single Ptr4CL5. The mathematical model was developed to study their kinetics at different ratios and activation components. Using this model and experimental analysis like microdissection, co-immunoprecipitation, chemical cross-linking, bimolecular fluorescence complementation and mass spectrometry, showed that Ptr4CL5 has a regulatory role in Ptr4CL3-Ptr4CL5 tetrameric complex (Chen et al. 2014a). Mechanism of

4CL activity also studied in Nt4CL2 isoform from *Nicotiana tabacum*. The crystal structure of Nt4CL2 reveals that this enzyme present in two conformations during catalysis, namely adenylate-forming and thioester forming. During the catalysis process, enzyme forms two conformations in which substrate is converted to adenylate intermediate and then to thioester form (**Figure 1.2**) (Li and Nair 2015).

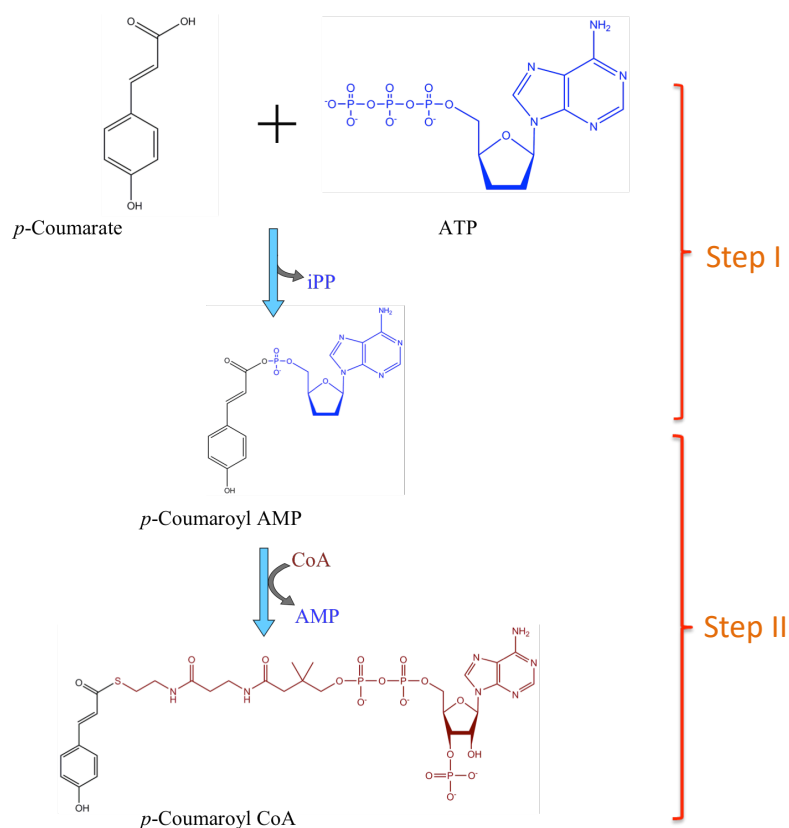


Figure 1.2 Mechanism of reaction catalyzed by 4CL.

1.3 Evolution and diversity of plant 4CLs

The evolution of phenylpropanoid pathway was crucial precondition for colonization of terrestrial plants on land. The products of this pathway such as flavonoids and lignin protect plants from UV light and provide structural support, respectively (Douglas 1996). Phylogenetic analysis of 4CL genes showed that the 4CL genes separated into two distinct clades from monocot and dicot plants. This is probably because 4CL genes evolve independently after the separation of monocots and dicots during the course of evolution. In dicot plants, 4CL genes grouped into two clusters: type I and type II (**Figure 1.3, Table 1.1**). Type I cluster is mainly involved in monolignol biosynthesis whereas type II is involved in phenylpropanoid

biosynthesis other than lignin. In *Arabidopsis thaliana*, At4CL1, At4CL2 and At4CL4 belong to type I and At4CL3 comes under type II cluster (Sun et al. 2013). Sun et al. (2013) categorized monocot 4CL's into two other clusters type III and IV with their similar functions as that of type I and II, respectively. In rice, Os4CL1, Os4CL3, Os4CL4 and Os4CL5 belong to type III, whereas Os4CL2 belongs to type IV. These different clusters formation suggest that the evolution of 4CL genes in dicot and monocot is an independent event after its separation. However, Pinta4CL3 isoform from *Pinus taeda* (a gymnosperm plant), is phylogenetically closer to type II angiosperm 4CLs than to type III Pinta4CL1 (Chen et al. 2014b). This suggests that type I and II might have diverged before the divergence of gymnosperm and angiosperm lineages (Cukovic et al. 2001). Amino acid sequences of At4CL1 and At4CL2 from *Arabidopsis thaliana* are more similar to each other (86% identical) as compared to At4CL3 sequence, which shows 71 and 73% identity with At4CL1 and At4CL2, respectively. These findings suggest that At4CL1 and At4CL2 are the recently evolved (Ehltng et al. 1999). Similarly, *Salvia miltiorrhiza* genome have ten *Sm4CL* related genes, out of these, only three (*Sm4CL1*, *Sm4CL2* and *Sm4CL3*) were clustered with bonafide 4CLs in phylogenetic analysis. *Sm4CL1* and *Sm4CL2* clustered in type I, which are involved in lignin biosynthesis whereas *Sm4CL3* clustered in type II 4CLs, which are involved in flavonoid biosynthesis (Wang et al. 2015).

The evolution of 4CL genes was studied in the 11 genera of *Larix* (family: Pinaceae) by comparing its copy number, GC content and codon usage, sequence divergence, and phylogenetic analysis (Wei and Wang 2004). The 4CL sequences were grouped into two paralogous clades, *4clA* and *4clB*. Both the clades have sequences from all studied 11 species of *Larix* genera, but some species have more than one sequence in the same clade. These findings support that two sister clades *4clA* and *4clB* is due to duplication in *Larix*'s common ancestor, thereby leading to co-existence of two alleles in all 11 species. After this duplication, they might have evolved separately because there is variation in values of mean distance of synonymous, non-synonymous, and nucleotide substitutions, and ratio of transitions to transversions. Thus, divergence of genera duplication and deletion in 4CL gene family has happened, which might be potential reason for variation in sequence and number of alleles. Therefore, two or three major 4CL loci might be present in Pinaceae family (Wei and Wang 2004).

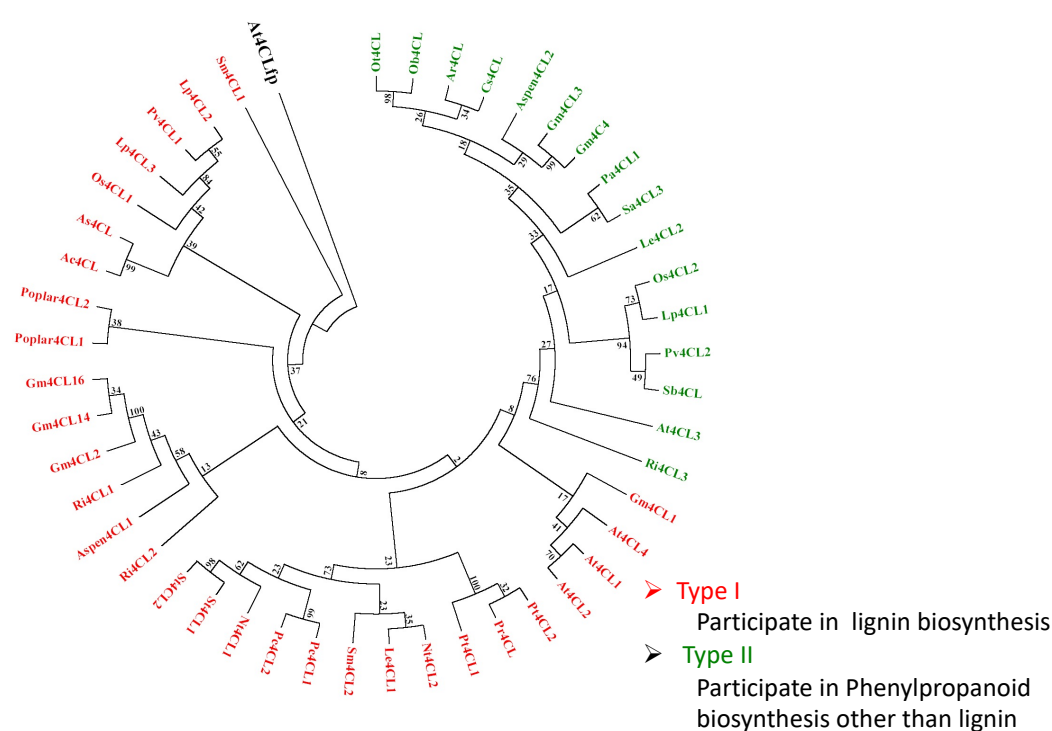


Figure 1.3 Phylogenetic tree showing the relationship among different 4CL proteins.

Table 1.1 Number of annotated and characterized 4CL genes in plants.

Plant	Annotated 4CLs	Characterized 4CLs	Class-I	Class-II	Reference
<i>Arabidopsis thaliana</i>	14	4	3	1	The Arabidopsis Genome Initiative, 2000; Costa et al, 2005
<i>Salvia miltiorrhiza</i>	10	3	2	1	Wang et al, 2015
<i>Populus trichocarpa</i>	20	6	5	1	Zhang et al, 2015
<i>Populus pruinosa</i>	20	5	4	1	Zhang et al, 2015
<i>Populus euphratica</i>	20	5	4	1	Zhang et al, 2015
<i>Salix suchowensis</i>	12	4	3	1	Zhang et al, 2015

List of annotated and characterized 4CL genes in various plant species such as *Arabidopsis thaliana*, *Salvia miltiorrhiza*, *Populus trichocarpa*, *Populus pruinosa*, *Populus euphratica* and *Salix suchowensis* (Table 1.1). In the above-mentioned plant species, only 3 to 6 4CLs are functionally characterized out of 10 to 20 annotated 4CLs.

1.4 Gene structure and genomic location of plant 4CLs

The 4CL genes have been studied from numerous plants and it is observed that they exist in small gene families, where they encode identical, nearly identical, or divergent proteins. In *Arabidopsis thaliana*, fourteen genes annotated as a putative 4CLs by *in-silico* genome analysis, out of which only eleven were studied for functional characterization. While, rest three has peroxisome targeting sequence (based on C-terminal signal peptide analysis), which is absent in bonafide 4CLs (Costa et al. 2005). Genomic location of *Arabidopsis thaliana* 4CLs showed a wide distribution on chromosome. Out of eleven functionally characterized 4CLs, *At4CL1*, *At4CL3*, *At4CL9* and *At4CL10* present on chromosome 1; *At4CL2*, *At4CL5* and *At4CL8* is on chromosome 3; *At4CL6* and *At4CL7* is on chromosome 4; and *At4CL4* and *At4CL11* is on chromosome 5. However, only *At4CL2* and *At4CL5* are clustered together on chromosome 3. Among these, *At4CL1*, *At4CL2*, *At4CL3* and *At4CL5* are functionally active (Costa et al. 2005). Sequence analysis of *At4CLs* shows that *At4CL1*, *At4CL2* have three introns, whereas *At4CL3* have six introns (**Figure 1.4**). Out of these three extra introns of *At4CL3*, first intron resulted from interruption of the first exon and remaining two introns resulted from interruption in the second exon of the *At4C11* and *At4C12* (Ehlting et al. 1999). There is variation in the sequence of 5' untranslated region and putative promoter region of *At4CL1*, *At4CL2*, *At4CL3*. Putative TATA boxes are located at -124, -103 and -116bp upstream of translation start codon ATG of *At4CL1*, *At4CL2* and *At4CL3*, respectively (Ehlting et al, 1999). For genes involved in phenylpropanoid pathway, like *PAL*, *C4H* and *4CL*, the promoter region have conserved box P, box A and box L. Promoter region of *At4CL1* and *At4CL2* contain only boxes P and L whereas *At4CL3* have all three boxes (Ehlting et al, 1999).

Consequently, in rice five 4CLs are reported namely *OsCL1*, *OsCL2*, *OsCL3*, *OsCL4*, and *OsCL5*. Sequence analysis of these 4CLs shows that *Os4CL1* and *Os4CL5* contain five exons and four introns. The position of introns is conserved; however, there are differences in intron length and sequence. *Os4CL2* have one additional intron, which results from the interruption of third exon of *Os4CL1* and *Os4CL5* (Sun et al. 2013) (**Figure 1.4**). In *Physcomitrella patens*, *Pp4CL1* and *Pp4CL4* have five exons and four introns, while *Pp4CL2* and *Pp4CL3* have four exons and three introns of similar size (Silber et al. 2008). Sequence analysis of various 4CL isoforms shows that difference in number of exon and intron has resulted from interruption of exon. Variation in number of isoform most probably resulted from gene duplication events (Ehlting et al. 1999) **Figure 1.4**.

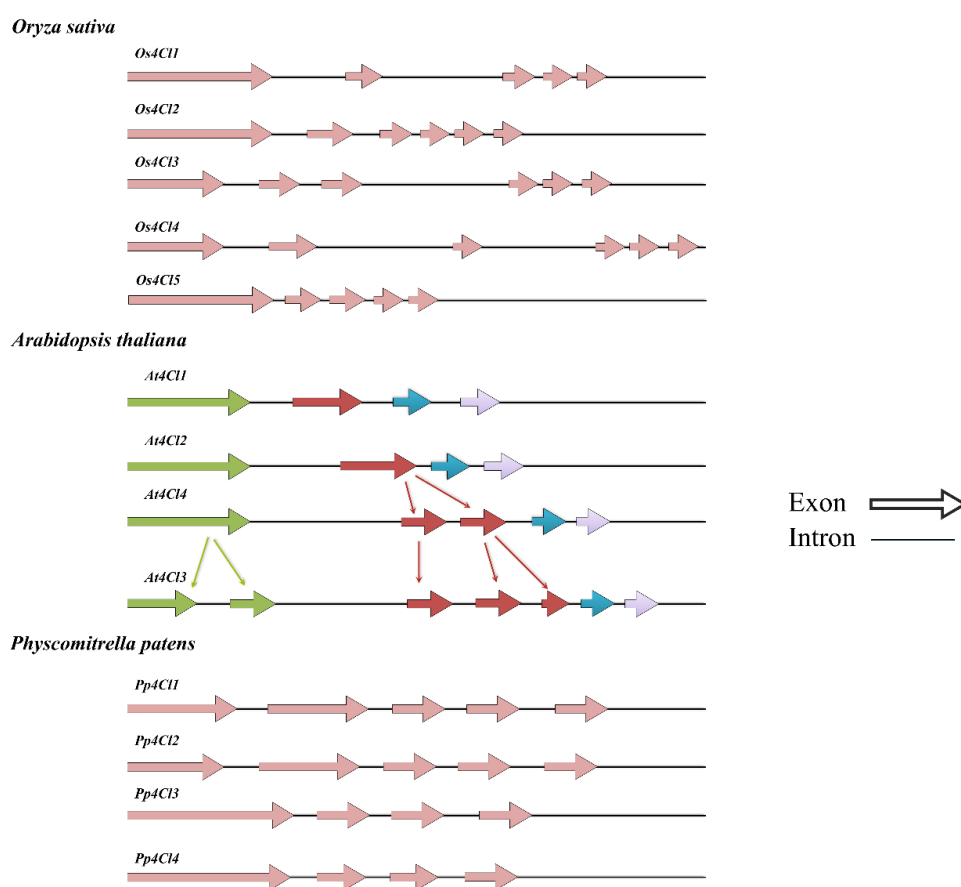


Figure 1.4 Gene structure of 4CLs from *Oryza sativa* ssp. japonica, *Arabidopsis thaliana*, and *Physcomitrella patens*.

1.5 Contribution of 4CL in lignification, flavonoid biosynthesis and stress response

The numerous products of phenylpropanoid pathway play a vital role in plants, viz. adaptation, growth and development, reproduction, protection against biotic and abiotic stresses, etc. Here we emphasize role of 4CL in biosynthesis of lignins, flavonoids, phenylpropanoids and other compounds for mechanical support and protection from biotic and abiotic stresses.

1.5.1 Lignification

Lignin is the second most abundant polymer after cellulose and present in secondary cell wall of all vascular plants. Complexity of lignin structure depends on the proportion of its constituent monolignol derivatives, namely H, G, and S. These monolignol derivatives synthesized via multiple routes in phenylpropanoid pathway (Naik et al. 2018). Effect of

individual 4CL and 4CL complex (Ptr4CL3-Ptr4CL5; *Populus tremuloides*) on lignin content was studied for its different aspects such as steady-state flux distribution, robustness and homeostatic properties by using mathematical modeling. Results of this mathematical modeling suggest that robustness and stability of pathway for S and G monolignol biosynthesis increased in the presence of Ptr4CL3-Ptr4CL5 complex (Naik et al. 2018). In aspen (*Populus tremuloides*), *Pt4CL1* gene is specifically expressed in lignin-containing tissue such as xylem, whereas *Pt4CL2* expressed in epidermal layers of stem and leaf. *Pt4CL1* shows the highest activity with 5-hydroxy ferulic acid, whereas *Pt4CL2* is inactive with 5-hydroxyferulic acid and has highest utilization rate with coumarate. The compartmentalized expression and substrate preferences suggest that *Pt4CL1* and *Pt4CL2* are involved in biosynthesis of lignin and other phenylpropanoids, respectively (Hu et al. 1998; Sutela et al. 2014). Three 4CL protein isoforms were detected in hybrid poplar (*Populus trichocarpa* and *Populus deltoides*) using fast-protein liquid chromatography. All of them utilize hydroxycinnamic acid but are inactive against sinapic acid (Allina et al. 1998). In this hybrid, 4CL1 is preferably expressed in old leaves, green stem and xylem while 4CL2 is expressed in young leaves (Allina et al. 1998). Five Pto4CL isoforms were characterized in *Populus tomentosa* by expression analysis and activity of recombinant enzymes with different substrates. All five isoforms have different substrate specificities and turnover rates. None of them is able to utilize sinapate as substrate. Over-expression of all these five isoforms leads to significant increase in level of lignin. In case of Pto4CL4, over-expressed transgenic tobacco increase in naringenin content was observed (Rao et al. 2015). Crude proteins extracted from developing xylem of *Robinia pseudoacacia* have three 4CL isoforms. The isoform Rp4CL1 preferably utilize *p*-coumarate as substrate, but unable to utilize ferulate and sinapate. Rp4CL2 and Rp4CL3 utilize sinapate and also show high activity with caffeate and *p*-coumarate. The crude extract from the shoots also has very similar substrate preference pattern. These results suggest that sinapate activating Rp4CL isoforms are constitutively expressed in lignin-forming cells (Hamada et al. 2004).

1.5.2 Flavonoid biosynthesis

Plant produces flavonoids, hydroxycinnamic acids and their related compounds. These compounds have wide range roles in plant including attractant, deterrent, symbiotic and allelopathic interaction, and crucial for protection from UV radiation (Mierziak et al. 2014). These compounds are primarily present in epidermis of leaves, stems, apical meristem and pollen. Flavonoids with conjugated double bonds are potent antioxidant than single conjugated bond. Furthermore, they undergo functional group modification (methylation and

glycosylation) to alter the reactivity, solubility and stability. In addition, they are known to protect from reactive oxygen species (ROS) by suppressing singlet oxygen, inhibiting enzymes that generate ROS (cyclooxygenase, lipoxygenase, monooxygenase, xanthine oxidase), chelating ions, quenching of free-radical and recycling of other antioxidants. All UV radiations induce synthesis of protecting flavonoids but comparatively, UVC and UVB induce more than UVA (Saewan and Jimtaisong 2013; Mierziak et al. 2014; Panche et al. 2016; Surjadinata et al. 2017; Zhao et al. 2017). Product of 4CL enzyme is required for flavonoid biosynthesis. Condensation of p-coumaroyl-CoA/cinnamoyl CoA molecule with three molecules of malonyl-CoA to yield chalcone, this reaction catalyzed by CHS. Then chalcone is isomerized to flavanone by CHI. Flavanone further utilized by several branch pathways to synthesize various flavonoids, including aurones, dihydrochalcones, flavanonols (dihydroflavonols), isoflavones, flavones, flavonols, leucoanthocyanidins, anthocyanins and proanthocyanidins (Hahlbrock and Scheel 1989; Dixon' and Paiva 1995; Holton and Cornish 1995; Mierziak et al. 2014). Specific 4CL isoform is found to be responsible for flux diversion toward flavonoid biosynthesis. In *Arabidopsis thaliana*, *At4CL3* is specifically expressed in light- exposed tissues such as leaves and flowers and positively correlated with flavonoid content of these tissues (Lee and Douglas 1996). Similar expression pattern is observed in case of rice *Os4CL2* gene in rice. *Os4CL2* was specifically expressed in the anther and expression is escalated by UV irradiation, suggesting its potential involvement in flux diversion for flavonoid biosynthesis (Sun et al. 2013). *Pueraria lobata* used as an herbal drug to prevent migraine, hypertension, alcoholism and cardiovascular disorder. Its main active ingredient includes isoflavonoids such as puerarin, daidzin, genistin and other compounds. In this plant, *Pl4CL* expression is found to be highest in root tissue. The level of puerarin is also highest in root. Upon treatment with methyl jasmonate (MeJA), expression level of *Pl4CL1* is upregulated and content on puerarin is also increased over three-fold. This suggests that *Pl4CL1* is responsible for biosynthesis of these isoflavonoids in *Pueraria lobata* (Li et al. 2014).

1.5.3 Protection from biotic and abiotic stresses

Various reports show that the level of 4CL increases upon biotic and abiotic stresses. Protection from this is mainly achieved by modulating the level of lignin, flavonoid and other secondary metabolites. In *Arabidopsis thaliana*, *At4CL* expression is studied in response to various stresses. In wound treatment the expression level of *At4CL1* and *At4CL2* was increased (Lee and Douglas 1996; Ehltng et al. 1999), while the level of *At4CL3* decreased (Soltani et al. 2006). *At4CL3* is involved in flavonoid biosynthesis while *At4CL1* and *At4CL2* involved in lignin

biosynthesis (Ehltling et al. 1999). In *Ocimum basilicum*, one of the *Ob4CL* was decreased in response to drought stress, and this might be because of 4CL involvement in the biosynthesis of metabolites other than lignin (Abdollahi Mandoulakani et al. 2017). *Physaria pruinosa* is salt tolerant species of poplar, while *Populus trichocarpa* is salt sensitive. When callus grown in salt stress condition, expression of Pp4CL2, Pp 4CL11, and Pp4CL12 is induced significantly in the resistant species compared to sensitive ones (Zhang et al. 2015). During infection of *Alternaria solani* to tomato (*Solanum lycopersicum* L.), the transcript level of 4CL gene is found to be upregulated (Shinde et al. 2017). These data suggest that in plant, 4CL play a role during biotic and abiotic stress.

1.6 Activity and expression of 4CL tightly regulated at various levels

1.6.1 Regulation at transcriptional level

The transcriptional control of phenylpropanoid enzymes is a key factor for regulation (Dixon' and Paiva 1995). Promoter region of *Ec4CL1* from *Eucalyptus camaldulensis* studied to investigate its role at transcriptional level regulation. The 1127 bp 5' upstream sequence contains various regulatory elements, including cis-regulatory and cis-acting regulatory elements. These elements include light-responsive, low temperature-responsive, abscisic acid-responsive, fungal elicitor responsive, meristem-specific activation and element for restricting vascular expression to the xylem tissue (Hue et al. 2016). Methylation pattern in the upstream promoter region of *St4CL1* and *St4CL1a* genes is studied in *Solanum tuberosum*. No difference is observed in the methylation pattern between elicitor stimulated (*Phytophthora infestans* culture filtrate) and non- elicitor stimulated (Becker-Andre et al. 1991). In case of *Pc4CL* promoter from *Petroselinum crispum*, a change in methylation pattern is observed when treated with UV light, but a similar change is not observed in case of *St4CL* promoter (Douglas et al. 1987; Becker-Andre et al. 1991). This difference in gene expression and methylation pattern is probably due to lack of two motifs in *St4CL*, which are present in *Pc4CL* promoter region centered around -57 and -127 relative to transcription start site (Becker-Andre et al. 1991) (**Figure 1.5; Table 1. 2**). Transcription factor MYB is known to regulate phenylpropanoid pathway enzymes including 4CL expression. Over-expression of *AtMYB4* in tobacco reduces the basal transcript levels of *C4H*, *4CL1* and *CAD* genes.

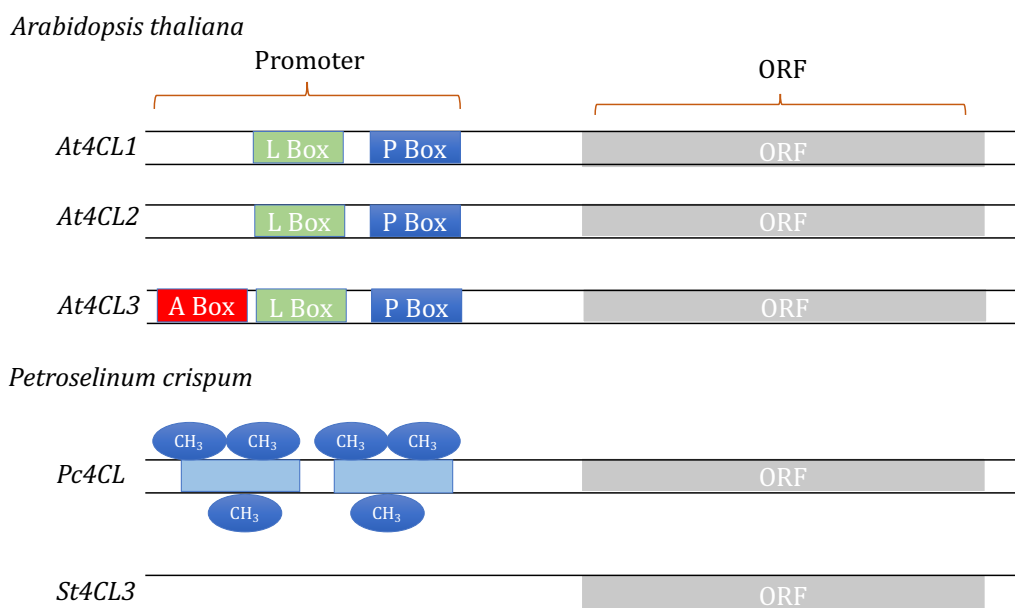


Figure 1.5 Structural features of promoter might be involved in regulation of 4CL gene expression. Presence of different boxes and methylation sites in promoter region of 4CLs might be involved in regulation of differential gene expression.

In addition, an *AtMYB4* mutant *Arabidopsis thaliana* shows enhanced tolerance to UV-B relative than wild type. Expression of *AtMYB4* is reduced upon exposure to UV-B light and wounding, leading to de-repression of C4H, resulting in higher synthesis of protecting sinapate esters. Upon UV-B exposure, over expressing *AtMYB4 Arabidopsis thaliana* plants are more sensitive, leading, leading to death (37% plants) while no death in wild-type plants. In addition, these lines have reduced sinapate ester level and unchanged flavonoid composition, this suggest that *AtMYB4* negatively regulates biosynthesis of UV protectant metabolites (H Jin 2000). In *Ipomoea batatas* (sweet potato), *IbMYB1a* transcription factor regulates expression of anthocyanin biosynthesis genes along with 4CL. *IbMYB1a* positively regulate multiple anthocyanin biosynthetic genes. This is confirmed by over-expression of *IbMYB1a* under different promoters in tobacco. These transgenic lines showed increased expression of genes encoding PAL, C4H, 4CL, CHS, CHI, F3H, DFR, and ANS (An et al. 2015) (Table 1.2).

Developmentally regulated and wound induced gene expression of 4CL is studied using fusion of 4CL promoter (truncated and full) with GUS (beta-glucuronidase). In *Arabidopsis thaliana* transformed with GUS coding region under control of 4CL promoter (full/truncated), expression of GUS::At4CL1 and GUS::At4CL2 is restricted to vascular tissues of root and aerial organs. The At4CL3::GUS expression is higher in non-vascular tissue (leaf and

cotyledons), upper part of hypocotyls, and roots whereas the *At4CL4::GUS* expression is restricted to roots only. Regulatory element analysis shows that -950 to -750bp region of the *At4CL2* promoter regulates early wound response, while region from -950 to -1600 bp negatively affects early wound response. Late wounding response of *At4CL2* may be because of presence of late wound response element in intron 1/2/3. The constructs with all intron show a strong response after 72 h of wound compared to a construct with less introns and without introns (Soltani et al. 2006).

1.6.2 Regulation by plant hormones

Hormones modulate metabolic pathways including phenylpropanoid pathway by regulating *4CL* expression (**Table 1.2**). For example, two-month-old *Plagiochasma appendiculatum* callus treated with abscisic acid (ABA), salicylic acid (SA) and MeJA, *Pa4CL1* showed upregulation in case of SA and MeJA treatment while downregulation in response to ABA treatment (Gao et al. 2015). *Hc4CL* from *Hibiscus cannabinus* showed downregulation when treated with MeJA and SA. However, in ABA treatment, *Hc4Cl* transcript level slightly decreased in 1 h and then subsequently increased to maximum in 24 h (Chowdhury et al. 2013). Expression of *Nt4CL1* and *Nt4CL2* (*Nicotiana tabacum*) was induced upon wounding and MeJA treatment (Lee and Douglas 1996). In *Salvia miltiorrhiza*, MeJA responsive element present in *SmC4H1*, *Sm4CL2*, *Sm4CL3*, *Sm4CL-like1*, *Sm4CL-like2*, *Sm4CL-like3*, *Sm4CL-like6*, *Sm4CL-like7* and their expression may be regulated by MeJA (Wang et al. 2015). Information about the role of hormonal regulation in *4CL* genes from *Arabidopsis thaliana*, *Pennisetum purpureum*, *Morus notabili*, *Camellia sinensis* are reported in **Table 1.2**. Specific *4CL* isoform get upregulated or downregulated in response to particular type of hormone in different plants. For example, in response to SA treatment, *Pa4CL* in *Plagiochasma appendiculatum* is upregulated while, *Hc4CL* is downregulated in *Hibiscus cannabinus*. These findings suggest that each *4CL* isoform has differential response to stimuli.

Table 1.2 List of reported factors/elements that regulate 4CL in plants.

	Factor/Element	Effect	Reference
<i>Ip4CL</i>	IbMYB1a	Over-expression leads to anthocyanin accumulation	(An et al. 2015)
<i>Os4CL2</i>	Wounding UV radiation	Down-regulation Up-regulation	(Sun et al. 2013)
<i>Pa4CL</i>	SA and MeJA ABA	Up-regulation Down-regulation	(Gao et al. 2015) (Gao et al. 2015)
<i>Hc4CL</i>	SA and MeJA ABA	Down-regulation Up-regulation	(Chowdhury; et al. 2013) (Chowdhury; et al. 201)
<i>At4CL1</i> , <i>At4CL2</i>	Wound and MeJA <i>Peronospora parasitica</i> infection and wounding	Up-regulation Up-regulation	(Lee and Douglas 1996) (Ehltling et al. 1999)
<i>At4CL3</i>	<i>Peronospora parasitica</i> infection and wounding	No change	(Ehltling et al. 1999)
<i>Ob4CL</i>	Drought stress	Down-regulation	(Mandoulakani et al. 2017)
<i>Pp4CL</i>	ABA, MeJA and GA	Up-regulation	(Peng et al. 2016)
<i>Pe4CL2</i> , <i>Pe4CL11</i> , <i>Pe4CL12</i>	NaCl stress	Change depends upon genotypes	(Zhang et al. 2015)
<i>Pe4CL5</i>	NaCl stress	Down-regulation	(Zhang et al. 2015)
<i>Pe4CL9</i> , <i>Pe4CL10</i>	NaCl stress	No change	(Zhang et al. 2015)
<i>Ma4CL3</i>	Wounding, salicylic acid, and ultraviolet treatments	Up-regulation	(CH Wang 2016)
<i>Mt4CL</i>	Aluminium stress	Up-regulation	(Chandran et al. 2008)
<i>Cs4CL</i>	Catechin treatment Drought stress, ABA and GA3 Wounding	Down-regulation Down-regulation, Decrease in Catechin content Up-regulation, Increase in Catechin content	(Rani et al. 2009) (Rani et al. 2009) (Rani et al. 2009)
<i>Sa4CL3</i>	Light exposer	Up-regulation	(Gaid et al. 2011)
<i>St4CLs*</i>	<i>Phytophthora infestans</i> infection and arachidonic acid	Up-regulation	(Fritzemeier et al. 1987)
<i>Ps4CLs*</i>	<i>Phytophthora megasperma</i> glycinea infection	High expression in epidermal cells, oil- duct epithelial cells and developing xylem	(Schmelzer et al. 1989)
<i>Pv4CL</i>	<i>Sclerotinia sclerotiorum</i>	Up-regulation	(Oliveira et al. 2015)

Ib, *Ipomoea batatas* L.; *Os*, *Oriza sativa* L.; *Pa*, *Plagiochasma appendiculatum* Lehm. & Lindenb.; *Hc*, *Hibiscus cannabifolius* L.; *At*, *Arabidopsis thaliana* L.; *Ob*, *Ocimum basilicum* L.; *Pp*, *Pennisetum purpureum* Schumach. ; *Pe*, *Populus euphratica* Oliv.; *Ma*, *Morus notabilis* L.; *Mt*, *Medicago truncatula* L.; *Cs*, *Camellia sinensis* L.; *Sa*, *Sorbus aucuparia* L.; *St*, *Solanum tuberosum* L.; *Ps*, *Petroselinum sativum* L.; *Pv*, *Phaseolus vulgaris* L.; * analysis using northern blot technique.

1.6.3 Regulation in response to biotic and abiotic stresses

Several reports suggested that the *4CL* gene expression is regulated by various abiotic stresses such as drought, salinity, temperature, etc. (**Table 1.2**). In *Arabidopsis thaliana*, upon wounding, expression of *At4CL1* and *At4CL2* increased after 2.5 h and then reversed to the basal level of expression, but again increased to maximum after 48 h (Ehltling et al. 1999). Similar gene expression patterns like *At4CL1* and *At4CL2* were shown by *Hc4CL* in *Hibiscus cannabinus* plant in response to wounding (Choudhary et al, 2013). The *At4CL3* expression level showed downregulation that became normal by 4 h and then simultaneously increased up to 72 h post wounding (Soltani et al. 2006). *At4CL4* expression also increased up to 2.5 h and remained higher up to 12 h and then fall down to basal level. Expression pattern of *At4CL3* show by is Soltani et al. (2006) is different from pattern shown by Ehltling et al. (1999). In *Oryza sativa japonica*, upon wounding *Os4CL3*, *Os4CL4*, *Os4CL5* and other genes of phenylpropanoid biosynthesis pathway were significantly upregulated, whereas *Os4CL1* and *Os4CL2* were downregulated. When dark-adapted rice plants were illuminated with UV light, *Os4CL1*, *Os4CL3*, *Os4CL4*, and *Os4CL5* were downregulated and *Os4CL2* upregulated. It concluded that the *Os4CL2* is probably involved in biosynthesis of flavonol (Sun et al. 2013). When *Ocimum basilicum* plants were subjected to controlled drought stress condition, the expression levels of *Ob4CL* and *ObC4H* genes decreased, while the expression levels of *CVOMT* and *EOMT* were increased and *CAD* expression was relatively unchanged. This has been correlated with essential oil compounds and observed that highest amount of methyl chavicol, methyl eugenol, α -bergamotene, and β -myrcene. This suggests that altered gene expression in response to drought stress increases methylchavicol and methyleugenol content (Mandoulakani et al. 2017). In response to NaCl treatment *PAL*, *CCoAOMT*, *C3H*, *HCT* and *F5H* genes were upregulated, which are involved in lignin biosynthesis (Choudhary et al, 2013; Kim et al, 2013). It clearly indicated that 4CL has a significant role in countering various abiotic stresses. This has been also proven in other plants like *Populus euphratica*, *Morus notabili*, *Medicago truncatula*, *Camellia sinensis*, and *Sorbus aucuparia* (**Table 1.2**).

In response to biotic stresses like bacterial/fungal/virus infection and insect infestation plant combat through phenylpropanoid modulation. The 4CL of phenylpropanoids pathway is one of the key enzymes and its expression altered in response to biotic stresses (**Table 1.2**), indicating that 4CL has a significant role in counteracting various biotic stresses. Potato leaves infected with *Phytophthora infestans* leads to rapid accumulation of *PAL* and *4CL* and other genes from phenylpropanoid pathway. *Phytophthora infestans* is one of the most destructive fungal pathogens and leads to rapid browning and hypersensitive cell death around the

infection site. In this study, mRNA of both *PAL* and *4CL* gene has maximum translational activity after 2 h and enzyme activity increased up to two-fold in 6 to 12 h after infection (Fritzemeier et al. 1987). Similar type upregulation of *4CLs* gene observed in case of following plant-pathogen interaction *Arabidopsis thaliana-Peronospora parasitica*, *Solanum tuberosum*, *Phytophthora infestans*, *Petroselinum sativum*, *Phytophthora megasperma*, *Phaseolus vulgaris* and *Sclerotinia sclerotiorum* (**Table 1.2**). Grapevine red blotch-associated virus is major problem to cultivated grapevines (*Vitis vinifera*). Infection of this virus leads to downregulation of phenylpropanoid pathway genes along with *4CL* at the ripening stage (Blanco-Ulate et al. 2017).

1.7 Functional specificity of 4CL isoforms: a case study

Co-expression analysis of *At4CL1*, *At4CL3* from *Arabidopsis thaliana* and *Gm4CL4* from *Glycine max* was performed individually using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, all the parameters were kept as default except number of interactions to show was set to 20) database and merged (Szklarczyk et al. 2015) (**Figure 1.6**).

At4CL1 co-expressed with lignin biosynthesis genes while *At4CL3* with flavonoid biosynthesis genes. This shows that each *4CL* isoform may be responsible for metabolic flux diversion. Both *At4CL1* and *At4CL3* also show co-expression with genes of other metabolite biosynthetic pathway (other than lignin and flavonoid biosynthesis genes). *Gm4CL4* shows co-expression with genes for biosynthesis of lignin, flavonoid, other metabolites and uncharacterized genes (Figure 5). Based on *4CL*'s co-expression analysis from *Arabidopsis thaliana* (*At4CL1* and *At4CL2*) and *Glycine max* (*Gm4CL4*), it showed that some plants may have single isoform for both lignin and flavonoid and some plant may have separate isoform for each of the class.

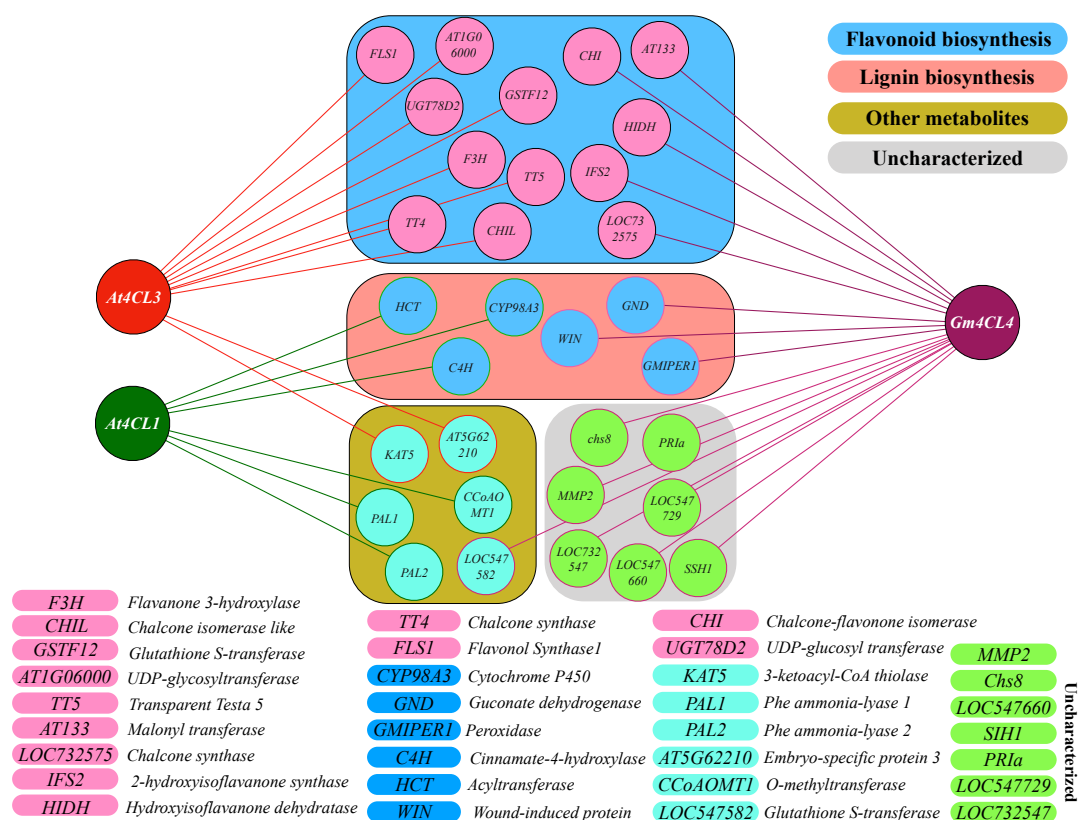


Figure 1.6 Co-expression analysis of *At4CL1*, *At4CL3* and *Gm4CL4*.

1.8 Scope of 4CLs in metabolic pathway engineering for various products

Enzymes of phenylpropanoid and flavonoid metabolism present in the form of membrane-associated multienzyme complexes as a metabolon on the cytosolic face of the endoplasmic reticulum (Hrazdina and Wagner 1985). These organized multienzyme complexes are efficient in channeling of intermediates between enzymes of sequential reactions (Winkel-Shirley 1999; Burbulis and Winkel-Shirley 1999). Interactions and organization of metabolon may vary in different plant species (Fujino et al. 2018). As per reviewed by Chemler and Koffas (2008) it is possible to synthesize flavonoids by gene cloning and expression of 4CL and other enzymes involved in its biosynthesis in bacterial (*Escherichia coli*) and yeast (*Saccharomyces cerevisiae*). 4CLs has been used for modification of crop metabolites for economic benefit and also used for biosynthesis of commercially important metabolites.

1.8.1 Fuel and paper industry

Lignin along with cellulose microfibrils provides rigidity to plant. Lignin content affects utilization of plant material for chemicals, fiber and energy production, etc. Lignin is one of

the products of phenylpropanoid pathway and expression of specific 4CL isoforms plays role in lignin biosynthesis (Costa et al. 2005; Shigeto et al. 2017). Transgenic aspen (*Populus tremuloides* Michx.) in which *Pt4CL1gene* was silenced showed downregulation of lignin biosynthesis. Such plants showed reduced lignin up to 45% while 15% increased cellulose content. Silenced lines have a thicker stem, longer internodes, larger leaves and an overall enhanced growth rate than the control (Hu et al. 1999). Sugarcane is a best raw material for bioethanol production. Two 4CL isoforms are characterized from *Saccharum* spp. hybrids. Out of two, *Sh4CL1* is involved in lignin biosynthesis based on phylogenetic analysis and RNAi silencing. RNAi suppression of *Sh4CL1* leads to a decrease in 16.5% lignin content and shows improvement of 52 to 76% saccharification rate compared to wild type control. This suggests that 4CL silencing can be used to improve lignocellulosic raw material for biofuel production (Jung et al. 2016).

1.8.2 Natural product industry

Various valuable natural products (NPs) are available from the general phenylpropanoid pathway. The use of 4CL to produce high-value NPs has already been proven in bacterial, fungal and plant systems. Resveratrol is an important secondary metabolite, which has health-promoting properties and found in red wine (Giovinazzo et al. 2012). Fusion of *At4CL1* from *Arabidopsis thaliana* and stilbene synthase (VvSTS) from *Vitis vinifera* was used to improve resveratrol synthesis in yeast. With this fusion, 15-fold more resveratrol level achieved compared to yeast individually expressing these enzymes (Wang et al. 2011). Rosmarinic acid is also one of the important products of phenylpropanoid pathway produced by fusion of 4-hydroxyphenyllactic acid with 4-coumaroyl-CoA, this reaction is catalyzed by hydroxycinnamoyl-CoA: hydroxycinnamoyl transferase. It has various biological activities including antiviral, antibacterial, antioxidant and anti-inflammatory (Petersen and Simmonds 2003). Rosmarinic acid accumulation increased from 2.1 to 3.9-fold in response to MeJA treatment in cell cultures of *Agastache rugosa* Kuntze. Transcript level of *ArPAL*, *Ar4CL*, and *ArC4H* were increased up to 4.5-, 3.4- and 3.5-fold, respectively, compared to untreated culture cells (Kim et al. 2013). Plants belonging to order Zingiberales produce curcuminoids that has wide range of applications including food additive, stimulant, food colouring agent, anti-tumour, antioxidant and hepatoprotective, etc (Amalraj et al. 2017). Biosynthesis of curcuminoids in *Escherichia coli* was achieved by cloning of enzymes: 4CL from *Lithospermum erythrorhizon* and curcuminoid synthase (CUS) from rice. Ferulic acid from rice bran pitch, an industrial waste during rice edible oil production was used during growth of

recombinant *Escherichia coli* expressing 4CL and CUS for production of curcuminoids (Katsuyama et al. 2008). Os4CL in *Ocimum sanctum* is responsible for partition of metabolite flux towards eugenol biosynthesis (Rastogi et al, 2013). Eugenol inhibits advanced glycation end products in diabetes (Singh et al. 2016). 4CL has a prominent role in natural product synthesis improvement.

1.8.3 Food technology

The expression pattern of genes involved monolignol and flavonoid biosynthesis are studied during the fruit ripening in apple. Differential expression of PAL and 4CL isoforms leads to suppressing lignin biosynthesis and strong induction of flavonoid biosynthesis during fruit ripening (Baldi et al. 2017). Results show that fine-tuning of the pathway was achieved by differential expression of branch point enzyme isoforms such as PAL and 4CL. The differential expression of a specific 4CL isoform was observed at different stages of pear fruit development (Cao et al. 2016). Thus, using 4CL manipulation for metabolic pathway engineering is useful in post-harvesting and food technology-related industries.

1.9 Conclusions and future perspectives

1. The 4CL is essentially involved in channeling precursors for different phenylpropanoid biosynthesis. Based on phylogenetic analysis, 4CL divided into four clusters: type I involved in monolignol biosynthesis, type II involved in phenylpropanoid biosynthesis other than lignin, type III and IV are present in monocots with similar functions as type I and II, respectively. The 4CL genes in monocots and dicots are evolved independently after their separation. However, type I and II might have been diverged before the divergence of gymnosperm and angiosperm lineages. The 4CL activity and expression are tightly regulated spatio-temporal and in response to biotic and abiotic stresses.
2. The role of amino acids in SBP and domains of 4CL is well studied with the help of techniques such as bioinformatics tools, crystallography, site directed mutagenesis, etc. Its activity has been successfully improved by site- directed mutagenesis and domain exchange approaches.
3. Co-expression analysis of characterized 4CL from various plants suggests that single 4CL is required for the lignin and flavonoid biosynthesis, whereas some plants have separate 4CL's for these biosynthetic pathways.
4. The metabolic pathways are regulated at entry point, branch point and/or in some cases by intermediate enzymes for its final product biosynthesis. However, terminal enzymes

are well studied in most of metabolic pathways. The 4CL enzymes are mostly studied in the context of functional characterization and regulation by external factors. The effect of internal factor/triggers on 4CL function and regulation remains enigmatic, which is an important aspect of harnessing further applications in metabolic engineering. It has been established that specific PAL isoforms also play an essential role in the regulation of phenylpropanoid pathway, which is the first committed step of this pathway.

5. The detailed validation is a prudent necessity for efficient metabolic pathway engineering: (i) the role of intermediate and end product metabolites during specific 4CL isoform activity; (ii) involvement of transcription factors in regulating the levels of individual 4CL isoforms; (iii) cross-talk with other pathways; (iv) splicing mechanism in the formation of multiple isoforms; (v) regulation at -post-translational level; (vi) activity optimization in bacterial or yeast host system, etc. Such information is highly imperative to achieve efficient metabolic pathway engineering for commercially important candidate metabolites.

1.10 Objectives of the work

Ocimum is known for several important medicinal properties, including antimicrobial activity, anti-inflammation activity, anti-stress activity, antidiabetic, antioxidant and wound healing, etc. Exact compounds or groups of compounds associated with these activities are not known in *Ocimum* species. Metabolic analysis revealed that *Ocimum* species are rich in either phenylpropanoids (*O. gratissimum*) or terpenoid (*O. kilimandscharicum*) or they may have a combination of both compounds in substantial amounts (*O. canum*) (Singh et al. 2015; Gurav et al. 2021). Most of the *Ocimum* species of 4CL enzymes are not characterized besides having an important role in the phenylpropanoid biosynthesis pathway. Functional characterization of 4CL isoforms in *Ocimum* will help to understand the role of individual isoforms in phenylpropanoid diversity and plant physiology. Knowledge about the contribution of Ok4CLs in the modulation of flux towards either lignin or non-structural phenylpropanoids will be useful to improve plant physiology or essential metabolite content in this important plant species. Further, this knowledge could be used for developing or screening variety with high essential oil production. Based on available information, the major objective of this study is the characterization of 4-Coumarate CoA-Ligase isoforms from *Ocimum kilimandscharicum*.

On this background, we have defined the objectives of the thesis work as follows:

1. Functional insights into two 4-coumarate CoA ligase from *O.****kilimandscharicum***

- Selection of putative *Ok4CLs* genes from *O. kilimandscharicum* transcriptome data
- Gene expression analysis of *Ok4CLs* in different tissues of *O. kilimandscharicum*
- Gene cloning, expression, purification and *in vitro* functional analysis of *Ok4CLs*

2. Determination of active site residues of *Ok4CLs* responsible for substrate specificity

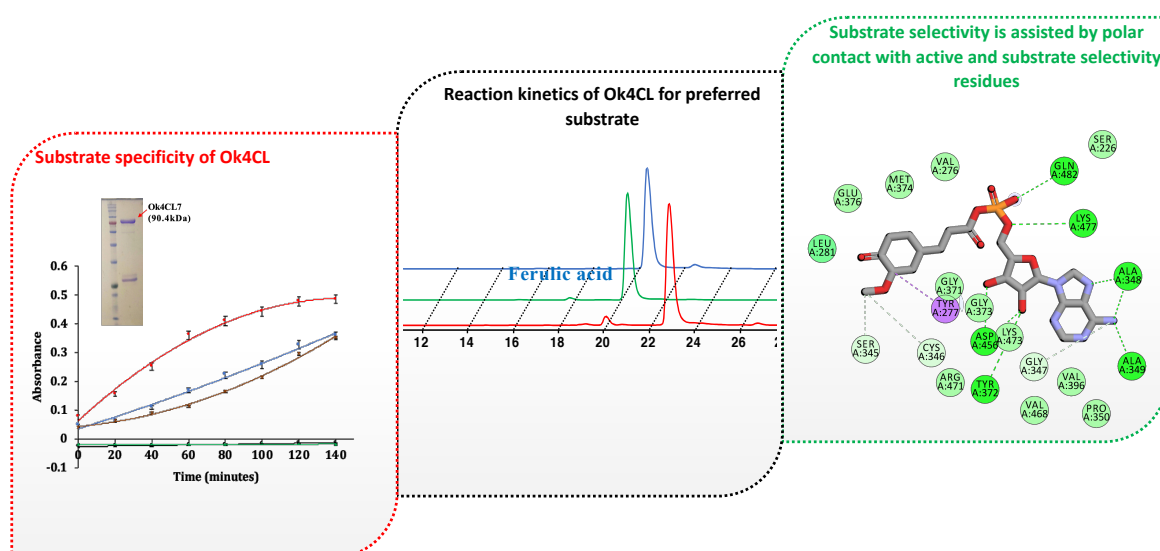
- *In silico* analysis of enzyme-substrate interaction using Molecular docking
- Development of mutant models to modulate the enzyme-substrate interaction
- Development of mutant *Ok4CL* enzymes and analysis of their interaction with substrates using enzyme kinetics study

3. Functional characterization of *Ok4CLs* in *Nicotiana benthamiana*

- Construct preparation for localization and plant transformation
- Development of *N. benthamiana* transgenic lines for *Ok4CL* gene overexpression
- Morphology, gene expression and metabolite analysis of transgenic *N. benthamiana* lines

Chapter 2

Diversity of 4-coumarate-CoA ligases in *Ocimum kilimandscharicum* and functional characterization of two candidate Ok4CLs



Contents of Chapter 2 have been published in the research article listed below.

Lavhale, S.G., Joshi, R.S., Kumar, Y., Giri, A.P., Functional insights into two *Ocimum kilimandscharicum* 4-coumarate-CoA ligases involved in phenylpropanoid biosynthesis. International Journal of Biological Macromolecules 2021, 181: 202-210. (Cited 3 times)

2.1 Introduction

Phenylpropanoids protect plants from various biotic or abiotic stress conditions. Many of them are known to have or being explored for medicinal properties. 4-coumarate CoA ligase (EC 6.2.1.12) (4CL) is a crucial enzyme involved in the phenylpropanoid pathway. It serves as the main branch point in the pathway (Gui et al. 2011). They are encoded by a multigene family of adenylate forming enzymes. 4CLs convert hydroxy or methoxy cinnamic acid derivatives to the corresponding activated thioesters. Products of 4CL are utilized by various oxygenase, reductases, and transferases for the biosynthesis of lignin, flavonoids, anthocyanins, tannins, aurones, stilbenes, coumarins, suberin, cutin, sporopollenin, etc. (Vogt 2010). It has been reported that 4CLs might play a crucial role in lignin biosynthesis (Kollmann et al. 1968; Boerjan et al. 2003; Vanholme et al. 2010). Lignin is the polymer of H-lignin, G-lignin and S-lignin monomers. Monolignols of H-lignin, G-lignin and S-lignin are synthesized from *p*-coumaroyl CoA, feruloyl CoA and sinapoyl CoA, respectively (Whetten⁹ and Sederoff^{avb} 1995; Boerjan et al. 2003; Vanholme et al. 2010). Also, several non-structural phenylpropanoids synthesized *via* 4CLs exhibit diverse functions in plant physiology. Amongst these, flavonols and isoflavones negatively regulate the transport of auxin hormone (Deng and Lu 2017). They are also involved in the attraction of symbiotic bacteria for nitrogen fixation (Hassan and Mathesius 2012; Dastmalchi and Dhaubhadel 2015). Moreover, flavonoids protect the plant against UV-B irradiation, sugar stress, nutrition deficiency (low phosphate/iron/nitrogen), low temperature, drought, pathogen infection, and herbivores attack (Shah and Smith 2020). Anthocyanins, another metabolite from this class, shows various colors ranging from orange to pink, red, and purple. Anthocyanins impart pigmentations to flowers, fruits and seeds to attract pollinators and seed dispersal (Shi and Xie 2014). These potential features of 4CL in plant physiology urge the importance of molecular investigation in the functionality of 4CLs in phenylpropanoid rich medicinal plants like species of *Ocimum*.

Ocimum has remarkable diversity in metabolite contents with a variety of medicinal properties (Singh et al. 2015; Gurav et al. 2022). Functions of *Ocimum* 4CLs beyond phenylpropanoid biosynthesis are still enigmatic. Gene expression profiles and quantification of metabolite contents showed that eugenol and camphor biosynthesis is tissue-specific in *O. kilimandscharicum* (Singh et al. 2020). A particular 4CL isoform was co-expressed with *EUGENOL SYNTHASE 1* in leaf tissue (Singh et al. 2020). This chapter deals with the analysis of 4CL diversity using the phylogenetic similarity search network and gene expression in various tissues. Following this, we have performed the structural-functional characterization

of two recombinant 4CL isoforms from *O. kilimandscharicum* (Ok4CL7 and -15). Characterization of two recombinant Ok4CL7 and -15 might shed light on their putative functions in the *O. kilimandscharicum* phenylpropanoid and lignin pathway.

2.2 Materials and methods

2.2.1 Plant material and chemicals

O. kilimandscharicum plants were grown in the field at CSIR-National Chemical Laboratory, Pune, India. After flowering, tissues (flowers, young leaves, roots, and trichomes) were harvested and stored in liquid nitrogen for further analysis. Substrates of 4CL, such as *p*-coumaric acid, cinnamic acid, ferulic acid, caffeic acid and sinapic acid (Sigma-Aldrich, St. Louis, USA), Coenzyme A (Sisco Research Laboratories, Mumbai, India), complementary DNA (cDNA) synthesis kit (Applied Biosystems, Waltham, USA), RNA isolation total plant RNA extraction-Spectrum kit (Sigma Aldrich) were purchased. The plasmids like pGEM-T, pGEX 4T, pET28a and pRI101AN were purchased from Takara (Takara, Kyoto, Japan) and the solvents for High-Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS) from Merck (Kenilworth, NJ, USA).

2.2.2 Identification of 4-CLs from *O. kilimandscharicum* transcriptome datasets

Total RNA extracted from 100 mg tissue of flowers, young leaves, roots, and trichome using Spectrum Plant RNA Isolation kit (Sigma-Aldrich). The quality of RNA was checked on the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The transcriptome was sequenced using the Illumina *NextSeq500* next-generation sequencing (NGS) (San Diego, USA) platform. De novo assembly was performed using Trinity software without a reference genome (Grabherr et al. 2011) and transcript annotation using NCBI-BLAST-2.2.29+ (Altschul et al. 1990). Transcriptome Sequencing and analysis is outsourced at GenotypicTechnology, Bangalore, India. Annotated 4-Coumarate CoA Ligase (4CL) encoding genes were selected from the de novo assembled *O. kilimandscharicum* transcriptome. Open Reading Frame (ORF) analysis was performed using the ORF Finder web tool from NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>). Sequences having the start and stop codons as well as the presence of two conserved signature motifs (Box-I and Box-II) were selected for further study (**Figure 2.1 A and B**). All those enzymes having Box-I (SSGTTGLPKG_V) are grouped into a superfamily of adenylate-forming enzymes. These enzymes produce adenylated intermediate during the catalysis. Fifteen ORF sequences from

O. kilimandscharicum transcriptome had these two conserved motifs, and they were selected for further analyses (Ok4CL1 to 14).

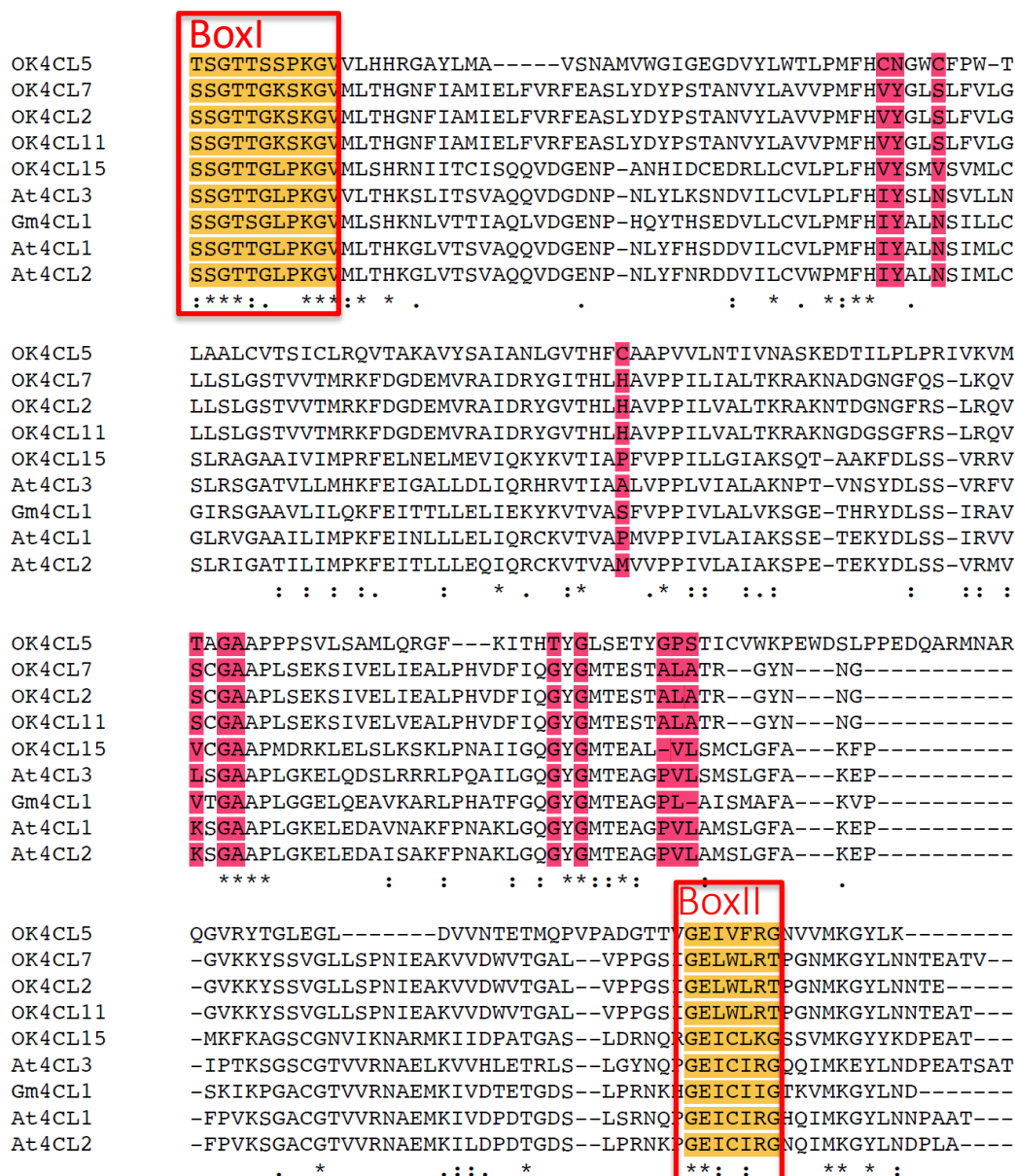


Figure 2.1 A. Sequence Alignment of 4CLs from *O. kilimandscharicum* and a representative from *A. thaliana G. max*. Substrate binding residues highlighted in pink colour; Box-I: Adenylate forming domain and Box-II: Required for catalysis.

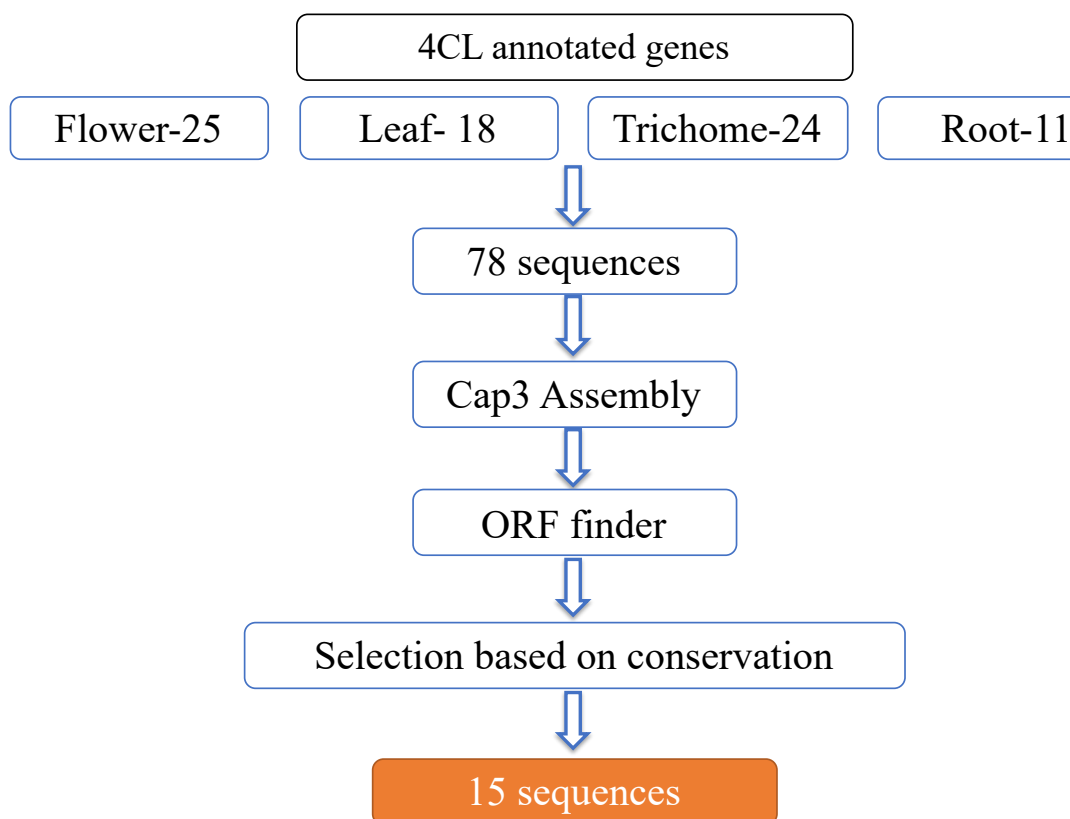


Figure 2.1 B Flowchart of analysis performed to select putative 4CL sequences from *O. kilimandscharicum* transcriptome data.

2.2.3 Phylogenetic, sequence similarity network and domain analysis

Phylogenetic analysis of 4CLs was performed using MEGA 6 software (Tamura et al., 2013). Deduced amino acid sequences of all putative 4CLs from *O. kilimandscharicum* and reported 4CLs were aligned using the MUSCLE alignment tool (Edgar 2004). Reported 946 plant 4CL sequences were retrieved from the UniProt protein database (partial, hypothetical and putative sequences were excluded). Aligned sequences in fasta file format used for phylogenetic tree construction using the neighbour-joining method with 1000 bootstrap value and by keeping the rest of the parameters default. The relation among 4CL sequences was also studied by Sequence Similarity Network (SSN) using SSNpipe with default protocol (<https://github.com/ahvdk/SSNpipe>, <https://github.com/ahvdk/SSNpipe/wiki/SSNpipe-Usage-Examples>). BLAST e-value used for edge distance calculations. Cut-off value $1e^{140}$ was used

for network building as it is the highest to achieve confident clustering. The output network was visualized using Cytoscape 3.7.2. Domain architecture of selected 4CLs studied using Conserved Domains Database (CDD) (Lu et al. 2020).

2.2.4 Gene expression analysis

Gene expression of putative 4CLs studied in different tissues from *O. kilimandscharicum*. Total RNA was isolated from 100 mg of freshly collected tissue, using Spectrum Plant RNA extraction kit as described (Anand et al. 2016) (Figure 2.2). Genomic DNA contamination in isolated RNA was removed using RNase free DNase treatment.



Figure 2.2 RNA isolation from different tissue of *O. kilimandscharicum* plants.

Two μg RNA was used for cDNA synthesis using SuperScript III reverse transcriptase system (Invitrogen, Carlsbad, USA). cDNA synthesis was confirmed by PCR using a set of primers that amplifies a short sequence elongation factor 1 α (EF1 α). qRT-PCR primers were designed for all putative 15 *Ok4CL* genes using Primer 3.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and Oligo Analysis online tool (<http://www.operon.com/tools/oligo-analysis-tool.aspx>). qRT-PCR primers were synthesized (Eurofins, Bangalore, India) and used for analysis. qRT-PCR was performed on an Applied Biosystems 7500 Fast Real-Time PCR System using the SYBR green protocol. Three biological and technical replicates were used for each sample. The reaction mixture contains 5 μl of SYBR green master-mix, 0.5 μl of 10 μM forward and reverse gene-specific primers (Table 2.1) and 1 μl of diluted cDNA (1:2) with nuclease-free water added to make up a volume to 10 μl . Standard plots of putative 4CLs were generated using a gene-specific set of primer pairs. Different cDNA dilutions (1:2, 1:3, 1:8, 1:16 and 1:32) were used to generate standard plots. The qRT-PCR reaction for sample cDNA was optimized using different concentrations of primer and cDNA. Absolute gene expression was performed using the standard plot method. Heatmap of *Ok4CL* genes expression across various tissues of *O.*

kilimandscharicum was illustrated using normalized z-score. Multiple Experimental Viewer (MeV) (<http://mev.tm4.org/#/welcome>) software was used to plot the heatmap.

2.2.5 Cloning, recombinant protein expression and purification of Ok4CL7 and -15

As the representative of different clades, three candidate Ok4CLs, Ok4CL11, -7 and -15, were selected for heterologous expression and *in vitro* characterization. Putative full-length coding sequences of *Ok4CL7* (1731 bp), *Ok4CL11* (1731 bp), *Ok4CL15* (1608 bp) were amplified from cDNA by PCR using a high-fidelity Q5 polymerase (New England Biolabs, Ipswich, USA). Amplified sequences were cloned into expression vectors (*Ok4CL7* in pGEX 4T and *Ok4CL15* in pET 28a) using T4 DNA Ligase (New England Biolabs). *NdeI* and *EcoRI* restriction sites used for *Ok4CL7* and 11 cloning, while *NdeI* and *Sall* used for OK4CL15. These restriction sites were included in forward and reverse primers, respectively (**Table 2.1**).

Table 2.1 Primers used for qRT PCR and full-length cloning of *Ok4CL7* and -15.

Primer name	Nucleotide sequence
OK4CL1_RT_F	CCTCCATTATGAAAGGTTACTTC
OK4CL1_RT_R	GGTGAGTAAGCAACAAAGCC
OK4CL2_RT_F	TTAGGTAATGCTTTGGAGGGTGTT
OK4CL2_RT_R	AAAATCGCGGCGGTGTCGTTT
OK4CL3_RT_F	ATGTGACGAAGGGCTACAAGAAT
OK4CL3_RT_R	ATGTGACGAAGGGCTACAAGAAT
OK4CL4_RT_F	CCAAAGGCGTTGTTTACAGC
OK4CL4_RT_R	TTTGTGAGTGGCTATGCTGTC
OK4CL5_RT_F	GCATATCTCATGGCCGTAAG
OK4CL5_RT_R	AGCACAAAAATGGGTGACACC
OK4CL6_RT_F	GATGACGAAGTCGGATTTGG
OK4CL6_RT_R	AGGCTCAGTTGTACGAGTG
OK4CL7_RT_F	TTAGGTAATGCATTGGAGGGTTGC
OK4CL7_RT_R	AGTATAAAATCGCGGCGGTGTCCT
OK4CL8_RT_F	TATGGGATGACTGAAGCGGGGCCA
OK4CL8_RT_R	GTAGACTCCGGATCATTGAGAT

OK4CL9_RT_F	GGACAGGGTTATGGGATGACC
OK4CL9_RT_R	GTAAGAAACAAAATACGTTG
OK4CL10_RT_F	AGCAGCATTTCAGAGCAATAAATC
OK4CL10_RT_R	TGAAACCAAGGCTTCGGAG
OK4CL11_RT_F	GTTTTGGATTTGAGCTGTGAGAA
OK4CL11_RT_R	CAGTAGTGCCAGACGAGTAC
OK4CL12_RT_F	TGTTGACCGCATCAAAGAGC
OK4CL12_RT_R	ACATCTTCTTCAGTCAGTGAGC
OK4CL13_RT_F	TGGAAGCCCTAATAGTGGATTC
OK4CL13_RT_R	ACGGTCCACAACATAGAGTC
OK4CL15_RT_F	ATTTGCCTCAAAGGAAGCTCC
OK4CL15_RT_R	ACAGCAGCATCAGATATGGAG
OK4CL16_RT_F	GACCTTGTTTGATAAACGGAGTG
OK4CL16_RT_R	AGGAAGGCGAAGACAACC
Ok4CL7 NdeI-FF	AATACATATGATGGCTGCATTAATAAAAGCCC
Ok4CL7 EcoRI-FR	AATAGAATTCCTAAAGCCTGGAAACAAACAAGTTC
Ok4CL15 NdeI-FF	AATACATATGATGGAGCTGAAAGAAGAGAAG
Ok4CL15 Sall-FR	AATAGTCGACTTACATTCTAGCTCTCAGATTCTTCC

The constructs having either of the *4CL* genes were transformed in ArcticExpress (DE3) competent cells (Agilent Technologies, Santa Clara, USA). A single colony of ArcticExpress cells having expression construct/plasmid was inoculated in primary culture and grown in a shaker incubator at 37°C overnight (~14 hrs.). 10 mL of inoculum used from primary culture for 1 litre of secondary culture. Secondary culture allowed growing up to log phase ($OD_{600} \sim 0.45$). Once the culture reaches to log-phase, then it is kept at 4°C for chilling. Gene expression was induced by adding 0.5 M IPTG to 1 L culture ($OD_{600} \sim 0.45$) and cultures grown at 12°C in a shaker incubator for overnight (~14 hrs.). The bacterial cell pellet was obtained by centrifugation of culture at 3500 g for 10 min. Pellet was resuspended in 10 mL lysis buffer (Tris [50 mM], NaCl [200 mM], glycerol [5% v/v] and lysozyme [0.2 mg/mL] pH 8.0), followed by sonication using VibraCell (Sonics, Newtown, USA) (10 seconds on and off cycle with an amplitude of 45%) for cell lysis. For recombinant proteins, buffer containing Tris

(50 mM), NaCl (200 mM), glycerol (5% v/v) was used throughout the purification process. Cell lysate/cell debris was separated using centrifugation at 11200 g for 60 min at 4°C. The supernatant was collected in a fresh tube. Supernatant further filtered using 0.2-micron syringe filters for removal of any solid matter. Clear supernatant along with slurry was kept for binding on a shaker for 3 hrs at 4°C. Ok4CL7 and 11 were purified using Glutathione Sepharose 4B (GE Healthcare, Chicago, USA), whereas Ok4CL15 with Ni-NTE slurry (Roche, Basel, Switzerland). For elution of Ok4CL7 and 11, reduced glutathione (20 µM; pH 8.0) was used, whereas imidazole (200 µM; pH 8.0) used for Ok4CL15. Size of purified proteins confirmed by separating on 12% SDS-PAGE gel. Salts were removed using dialysis against Milli Q water. Purified recombinant Ok4CL7 and OK4CL15 protein aliquots were stored at -80 °C after the snap freeze in the liquid nitrogen until further use. We could not successfully obtain recombinant protein of Ok4CL11.

2.2.6 Enzymatic assays of recombinant Ok4CL7 and -15

All enzyme assays were performed in three replicates. The activity of recombinant purified Ok4CL enzymes was studied with five substrates: *p*-coumaric acid, caffeic acid, cinnamic acid, ferulic acid and sinapic acid as described previously (Gui et al. 2011; Gao et al. 2015). The reaction mixture of 200 µl contains substrate (0.2 mM), adenosine triphosphate (2.5 M), coenzyme A (0.2 mM) and enzyme (5 µg). All assays were performed in Tris-HCl buffer (200 mM; pH 8.0) having MgCl₂ (25 mM). Blanks used for all assays contain heat-inactivated enzymes and all other components. The absorbance change was monitored with a UV-visible spectrophotometer (Labindia Pvt Ltd., Mumbai, India). Optimum pH and temperature required for recombinant Ok4CL7 and Ok4CL15 activity were determined using ferulic acid and sinapic acid, respectively. Range of pH 5 to pH 9 used for determination optimum pH, while for finding out optimum temperature range from 20°C to 80°C were used. The absorbance was measured at absorption maxima at 311, 333, 345, 346, and 352 nm for the corresponding cinnamoyl-CoA, 4-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, and sinapoyl-CoA products, respectively (Stöckigt and Zenk 1975; Schatz and Grisebach 1982). Substrate concentration used for enzyme kinetics ranged from 10 to 400 µM for coumaric acid and caffeic acid, whereas 25 to 800 µM for ferulic acid. All other components of the reaction were kept constant (ATP [2.5 mM]; CoA [0.2 mM], Enzyme [5 µg]). The average value of three replicates used to plot the graph. Standard deviation calculated using three replicate values and plotted as an error bar of respective value in the graph.

2.2.7 Characterization of recombinant Ok4CL products on HPLC and LC-MS/MS

Ok4CL assay products for substrates 4-coumaric acid, caffeic acid, ferulic acid and sinapic acid were analyzed on Waters HPLC using reversed-phase C18 column at 24°C with 0.7 mL/min flow rate. Photodiode array (PDA) detector used to analyze the change in absorption. 20 µl of sample injected from each reaction. Three runs of each enzyme with each substrate. i) standard substrate; ii) heat-inactivated enzyme and all other components; iii) reaction mixture with the active enzyme. Product peak was observed at absorbance 333 nm (*p*-coumaroyl CoA), 345 nm (caffeoyl CoA), 346 nm (feruloyl CoA) and 352 nm (sinapoyl CoA) (Knobloch and Hahlbrock 1977; An et al. 2015). Acetonitrile and ammonium acetate (0.1M; pH 4.5) was used as the mobile phase. Previous protocol with modifications was used for HPLC analysis (Costa et al. 2005). The gradient used ranges from 0 to 100% ACN. ACN reaches 65% in approximately 25 min and 100% in about 28 min. 100% ACN wash was given for 30 min. For equilibration of the system before the next run, 100% ammonium acetate (pH 4.5) for 8 min. The product peak was collected and dried using a lyophilizer. Purified metabolites were dissolved in 50% acetonitrile and 50% water and used for characterization by LC-MS/MS. Orbitrap Fusion mass spectrometer (Thermo Scientific) coupled with the heated electrospray Ion source was used for data acquisition. For MS1 mode, the mass resolution was kept at 120,000 and for MS2 acquisition, the mass resolution was 30,000. Mass range of data acquisition was 60–900 da. Extracted metabolites were separated on UPLC ultimate 3,000 installed with the Xbridge Amide column. Accurate mass and fragmentation pattern were acquired for the mixture by separating them on the HILIC column and positive and negative ionization mode both. The purified mixture was separated by solvent A was 20 mM ammonium acetate in the water of PH 9.0 and mobile phase B was 100% acetonitrile. The elution gradient starts from 85% B to 10% B over 14 min, with a flow rate of 0.35 ml/min. LC-MS/MS acquired data has been processed using the Xcalibur software (Thermo Scientific) using the default setting.

2.2.8 Statistical analysis

At least three replicates were used for all statistical analysis. Significance of *Ok4CL7* and *Ok4CL15* gene expression calculated using multiple unparallel t-tests in GraphPad Prism version 9.0.0 for macOS, GraphPad Software, CA, USA, www.graphpad.com. (ns= $P > 0.05$; *= $P \leq 0.05$; **= ≤ 0.01 ; ***= $P \leq 0.001$; **** = $P \leq 0.0001$). Significant differences between data were determined using the One-way Analysis of Variance (ANOVA: single factor) tests. Error bars represent the mean \pm standard deviation. One-way ANOVA test suggested a

significant difference between data $p < 0.005$ for both the enzymes substrate preference experiments.

2.3 Results and discussion

2.3.1 Ok4CLs diverged depending on their biological roles

Fifteen annotated 4CL sequences were retrieved from *O. kilimandscharicum* transcriptome datasets. Domain analysis depicts that most of Ok4CLs contain AMP-BD, CaiC, and 4CL domains, indicating their homology and possible functional conservation (**Table 2.2**). The open reading frame of these Ok4CLs had Box-I and Box-II signature motifs. The phylogenetic tree was constructed using 946 plant 4CLs protein sequences, including 15 putative Ok4CLs. These 4CLs were grouped into nine major clades (**Figure 2.3**).

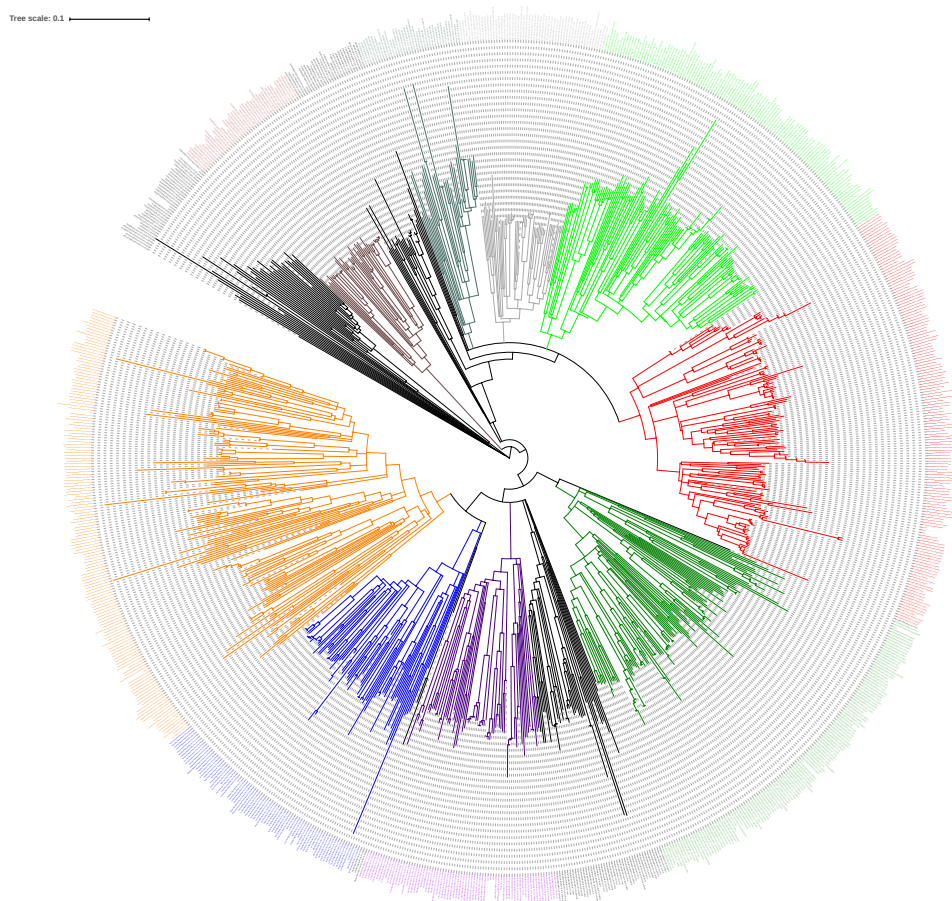


Figure 2.3 Phylogenetic analysis of plant 4CLs. The phylogenetic tree was constructed using 946 4CLs from the UniProt database and 15 4CLs from *O kilimandscharicum*. Plants 4CLs are distributed in 9 clades (highlighted in different colors).

Table 2.2 Classified features of Ok4CL enzyme (clade and domain arrangement) along with selected other plant 4CLs.

Name and Accession number	Length (aa)	Clade	Domain(s)
OK4CL1_ MW413287	555	Clade III	AMP-BD, CaiC, 4CL
Ok4CL2_ MW413288	477	Clade III	AMP-BD, CaiC, 4CL
OK4CL3_ MW413289	523	Clade III	AMP-BD, CaiC, FACL_fum10p_like
OK4CL4_ MW413290	484	Clade III	AMP-BD, CaiC, ttLC FACS AEE21 like
OK4CL5_ MW413291	614	Clade III	AMP-BD, CaiC, PLN02479, ttLC FACS AEE21 like
OK4CL6_ MW413292	380	Clade III	AMP-BD, CaiC
OK4CL7_ MW413293	576	Clade III	AMP-BD, CaiC, 4CL
OK4CL8_ MW413294	446	Clade I	AMP-BD, CaiC, 4CL
OK4CL9_ MW413295	425	Clade I	AMP-BD, CaiC
OK4CL10_ MW413296	454	Clade III	AMP-BD, CaiC
OK4CL11_ MW413297	576	Clade III	AMP-BD, CaiC, 4CL
Ok4CL12_ MW413298	424	Clade III	AMP-BD, CaiC, 4CL
OK4CL13_ MW413299	454	Clade III	AMP-BD, CaiC
OK4CL15_ MW413300	535	Clade I/II/III	AMP-BD, CaiC, 4CL
OK4CL16_ MW413301	412	Clade I	AMP-BD, CaiC
<i>N. tabacum</i> 4CL2_ A0A1S4AFV7	542	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>G. max</i> 4CL2_ Q8S5C1	547	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>P. tomentosa</i> 4CL_ Q941M4	536	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>V. vinifera</i> 4CL1_ A0A438CEN2	548	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>C. annuum</i> 4CL1_ A0A2G2YZH8	553	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>N. attenuata</i> 4CL2_ A0A314L3J8	558	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>A. thaliana</i> 4CL1_ A0A178W9C1	561	Clade I	AMP-BD, CaiC, 4CL, PLN02246, PLN02574
<i>A. sativum</i> 4CL_ G3CU71	545	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>Z. marina</i> 4CL1_ A0A0K9PSW3	552	Clade I	AMP-BD, CaiC, 4CL,
<i>Z. mays</i> 4CL1_ B4FQP4	555	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>P. patens</i> subsp. <i>patens</i> 4CL2_ B7SBA0	585	Clade II	AMP-BD, CaiC, 4CL, PLN02246

<i>P. taeda</i> 4CL_U5MY99	575	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>G. max</i> 4CL2_P31687	562	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>A. thaliana</i> 4CL3_Q9S777	561	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>V. vinifera</i> 4CL11_A0A438DX51	547	Clade III	AMP-BD, CaiC
<i>C. annuum</i> 4CL11_A0A1U8FQ18	554	Clade III	AMP-BD, CaiC, 4CL
<i>N. tabacum</i> 4CL11_A0A1S4B2L5	551	Clade III	AMP-BD, CaiC, 4CL
<i>A. thaliana</i> 4CL17_Q9M0X9	544	Clade III	AMP-BD, CaiC, 4CL
<i>G. max</i> 4CL_H2BER4	540	Clade III	AMP-BD, CaiC, 4CL
<i>A. annua</i> 4CL17_A0A2U1MBW6	535	Clade III	AMP-BD, CaiC, 4CL
<i>N. tabacum</i> 4CL1_A0A1S4A2D2	565	Clade III	AMP-BD, CaiC, 4CL
<i>C. cajan</i> 4CL1_A0A151SI61	608	Clade III	AMP-BD, CaiC, 4CL
<i>A. thaliana</i> 4CL1_Q84P24	566	Clade III	AMP-BD, CaiC, 4CL, PLN02574
<i>E. sagittatum</i> 4CL2_A0A0A7DMV6	550	Clade III	AMP-BD, CaiC, 4CL
<i>V. vinifera</i> 4CL1_A0A438K487	549	Clade III	AMP-BD, CaiC, 4CL
<i>Z. mays</i> 4CL14_A0A317YF40	551	Clade III	AMP-BD, CaiC, 4CL
<i>P. andersonii</i> 4CL_A0A2P5B929	556	Clade III	AMP-BD, CaiC, 4CL
<i>N. papyraceus</i> 4CL2_A0A346TLE9	560	Clade III	AMP-BD, CaiC, 4CL
<i>G. soja</i> 4CL19C_A0A445KB03	472	Clade III	AMP-BD, CaiC, 4CL

AMP-BD:AMP binding domain; CaiC:Acyl-CoA synthetase; 4CL:4 Coumarate CoA Ligase PLN02246:4-coumarate--CoA ligase; PLN02574:4-coumarate--CoA ligase- like; ttLC FACS AEE21 like: Fatty acyl-CoA synthetases similar to LC-FACS from *Thermus thermophilus* and *Arabidopsis*; FACL_fum10p_like: Subfamily of fatty acid CoA ligase (FACL) similar to Fum10p of *Gibberella moniliformis*.

Representative 4CLs selected from each clade and putative Ok4CLs were further used for phylogenetic tree construction (**Figure 2.4**). Selected 4CLs were grouped into three major clades belonging to known three classes (4CLs involved in lignin biosynthesis, non-structural phenolics biosynthesis and uncharacterized) as mentioned earlier. Ok4CL8, 9 and 16 grouped with other 4CLs in class-I, while all other Ok4CLs appeared in the uncharacterized clade (Class III) along with other 4CLs. Ok4CL3, 4 and 5 form a separate clade due to the fatty acyl CoA synthetase domain (**Figure 2.4**). Interestingly, Ok4CL15 is not clustered with any of the class. Comparatively, It is close to class I and class II than unclassified 4CLs.

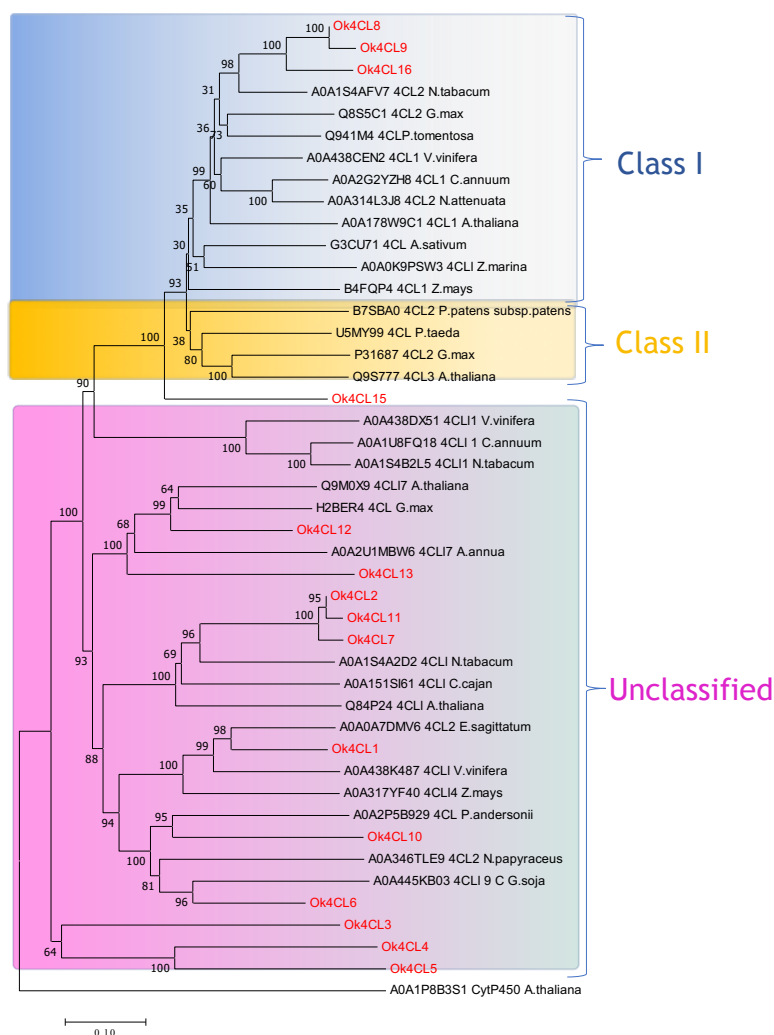


Figure 2.4 Phylogenetic analysis of 4CLs. Class-I 4CLs are involved in the biosynthesis of lignin and Class-II 4CLs are involved in the biosynthesis of non-structural phenylpropanoids.

A similar classification was reported for the 4CL isoforms in several plants (Ehltling et al. 1999; Lindermayr et al. 2002; Lavhale et al. 2018). *Piper nigrum* 4CLs distributed in class-I and class-II. Class-I 4CLs were active towards the lignin biosynthesis substrates like coumaric and ferulic. In addition to this, Class-II 4CLs are also active toward piperonylic acid, 3,4-(Methylenedioxy) cinnamic acid (MDCA), and piperic acid (Jin et al. 2020). 4CLs from monocot and dicot separated into two clades due to their parallel evolution after speciation (Sun et al. 2013). *Pinus taeda* 4CL (Pinta 4CL3, gymnosperm) isoform was grouped with Class-II angiosperm 4CLs, suggesting their evolutionary conservation (Chen et al. 2014). Some 4CLs contain conserved motifs as 4CLs and they are homologous with 4CL sequences

(Cukovic et al. 2001). These are also classified as 4CL-like and their functions are not yet clear (Raes et al. 2003). Further, similarity among the 4CLs was studied using the sequence similarity network (SSN). Our analysis showed three major clusters with an E value of $1e^{140}$ (Figure 2.5).

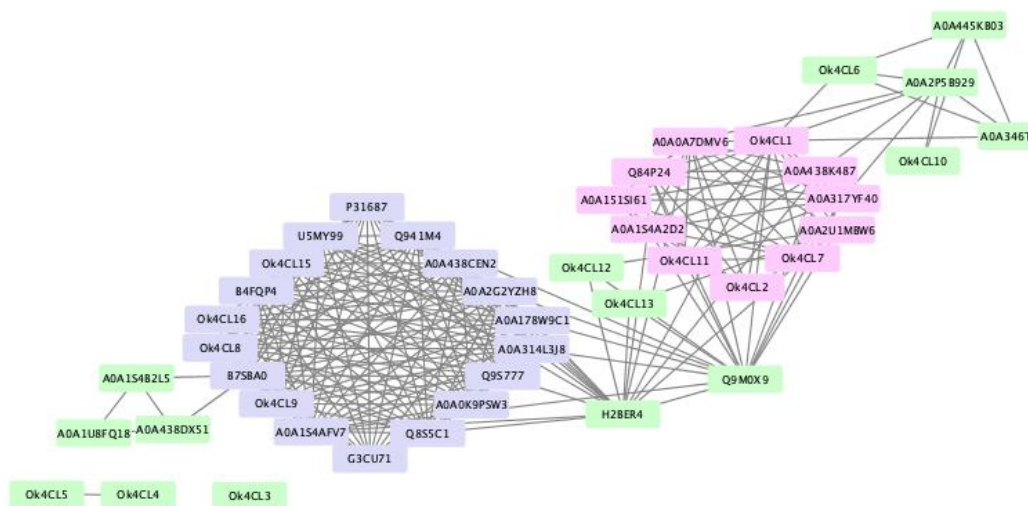


Figure 2.5 The sequence similarity network analysis of various plant 4CLs, including that from *O. kilimandscharicum* showed five clusters at the E-value of $1e^{140}$.

In the first cluster, Ok4CL7-15 were grouped with 4CL from *Epimedium sagittatum* (A0A0A7DMV6), *Prunus andersonii* (A0A2P5B929) and *Narcissus papyraceus* (A0A346TLE9). In the second cluster, Ok4CL2, -3, -8, -12, -13 and -5 were clustered with others 4CLs from *Nicotiana tabacum* (A0A1S4A2D2), (A0A1S4AFV7), *Zea mays* (A0A317YF40), *Vitis vinifera* (A0A338K487), *Arabidopsis thaliana* (Q9M0X9) (Q82P24), *Cajanus cajan* (A0A151SI61), *Glycine max* (H2BER4), *Artemisia annua* (A0A2U1MBW6). In the third cluster, Ok4CL1, -9, -10 and -16 were clustered with 4CLs from *Nicotiana attenuata* (A0A314L3J8), *G. max* (P31687), *Physcomitrella patens subsp. patens* (B7SBA0), *V. vinifera* (A0A438CEN2), *G. max* (Q855C1) *A. thaliana* (Q9S777), *C. annuum* (A0A1U8FQ18), *Allium sativum* (G3CU71G3CU71) *Zostera marina* (A0A0K9PSW3), *Populus tomentosa* (Q941M4), *A. thaliana* (A0A178W9C1), *Capsicum annuum* (A0A2G2YZH8) and *Pinus taeda* (U5MY99). Corroboration in phylogenetic and SSN clustering affirms the 4CLs clustering according to their evolutionary conservation.

The 4CL isoforms are characterized from several plants. These isoforms were grouped into two broad classes. Class-I is mainly involved in lignin biosynthesis and contains 4CLs

from dicotyledonous plants, such as At4CL1 and -2 (Lee and Douglas 1996; Vogt 2010). Many of the candidates from Class-II are possibly involved in the biosynthesis of non-structural phenolics, such as flavonoids and antitoxins and possess 4CLs from monocotyledonous, dicotyledonous and gymnosperm plants, e.g. At4CL3 (Lee and Douglas 1996; Hamberger and Hahlbrock 2004; Gui et al. 2011).

Selected Ok4CLs have conserved adenosine monophosphate binding sites and acyl CoA synthetase (CaiC) domain. 4CL domain was absent in Ok4CL3, -4, -5, -9 and -16. Out of these, Ok4CL3, -4 and -5 contain fatty acyl CoA synthetase (FACS) domain, proposing their potential role in fatty acid metabolism (Table 1). Ok4CL7, -8 and -15 have 4CL, CaiC and AMP binding domain (Figure 2.6).

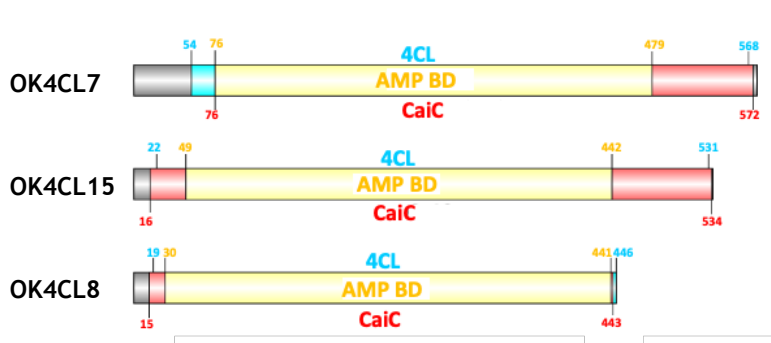


Figure 2.6 Domain analysis. Ok4CL7, -15 and -8 from *O. kilimandscharicum* have several domains, including AMP BD- AMP binding domain; CaiC (Acyl CoA Synthetase); 4CL (4-Coumarate-CoA Ligase).

2.3.2 Ok4CL7 and -15 exhibits differential expression in various tissues

Comparative analysis of gene expression values indicates differential and tissue-specific expression of 4CLs in *O. kilimandscharicum* (Figure 2.7A). Most of the 4CLs were expressed in the trichome, while the least expression was observed in floral organs. *Ok4CL11* was highly expressed in young and mature leaves, trichomes, androecium and petal. Whereas, *Ok4CL8* had high expression in stem tissue. *Ok4CL5* and -15 were abundant in inflorescence and root tissues (Figure 2.7A). Ok4CL7 is highly expressed in trichome, young leaf, mature leaf and root, while *Ok4CL15* shows expression in most of the *O. kilimandscharicum* tissues, with the highest levels in the root (Figure 2.7B).

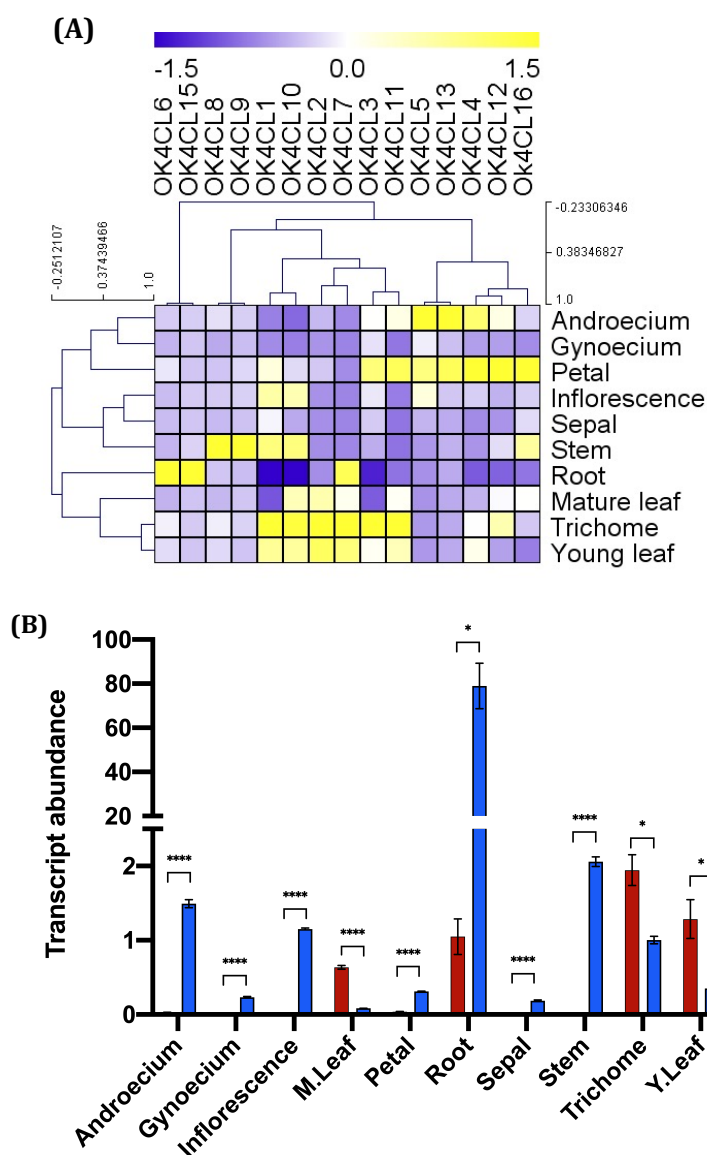


Figure 2.7 *Ok4CL* gene expression analysis in different tissues of *O. kilimandscharicum*. **A.** Gene expression analysis of putative fifteen *Ok4CLs* in ten different tissues of *O. kilimandscharicum*, such as androecium, gynoecium, inflorescence, mature leaf, petal, root, sepal, stem, trichome, and young leaf. **B.** *Ok4CL7* and *-15* expression analyses in all ten tissue types of *O. kilimandscharicum*.

The expression of 4CL isoforms studied in several plants (Lee and Douglas 1996; Di et al. 2012; Lavhale et al. 2018). In some plants, 4CL isoforms are expressed across all the tissues, while others have tissue-specific expressions, e.g. *Hibiscus cannabiuns*, *Isatis indigotica*, and *Rubus idaeus* (Kumar and Ellis 2003; Di et al. 2012; Emran et al. 2013).

Expression of 4CL isoforms also varies during plant development (Costa et al. 2005; Emran et al. 2013). In *H. cannabiuns*, 4CL expression in stem increases from 2 to 20 weeks of development (Emran et al. 2013). Phenylpropanoid pathway products protect plants from wounding, irradiation with UV light or pathogen attack (Lee and Douglas 1996). In *A. thaliana*, the expression of *At4CL1* and -2 were upregulated by wounding and the treatment with methyl jasmonate (Lee and Douglas 1996). *At4CL1* and -2 expression were upregulated by *Peronospora parasitica* infection and wounding (Ehlting et al. 1999). Tissue-specific expression of *Ok4CLs* might indicate their engagement in the specialized function through variable substrate specificities.

2.3.3 Recombinant Ok4CL7 and -15 proteins differ in biochemical properties and substrate specificity

Recombinant Ok4CL7 and -15 were purified separated on 12% SDS-PAGE (**Figure 2.8A and B**). The activity of recombinant Ok4CL7 and -15 proteins was studied with cinnamic acid, *p*-coumaric acid, caffeic acid, ferulic acid and sinapic acid. A single point assay revealed that Ok4CL7 and -15 preferred ferulic acid and sinapic acid, respectively. Moreover, optimum pH and temperature for Ok4CL7 were determined using ferulic acid as a substrate, while sinapic acid was used as a substrate for Ok4CL15. The observed optimum pH for Ok4CL7 and -15 activity were pH 8 and 7, respectively (**Figure 2.9A**). Both Ok4CL7 and -15 show maximum activity at 40°C (**Figure 2.9B**). Out of the five substrates, Ok4CL7 utilized *p*-coumaric acid, caffeic acid and ferulic acid (**Figure 2.10A**). Ok4CL15 showed activity with *p*-coumaric acid, ferulic acid and sinapic acid (**Figure 2.10B**). Both recombinant 4CLs proteins utilized *p*-coumaric acid and ferulic acid and inactive against cinnamic acid.

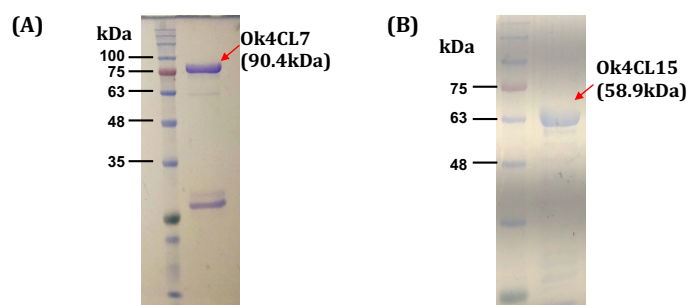


Figure 2.8 Recombinantly expressed (A) Ok4CL7 and (B) Ok4CL15 showed the predominant band on 12% SDS PAGE. Ok4CL7 and -15 proteins were observed at expected size 90.4 kDa and 58.9 kDa, respectively.

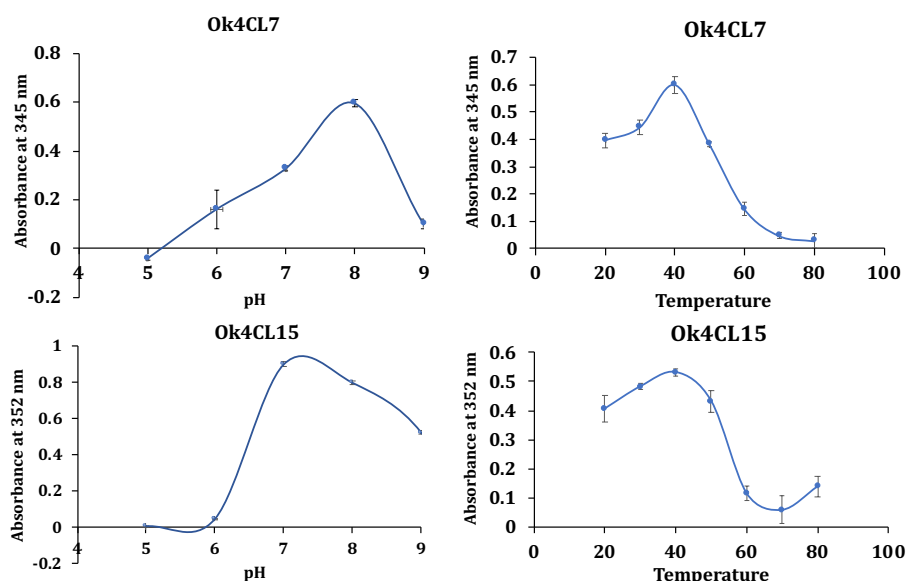


Figure 2.9 Determination of optimal conditions for 4CL7 and -15 activities. **B.** Determination of optimum pH for Ok4CL7 and -15. **C.** Determination of optimal temperature for Ok4CL7 and -15.

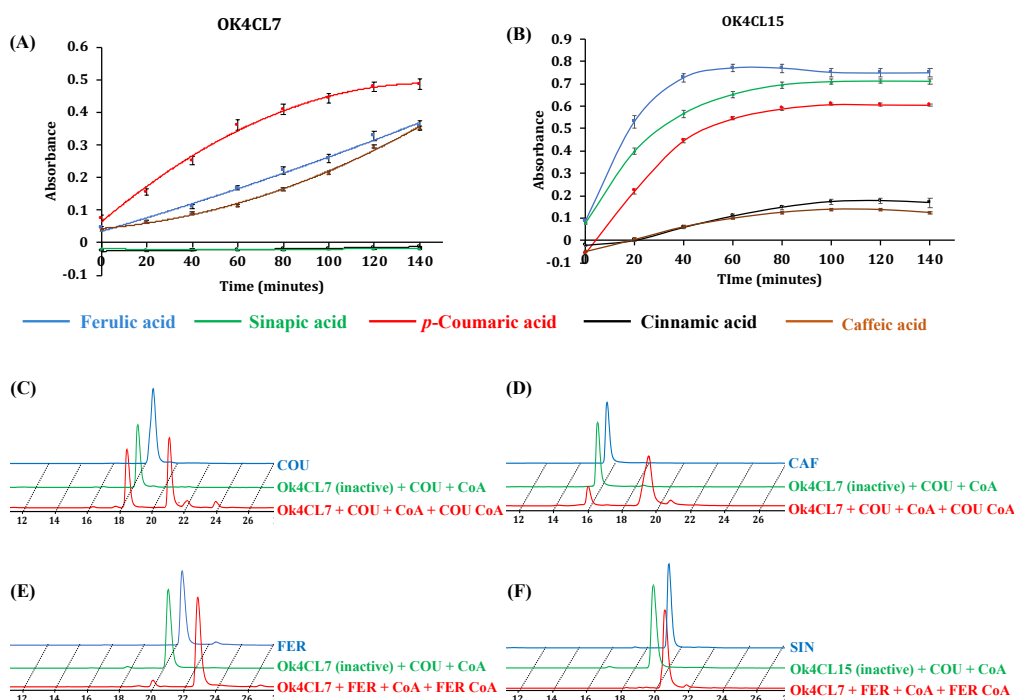


Figure 2.10 The activity of 4CL7 and -15 with different substrates. **A.** Ok4CL7 showed activity with *p*-coumaric acid (*p*-coumaroyl CoA: 333nm), ferulic acid (feruloyl CoA: 345 nm), and caffeic acid (caffeoyl CoA: 345 nm). **B.** Ok4CL15 showed activity with *p*-coumaric acid, ferulic acid, and sinapic acid (sinapoyl CoA: 352nm). **Analysis of enzyme assay reaction on HPLC.** Three runs of each enzyme with each substrate. i) standard

substrate; ii) heat-inactivated enzyme and all other components; iii) reaction mixture with the active enzyme. **C.** Enzyme activity with *p*-coumaric acid (COU), *p*-coumaroyl CoA product peak observed at RT 21.05. **D.** Enzyme activity with caffeic acid (CAF), caffeoyl CoA product peak observed at RT 19.02. **E.** Enzyme activity with ferulic acid (FER), feruloyl CoA product peak observed at RT 21.38. **F.** Enzyme activity with sinapic acid (SIN), sinapoyl CoA product peak observed at RT 20.7.

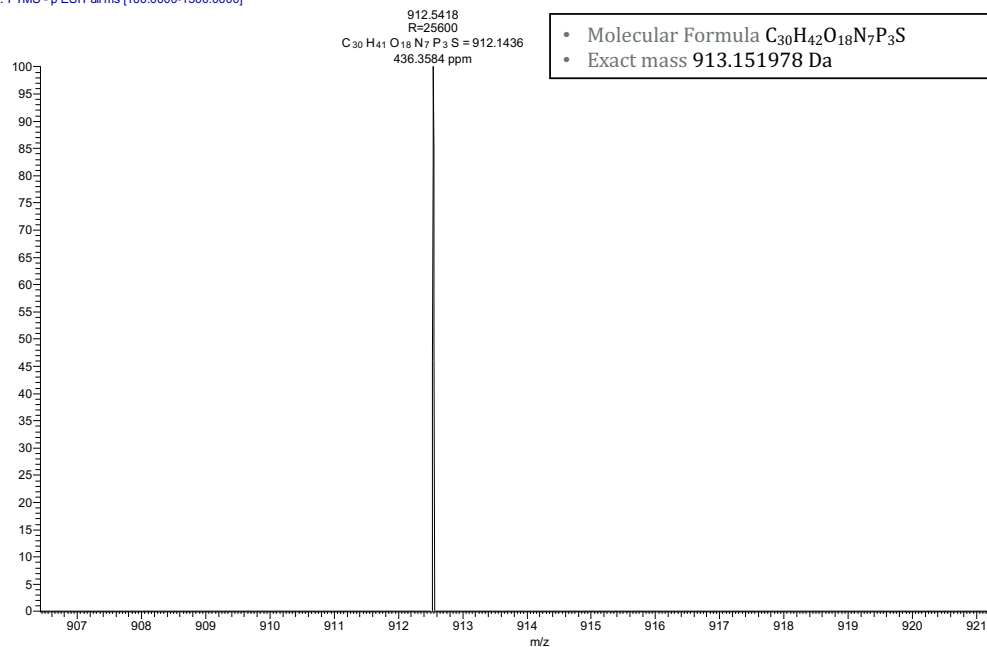
Ok4CL7 preferred substrates were caffeic acid, followed by *p*-coumaric acid and ferulic acid. Ok4CL15 has more preference for sinapic acid than *p*-coumaric acid and ferulic acid. The products formed were characterized and validated using HPLC (**Figure 2.10C, D, E, F**). The peaks of *p*-coumaroyl CoA, feruloyl CoA, caffeoyl CoA and sinapoyl CoA were observed at retention time 21.05, 19.02, 21.38 and 20.07 mins, respectively. The product was purified using HPLC and further confirmed by mass spectrometry using accurate mass and fragmentation pattern match (**Figure 2.11**). Different peaks of *p*-coumaroyl CoA (912.5418 Da for at 1.73 retention time) and feruloyl CoA (942.1552 Da at 3.89 retention time) were observed (**Figure 2.11**).

Substrate specificities among the 4CL isoforms define their role in plant growth, development and defense (Li et al. 2015; Lavhale et al. 2018). Sinapic acid converting 4CLs are very rare in angiosperms (Hamada et al. 2004). Alternative pathways for sinapoyl alcohol formation is reported in some plants, including *A. thaliana* (Li et al. 2000; Costa et al. 2005). In *A. thaliana*, syringyl lignins and sinapate/sinapyl alcohol are derived from methylation of 5-hydroxy ferulic acid instead of sinapic acid or sinapoyl CoA (Costa et al. 2005). Sinapic acid converting isoforms of 4CL4 are also characterized from *A. thaliana* and *Robinia pseudoacacia* (Hamberger and Hahlbrock 2004; Hamada et al. 2004). The isoform Rp4CL1 utilizes *p*-coumarate as a substrate, preferably over ferulate and sinapate. Rp4CL2 and Rp4CL3 can activate sinapate along with caffeate and *p*-coumarate. Crude protein extracts from the shoot and developing xylem showed similar substrate preferences as Rp4CL2 and Rp4CL3. This suggests that sinapate-activating isoforms might be involved in lignin biosynthesis (Hamada et al. 2004).

(A)

Coumaroyl CoA

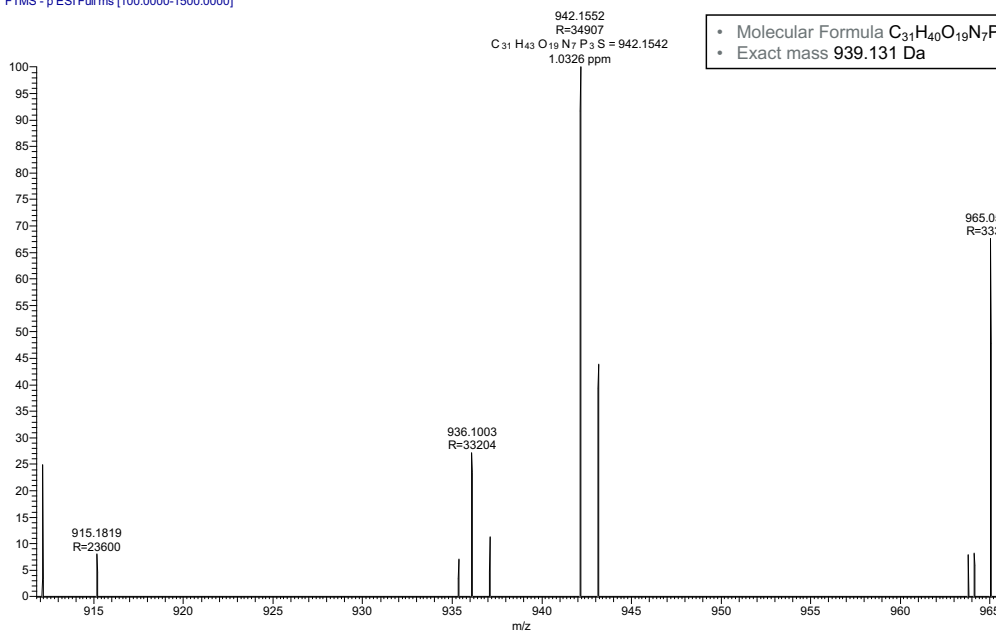
Coumaroyl_COA-neg #295 RT: 1.73 AV: 1 NL: 5.81E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]



(B)

Feruloyl CoA

Feruloyl_COA-neg #652 RT: 3.89 AV: 1 NL: 3.39E4
T: FTMS - p ESI Full ms [100.0000-1500.0000]



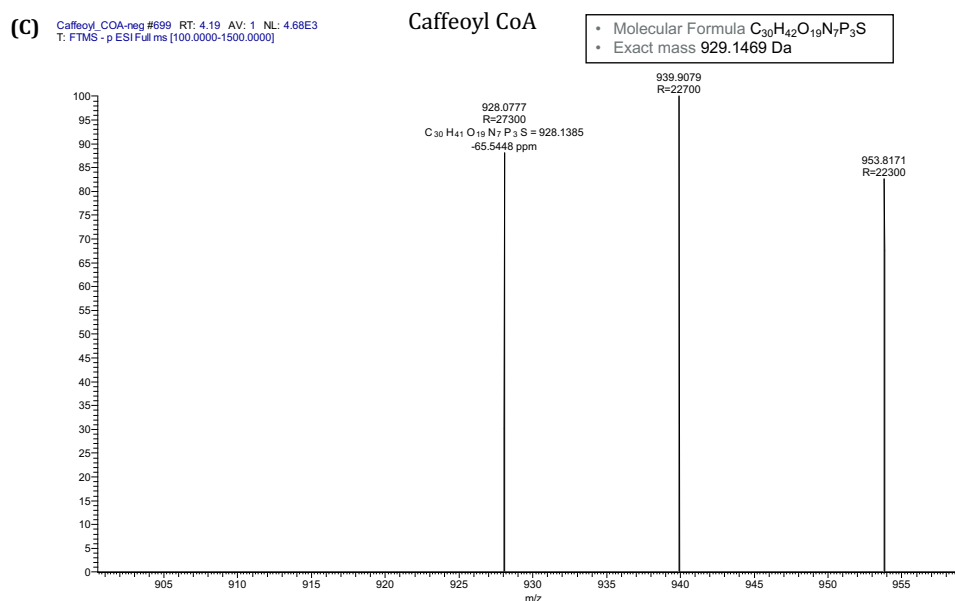


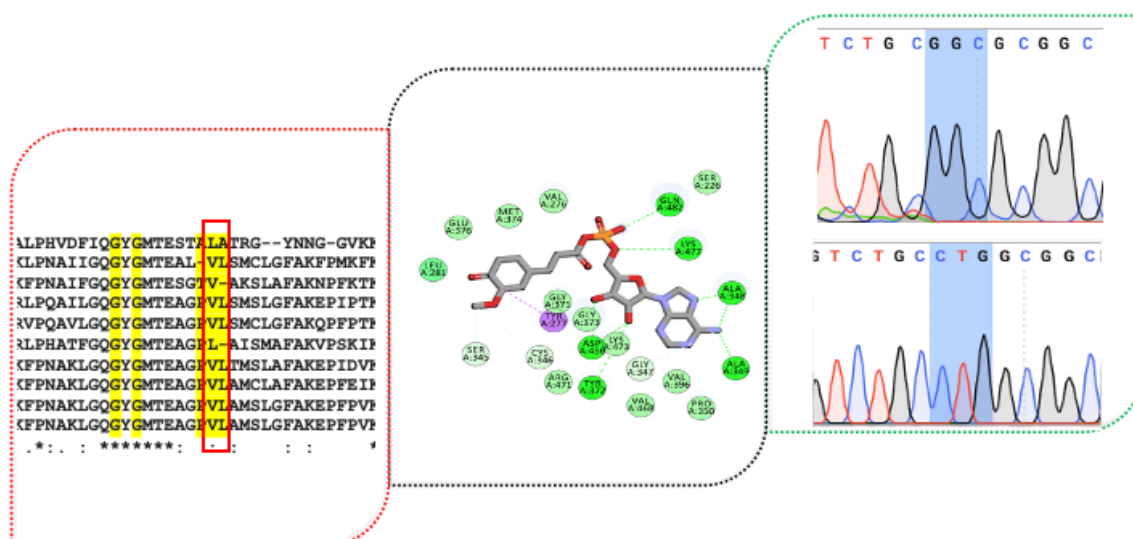
Figure 2.11 Analysis of 4CL enzyme assay products on LC-MS. A: Coumaroyl CoA; B: Feruloyl CoA; C: Caffeoyl CoA. Different peaks of *p*-coumaroyl CoA (912.5418 Da for at 1.73 retention time) and feruloyl CoA (942.1552 Da at 3.89 retention time) were observed.

2.4 Conclusions

Two isoforms of 4CL were functionally characterized from *O. kilimandscharicum* for the first time. While we could not successfully obtain recombinant protein of Ok4CL11. Gene expressions of *Ok4CL7* and *-15* were high in trichomes and root tissues compared to other tissue types tested. Optimum pH required for *Ok4CL7* and *Ok4CL15* activities was found to be 8 and 7, respectively. Both 4CLs showed maximum activities at 40°C. *Ok4CL7* catalyzed the conversion of *p*-coumaric acid, ferulic acid and caffeic acid into their corresponding CoA esters. Furthermore, *Ok4CL15* could catalyze the conversion of *p*-coumaric acid, ferulic acid and sinapic acid into respective CoA esters. However, both 4CLs were unable to utilize cinnamic acid. Specific non-covalent interactions between the substrate and binding pockets of Ok4CLs are essential for their catalytic activities. It appears that *Ok4CL15* could be involved in lignin biosynthesis due to its high expression in root tissues and catalytic activity to convert sinapic acid to sinapoyl CoA. The results of this study allow us to understand the function of 4CL isoforms of *O. kilimandscharicum* and its potential role in the diversification of the phenylpropanoid pathway.

Chapter 3

Molecular dissection of substrate specificity mechanism in Ok4CLs



3.1 Introduction

4CL belongs to the family of adenylate-forming enzymes. They are involved in plant specialized metabolite biosynthesis, protein biosynthesis, posttranslational modification, nucleotide, amino acid and fatty acid metabolism (Lavhale et al. 2018). They catalyze the substrate activation by ligating Coenzyme A in two steps, adenylate formation and thio-esterification (BERG 1956).

In vitro activity of Ok4CL7 and -15 shows differential substrate selection and activity. OK4CL7 preferably utilizes *p*-coumaric acid, ferulic acid and caffeic acid, while Ok4CL15 utilizes *p*-coumaric acid, ferulic acid and sinapic acid (Lavhale et al. 2021). Differential substrate selection by Ok4CLs may contribute to regulation flux in the phenylpropanoid pathway. 4CL is known for diverting flux in the phenylpropanoid biosynthesis pathway. 4CLs are broadly classified into two classes- class I and class II. Class I is involved in lignin biosynthesis, whereas class II is involved in the biosynthesis of nonstructural phenolics like flavonoids, coumarins, etc.

Molecular modeling and docking, followed by *in silico* analysis of enzyme-substrates interaction provided information about residues contacting the substrate in the binding pocket of enzyme. It helps in the identification of residues from the active site and binding pocket contributing to substrate preferences, also sheds light on structural features affecting enzyme activity (cavity size, orientation, etc.). Furthermore, docking analysis helps to predict binding affinity by calculating theoretical binding energy on enzyme-substrate interaction. This information could be used to predict the substrate preferences of enzymes. Further, the outcome from the *in silico* could be experimentally validated by performing assays.

In the current chapter, we have studied the interaction of Ok4CL7 and -15 with substrates to identify residues involved in the interaction. Based on the interaction study of the preferred and non-preferred substrate, mutant models were developed to validate the indispensability of selected residues in substrate preference. Based on outcomes from *in silico* experiments, 4CL mutants were prepared by site-directed mutagenesis and used to check their interaction by *in vitro* enzyme assay. The outcome of this study will generate information on active site residues in 4CL that contributes to substrate preferences.

3.2 Materials and methods

3.2.1 Comparison of 4CL protein sequences from *Ocimum kilimandscharicum* and reported 4CLs from other plants

Characterized plant 4CL sequences were selected for comparison of amino acid residues. Ok4CLs were compared with selected sequences, including plant 4CLs from *Arabidopsis thaliana*, *Nicotiana tabacum* and *Glycine max*. Uniport accession numbers of selected 4CLs- At4CL1(Q42524), 4CL2 (Q9S725), At4CL3(Q9S777), At4CL4 (Q 84P23), Gm4CL1 (Q8S564), Gm4CL2 (P31687), Gm4CL3 (Q8S5C2), and Nt4CL2 (O24146). Sequences were aligned using the multiple sequence alignment tool (<https://www.ebi.ac.uk/Tools/msa/>) (Madeira et al. 2019). Clustal Omega tool used for alignment and all other parameters set to default. Output format was set to ClustalW with character counts.

3.2.2 Molecular docking analysis for determination of residues interacting with the substrate in the active site

Interaction and substrate preferences of recombinant Ok4CL7 and -15 with selected substrates were studied using molecular docking. Three-dimensional structures of Ok4CL7 and -15 were predicted using the SWISS-MODEL server (Waterhouse et al., 2018). *Populus tomentosa* 4CL (PDB ID: 3NI2) and *Nicotiana tabacum* 4CL (PDB ID: 5BST) crystal structures as used as templates for prediction of Ok4CL7 and -15 structures, respectively. Structures were energy minimized using Maestro 10.1 Tools (Maestro, Schrödinger, LLC, New York, NY, 2020) and substrate binding pocket residues were predicted by superimposing models with template structure. Ok4CL structures were then prepared for docking by adding Kollman and Gasteiger charges using AutoDock Tools (Morris et al., 2009). After adding polar hydrogens to the protein structure, the docking grid was set around substrate-binding residues using the AutoGrid Tool. The prepared structure was saved in *.pdbqt format for further analysis. Structures of the feruloyl adenylate and cinnamoyl adenylate were generated in Marvin sketch software (<http://www.chemaxon.com>) and converted in *.pdbqt format using AutoDock Tool. Docking of substrates with Ok4CLs was performed using AutoDock Vina (Trott and Olson 2009). After docking simulations, ten docking poses were generated with each substrate. The binding pose with the lowest binding score was selected for further analysis. Substrate binding poses were analyzed using the BIOVIA Discovery Studio 4.5 software (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016).

3.2.3 Development of 4CL mutant models to alter substrate specificity

The enzyme's affinity or substrate preference depends on binding pocket residues of the enzymes. Preference for a particular substrate could be influenced by variation in binding site residues. Binding pocket residues of Ok4CL7 and -15 were identified by molecular docking analysis with substrates. These residues were subsequently mutated to confirm their functionality. Based on enzyme-substrate interaction analysis as well as previous reports, 27 amino acids of Ok4CL7 were selected for further study- S226, V276, Y277, S280, L281, H317, S345, C346, G347, A348, A349, P350, G371, Y372, G373, M374, E376, A379, L380, A381, V396, D456, V468, R471, K473, K477 and Q482. Similarly, 29 amino acids interacting with the substrate in the active site of Ok4CL15 were selected for further study- S189, H237, V238, Y239, V242, S243, P279, V306, C307, G308, A309, A310, P311, Q331, G332, Y333, G334, M335, T336, L339, V340, L341, M343, C359, T417, D419, I431, R434 and K525.

The DUET tool developed mutant models by replacing residue at selected positions with all other standard amino acids. Models were selected for molecular docking analysis based on predicted stability change ($\Delta\Delta G$). Molecular docking analysis of mutant models performed using AutoDock Vina with major substrates related to plant metabolism. Variants showed significant binding energy variation with a substrate selected for the experimental validation. *Ok4CL* mutants with specific position changes were generated using the QuikChange lightning site-directed mutagenesis kit. The standard protocol provided by the manufacturer was used for the mutant development. Mutation at the desired position was confirmed by DNA sequencing. Expression of the mutant proteins and functional characterization work is in progress.

CoA ligase enzymes belong to the family of adenylate-forming enzymes. They are involved in several metabolic pathways, including phenylpropanoid biosynthesis (Schmelz and Naismith 2009; Orlova et al. 2012). 4CL substrates from several metabolic pathways such as phenylpropanoid, ubiquinone, terpenoid, fatty acid, etc., biosynthesis pathways were used to study interaction with Ok4CL mutants (**Table 3.1**). Structures of the 4CL substrates were generated in Marvin sketch software (<http://www.chemaxon.com>) and further converted in *.pdbqt format using AutoDock Tool. Docking of substrates with Ok4CLs was performed using AutoDock Vina.

Table 3.1 List of the substrates used for interaction study with Ok4CL mutants and their respective metabolic pathways.

Substrate	Pathway
4-Coumaric acid, coumaroyl adenylate, Ferulic acid, Caffeic acid, Sinapic acid, sinapoyl adenylate, Cinnamic acid and cinnamoyl adenylate	Ubiquinone and other terpenoid-quinone biosynthesis Phenylpropanoid biosynthesis Biosynthesis of plant specialized metabolites
Oleic acid and Succinate	Fatty acid biosynthesis Citrate cycle (TCA cycle) Propanoate metabolism Microbial metabolism in diverse environments
<u>2-succinylbenzoate</u> Acetate	Ubiquinone and other terpenoid-quinone biosynthesis Glycolysis / Gluconeogenesis Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Propanoate metabolism Methane metabolism
4-Hydroxybutyrate	Butanoate metabolism Carbon metabolism
Biotin	Biotin metabolism
Pimelic acid	Biotin metabolism
Oxalate	Glyoxylate and dicarboxylate metabolism
Anthranilic acid	Phenazine biosynthesis Aminobenzoate degradation Biosynthesis of specialized metabolite Microbial metabolism in diverse environments

3.2.4 Cloning, expression and purification of mutant Ok4CL proteins

Based on binding scores of Ok4CL mutants with substrates as well as previous reports, amino acid residues were selected for replacement or deletion in Ok4CL7 (K477I, K477V and L380) and Ok4CL15 (G308L and K525I). Mutations were incorporated in *Ok4CL7* and *-15* genes cloned in the expression vector (pET28a and pGEX4T) using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent, USA). Primers were designed as per the manual to change nucleotide at the appropriate position in the DNA of *Ok4CL* genes (**Table 3.2**). Primers 25 to 45 bases in length, with a melting temperature (T_m) of ≥ 75 °C, were designed. Desired mutation point is kept at the middle of the primer. Plasmids having the Ok4CL gene were isolated from the *E. coli* top 10 cells. These cells are *dam*⁺ have the ability to methylate DNA at a specific site. Methylated plasmids are used as a template to synthesize mutant strands (plasmid). Mutant strand synthesis reaction mix contains 2.5 μ l QuikChange Lightning Multi reaction buffer (10 \times) dsDNA template 1 μ l (100 ng), 2 μ l of forward and reverse mutagenic

primers (100ng), dNTP 1µl, 1µl of QuikChange Lightning Multienzyme blend nuclease and free water added to make 25 µl final volume. A mutant strand is synthesized on the thermal cycle using this reaction mix. The cycle sequence consists of 1 cycle at 95 °C for 1 minute, 30 cycles at 60 °C for 4 minutes, and the final step 65 °C for 1 minute. *Dpn* I (1µl) restriction enzyme was added to each amplification reaction and kept at 37 °C for 1 minute. *Dpn* I cleaves explicitly DNA containing methylated adenine. Template DNA plasmid is degraded at this step. PCR products (~5µl) transformed into XL10-Gold ultracompetent cells (provided along with kit) using the heat-shock method. Transformed cells spread on LBA medium containing ampicillin x-gal and IPTG. After overnight incubation at 37 °C, white colonies inoculated in LB containing ampicillin (100µg/mL). The next day, the plasmid is isolated and mutation at desired position is confirmed by sequencing (**Figure 3.1**). Mutant plasmids transformed into Arctic DE3 competent cells for protein expression. Protein expression was induced in a secondary culture growing at log phase by adding 0.5 mL of 1M IPTG. Cultures were incubated at 12 °C overnight. Protein was purified from overnight grown cultures by affinity purification method.

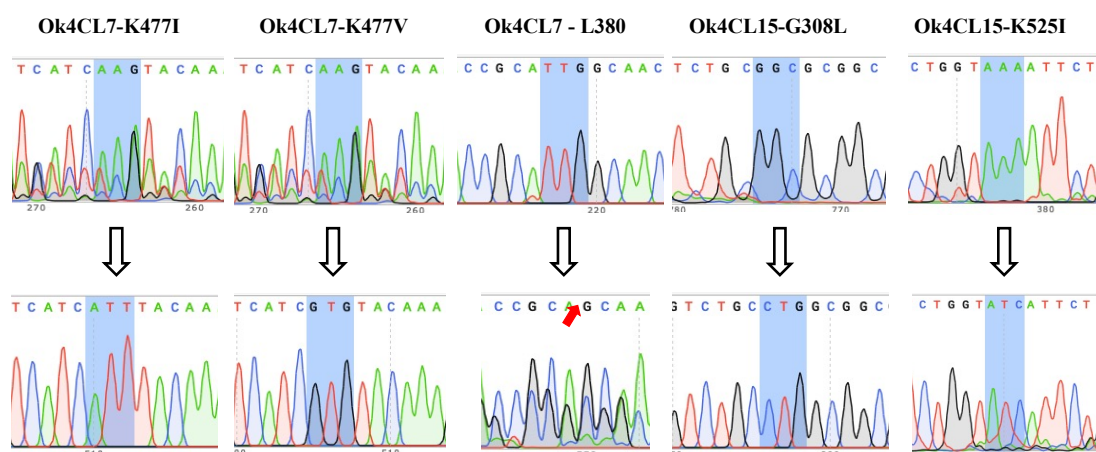


Figure 3.1 Analysis of DNA sequence for the desired mutation in *OK4CL7* and *-15* gene.

3.3. Results and discussion

3.3.1 Adenylate forming domain and conserved residues present in Ok4CLs

4CLs are functionally characterized from several model plants, such as *Arabidopsis thaliana*, *Glycine max* and *Nicotiana tabacum*. Active site residues of the 4CL enzymes are characterized by various tools like modeling, molecular docking, structure and crystallography. Schneider et al. (2003) studied At4CL2 structure by crystallography and identified 12 amino acids

responsible for the substrate specificity. These residues are present in the substrate-binding pocket (SBP) of At4CL2 (Schneider et al. 2003). Alignment of reported 4CLs and from Ok4CLs shows that adenylate forming domain (box I), box II and most of the conserved residues are present in both Ok4CL7 and -15 (**Figure 3.2**). According to modeling and site-directed mutagenesis study, valine is essential for activity. Deletion of a single amino acid valine (corresponding to V355 in At4CL2) from soybean Gm4CL2 and -3 leads to gain of sinapic acid converting activity (Lindermayr et al. 2003). Gm4CL1 (V is absent) (Lindermayr et al. 2003) and At4CL4 (L is absent) (Hamberger and Hahlbrock 1998) are naturally able to utilize sinapic acid. Naturally, Nt4CL2 does not utilize sinapic acid as a substrate (Li and Nair 2015). They developed sinapic acid using mutant Nt4CL2 by deleting V at the 341 position (Li and Nair 2015). Interestingly, V is present at the same place in Ok4CL15 and it can utilize sinapic acid. At the same time, V is absent in Ok4CL7, but it cannot use sinapic acid (**Figure 3.2**). These results suggest that some other residues are essential, along with the position of valine and leucine for the utilization of sinapic acid as a substrate. Amino acids highlighted in red interact with the substrate in the active site; residues shown in the red box are probably responsible for sinapic acid specificity (**Figure 3.2**).


```

Ok4CL7      SSGTTGKSKGVMLTHGNFIAMIELFVRFEASLYDYPSTANVYLAVVPMFHYGLSLFVLG
Ok4CL15    SSGTTGLPKGVMLSHRNIITCISQQVDGENPA-NHIDCEDRLLCVLPPLFHYVSMVSMLC
At4CL4     SSGTTGLPKGVMLTHKGLVTSIAQQVDGENPN-LNFTANDVILCFLPMFHYVALDALMLS
At4CL3     SSGTTGLPKGVMLTHKSLITSVAQQVDGDNPN-LYKSNVDVILCVLPPLFHYVSLNSVLLN
Gm4CL3     SSGTTGLPKGVMLTHKSLTTSVAQQVDGENPN-LYLTEDVLLCVLPPLFHYVSLNSVLLC
Gm4CL1     SSGTSLPKGVMLSHKNLVTITIAQLVDGENPH-QYTHSEVLLCVLPPLFHYVALNSILLC
Gm4CL2     SSGTTGLPKGVMLSHKGLVTSIAQQVDGDNPN-LYYHCHDTILCVLPPLFHYVSLNSVLLC
Nt4CL2     SSGTTGLPKGVMLTHKGLVTSVAQQVDGENPN-LYIHSVDVILCVLPPLFHYVSLNSVLLC
At4CL1     SSGTTGLPKGVMLTHKGLVTSVAQQVDGENPN-LYFHSDDVILCVLPPLFHYVALNSIMLC
At4CL2     SSGTTGLPKGVMLTHKGLVTSVAQQVDGENPN-LYFNRDDVILCVWPMFHYVALNSIMLC
          ****:*  ***::*  .: : :  *  :  :  *..*::*:..:  ::*

Ok4CL7      LLSLGSVVVTKRFDGDEMVRADRYGITHLHAVPPILIALTKRAKNADGNFGQSLKQVS
Ok4CL15    SLRGAALIVIMPRFELNELMEVIQKYKVTIAEFVPPILLGIKASQTA-AKFDLSSVRRVV
At4CL4     AMRTGAALIVPRFELNLMELIQRYKVTIVFVAPPVVLAFIKSPET-ERYDLSSVRIML
At4CL3     SLRSGATVLLMHKFEIGALLDLIQRHRVTIAALVPPVLIALAKNPTV-NSYDLSSVRFVL
Gm4CL3     ALRAGSAVLLMQKFEIGTLELIIQRHRVSVAMVPPVLVLAALAKNPMV-ADFDLSSIRLVL
Gm4CL1     GIRSGAAILIQKFEITLLELIEKHYKVTIVASFVPPVLAIVKSGET-HRYDLSSIRAVV
Gm4CL2     GLRAKATILLMPKFDINSLALIHKKVVTIAEFVPPVLAISKSPDL-HKYDLSSIRVLK
Nt4CL2     GLRVGAAILIMQKFDIVSFLELIQRKYKVTIGFVPPVLAIAKSPMV-DDYDLSSVRTVM
At4CL1     GLRVGAAILIMPKFEINLLELIIQRCKVTIVAMPVPPVLAIAKSSET-EKYDLSSIRVVK
At4CL2     SLRIGATILLIMPKFEITLLELIIQRCKVTIVAMPVPPVLAIAKSSET-EKYDLSSVRMVK
          :  : : : : : * :  :  * :  :  :  * :  : :  * :  : :  * :  : :

Ok4CL7      CGAAPLSEKSIVELIEALPHVDFIQGYGMTESTLAATRG--YNNG-GVKKYSVGLLSPN
Ok4CL15    CGAAPMDRKLLELSLKSPLNPAIIGQGYGMTEALVLSMCLGFAKFPKFKAGSGCENVIKN
At4CL4     SGAATLKKELEDAVRLKFPNAIFGQGYGMTESGTV-AKSLFAKPNFKTKSGACGTVVRN
At4CL3     SGAAPLGKELQDSLRRRLPQAILGQGYGMTEAGVLSMSLGFAPKEPIPTKSGCGCTVVRN
Gm4CL3     SGAAPLGKELVEALRNRVPAVLGQGYGMTEAGVLSMCLGFAKQPPFTKSGCGCTVVRN
Gm4CL1     TGAAPLGGELEAEVKARLPHATFGQGYGMTEAGPLAISMAFAKVPKIKPGACGTVVRN
Gm4CL2     SGGAPLGKELEDTLRAKFPNAKLGQGYGMTEAGVLTMSLAFAPKEPIDVKPGACGTVVRN
Nt4CL2     SGAAPLGKELEDTVRAKFPNAKLGQGYGMTEAGVLAACLFAFAKEPFEIKSGACGTVVRN
At4CL1     SGAAPLGKELEDAVNAKFPNAKLGQGYGMTEAGVLAAMSLGFAKEPFPVKSGACGTVVRN
At4CL2     SGAAPLGKELEDAISAKFPNAKLGQGYGMTEAGVLAAMSLGFAKEPFPVKSGACGTVVRN
          *.* :  :  :  .*..:  *****:  :  :  :  * :  : :  *

Ok4CL7      IEAKVVDVVTGALVPPGSI GELNLRIT
Ok4CL15    ARMKIIDPATGASLDRNQR GEICLKG
At4CL4     AEMKVVDVVTGALVPPGSI GELNLRIT
At4CL3     AELKVVHLETRELSLGYNQP GEICIRG
Gm4CL3     AELRVVDVVTGALVPPGSI GELNLRIT
Gm4CL1     AEMKIVDVTETGDSLPRNKH GEICIRG
Gm4CL2     AEMKIVDPETGHS LPRNQS GEICIRG
Nt4CL2     AEMKIVDPKTNLSLPRNQS GEICIRG
At4CL1     AEMKIVDPTGDSLPRNQP GEICIRG
At4CL2     AEMKILDPDTGDSLPRNKP GEICIRG
          . : : . *  :  . . ** :

```

Figure 3.2 Alignment of Ok4CL7 and -15 amino acid sequences from *O. kilimandscharicum* reported 4CL from *A. thaliana*, *G. max* and *N. tabacum*. Twelve amino acids highlighted in yellow are reported to interact with substrate in the active site of 4CLs. The conserved box I and II are highlighted in cyan color.

3.3.2 Differential interactions of Ok4CL7 and -15 with substrates contributes to variation in substrate preferences

Interaction of Ok4CL7 and -15 docked complexes illustrated the functional groups of substrates and their intermediates form several non-covalent interactions with binding site residues. Ok4CL7 has a binding score of -6.3 and -5.4 kcal/mol with feruloyl and cinnamoyl adenylate, respectively. While, Ok4CL15 has a binding score of -6.7 and -5.8 kcal/mol with feruloyl and cinnamoyl adenylate, respectively. Binding energy score pattern corroborating with the *in vitro* activity studies. It predicts that feruloyl adenylate is a more preferred substrate over cinnamoyl adenylate in Ok4CL7 and -15. In Ok4CL7, Asp456 and Lys473 interact with ribose, whereas Lys477 and Gln482 interact with α -phosphate (**Figure 3.3**). Feruloyl adenylate

(preferred substrate) showed Pi-sigma interaction with Tyr277 and carbon-hydrogen bond with Ser345 and Cys346 (**Figure 3.3**). These additional interactions are absent in cinnamoyl adenylate; hence, it is not preferred as a substrate by Ok4CL7. In Ok4CL15, Asp419 and Arg434 interact with ribose, whereas Lys525 forms a salt bridge with α -phosphate of adenylate. Ser243 and Cys336 make polar contacts with functional groups of feruloyl adenylate. These substrate-specific interactions are absent in the case of the Ok4CL15 and cinnamoyl adenylate complex (**Figure 3.3**).

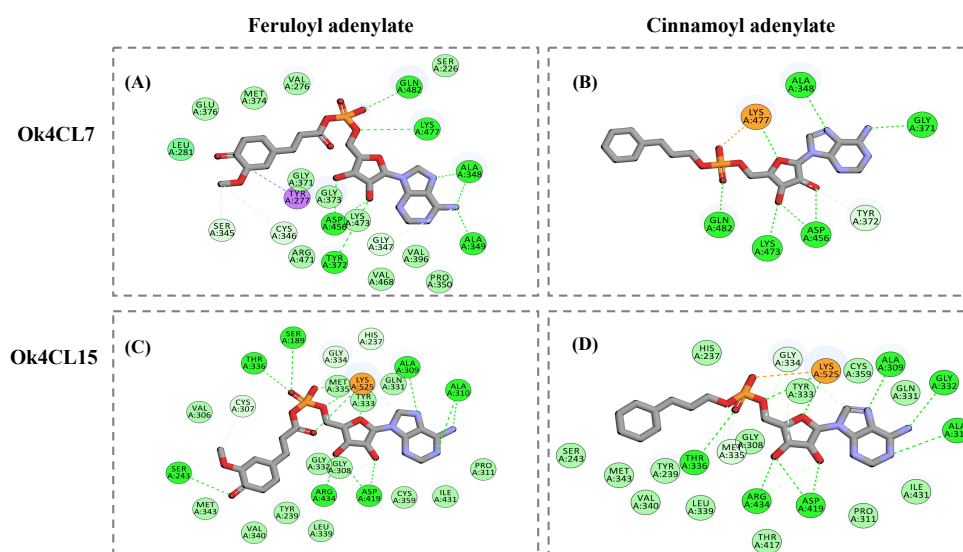


Figure 3.3 Substrate preferences of Ok4CLs. Molecular interaction of substrate binding pocket of Ok4CL7 with substrate-specific functional groups. **(A)** Feruloyl adenylate contributes to forming the salt bridge with Tyr277 hydrogen bonds with Ser345 and Cys346. **(B)** Unfavourable substrate cinnamoyl adenylate showed the absence of substrate-specific interactions. **(C)** Similarly, in Ok4CL15, feruloyl adenylate functional groups form hydrogen bonds with Ser243 and Cys307, whereas **(D)** Cinnamoyl adenylate does not have any specific interaction.

Most of the substrate-binding site residues of Ok4CL7 and -15 are conserved with 4CLs from other plants. In *Nicotiana tabacum* 4CL2, lysine interacts with α -phosphate of adenylate. Mutation of this lysine residue to alanine results in a complete loss of activity (Li and Nair 2015). Similarly, the indispensability for tryptophan present near the substrate has been validated in *N. tabacum* by mutagenesis. It has been observed that this mutant showed decreased activity. Tyr277 and Tyr239 also showed proximity with a substrate suggesting likelihood in their functionality with *N. tabacum* and *P. tomentosa* 4CLs. This interaction

analysis showed high corroboration with studies and highlights conservation of the substrate-specificity mechanism in different plants 4CLs.

3.3.3 Models of Ok4CL7 and -15 mutants show variation in affinity towards sinapic acid substrate compared to respective wild type enzyme

The molecular interaction study selected 27 and 29 positions in Ok4CL7 and -15 for mutation, respectively. Mutant models for each selected position in both 4CLs were prepared using the DUET server by replacing the original amino acid residue with all 19 standard amino acids. Models of Ok4CL7 and -15 were developed (**Table 3.3 and 3.4**). Folding energy ($\Delta\Delta G$ or Delta Delta G or DDG) is a predicted measurement for protein stability compared to protein with single point mutation. It gives an idea about the effect of single-point mutation on protein stability. Based on $\Delta\Delta G$ values provided by the DUET server, the top three stable replacements were selected for point mutation. Selected values or point mutations are circled in red (**Table 3.3 and 3.4**). All replacement with negative $\Delta\Delta G$ value from further study.

Interaction of selected mutants studied with different substrates (**Table 3.1**) using AutoDock Vina tool as described previously. Binding scores are normalized by calculating z-score and heatmap generated from normalized binding energy scores. Hierarchical clustering is used to cluster binding scores with a similar pattern among mutants and vice versa. An increase in the binding score of Ok4CL7 with sinapoyl adenylate was observed upon replacement of lysine (K) at 477 positions with leucine (L) or isoleucine (I) or valine (V) (**Figure 3.5**). Similarly, in the case of Ok4CL15, replacement of K at 525 positions with isoleucine leads to increase binding energy towards sinapoyl adenylate, coumaroyl adenylate and cinnamoyl adenylate substrates (**Figure 3.6**).

4CL mutant enzyme were expressed and purified using affinity chromatography. Kinetic parameters of purified enzymes were studied in comparison with wild 4CL. Ok4CL7-K477V shows no activity with any substrates, including sinapic acid, ferulic acid, *p*-coumaric acid, caffeic acid and cinnamic acid. This results suggest that K at 477 positions is crucial for enzyme activity. *In vitro* activity and validation of other selected mutants is in progress.

Table 3.2 Predicted stability change ($\Delta\Delta G$) of Ok4CL7 variants. Replacement of amino acids at selected with all other 19 standard amino acids leads to change in folding energy. Replacements showing positive $\Delta\Delta G$ chosen for further study (highlighted in red circle).

Amino acid	V276	V277	S280	H317	S345	G347	A348	G371	G373	A379	L380	A381	S226	L281	C346	A349	P350	Y372	M374	E376	V396	D456	V468	R471	K473	K477	Q482
A	-0.792	-2.07	-0.247	-2.802	0.063	-0.621	0	-0.447	-0.066	0	-2.883	0	-0.687	-2.839	-1.827	0	-0.088	-2.244	-0.12	-0.179	-1.618	-0.306	-1.867	-1.05	-0.801	-0.423	-0.376
C	-0.841	-0.698	0.093	-1.501	-0.02	-0.566	-0.421	-0.735	-0.407	-0.652	-2.333	-1.405	-0.186	-1.777	0	-0.396	-0.555	-1.298	-0.292	-0.109	-1.445	-0.14	-0.941	-0.813	-1.224	-0.643	0.401
D	0.555	-1.196	-0.686	-2.58	-1.826	-1.244	-0.345	-0.731	0.06	-0.449	-2.146	-1.98	0.361	-2.924	-0.972	-0.582	0.181	-0.429	1.098	-1.364	-0.921	0	-2.055	-0.264	-0.532	-1.035	-0.573
E	0.703	-0.865	-0.522	-2.157	-1.551	-1.245	-0.289	-0.618	-0.009	-0.35	-1.787	-1.924	0.338	-2.449	-0.864	-0.568	0.081	-0.083	1.173	0	-1.01	-0.565	-1.634	-0.025	-0.357	-0.524	-0.13
F	-0.586	-0.671	-0.6	0.203	0.077	-0.936	-0.715	-0.957	-0.55	-0.547	-1.525	-1.246	-1.326	-1.445	-1.151	-0.613	-0.716	-0.593	0.451	-0.148	-1.154	-0.343	-1.146	-0.871	-0.636	-0.21	-0.167
G	-1.482	-2.697	-0.985	-3.152	-0.631	0	0.28	0	0	-0.92	-3.008	-1.548	-0.946	-3.311	-2.226	-0.252	-0.211	-2.487	0.197	-0.858	-2.309	-0.697	-2.196	-1.151	-1.199	-0.628	-0.516
H	-0.833	-0.939	-1.231	0	-1.186	-1.346	-0.54	-0.803	-0.511	-1.318	-2.051	-1.907	-1.541	-2.51	-2.104	-0.372	0.168	-0.655	0.57	-0.645	-1.564	-0.597	-1.677	-1.517	-1.047	-1.142	-0.601
I	0.187	-1.072	0.716	-0.693	1.248	0.189	0.182	0.16	0.265	0.975	-0.475	0.161	-0.07	-0.262	-0.416	0.204	0.036	-0.898	0.649	1.185	-0.27	1.312	-0.095	0.105	0.24	0.817	0.853
K	-0.285	-1.095	0.195	-1.595	-0.405	-0.991	-0.655	-0.626	-0.429	-0.587	-1.689	-1.305	-0.308	-1.562	-1.027	-0.724	-0.122	-1.104	1.019	-0.274	-0.887	-0.879	-1.306	-0.918	0	0	0.031
L	0.326	-1.008	0.81	-0.807	1.15	0.08	0.163	0.054	0.244	0.844	0	0.099	-0.116	0	-0.572	0.058	0.033	-0.962	0.63	1.083	-0.165	0.983	-0.222	0.024	0.291	0.753	0.793
M	-0.249	-0.798	0.32	-1.343	0.549	-0.087	-0.253	-0.3	-0.161	-0.073	-1.437	-0.708	-0.208	-0.65	0.156	-0.272	-0.297	-1.117	0	1.117	-0.843	1.196	-0.348	-0.497	0.128	0.354	1.056
N	-0.269	-1.484	-0.454	-2.514	-1.148	-0.723	-0.345	-0.518	-0.113	-0.341	-1.717	-1.278	-0.226	-2.239	-1.24	-0.473	0.023	-1.879	0.786	-0.794	-0.849	-1.233	-1.612	-0.509	-0.657	-0.786	-0.686
P	-0.228	-1.814	0.057	-1.949	-0.039	-0.719	-0.415	-0.613	-0.443	-0.362	-2.017	-0.392	-0.696	-1.092	-1.243	-0.704	0	-1.657	-0.047	0.028	-0.707	0.139	-1.256	-0.97	-0.28	-0.193	-0.217
Q	-0.143	-1.3	-0.098	-1.983	-0.691	-0.827	-0.445	-0.651	-0.203	-0.379	-1.745	-1.328	-0.464	-1.802	-1.051	-0.622	0.007	-1.504	0.938	-0.462	-0.951	-0.875	-1.47	-0.519	-0.723	-0.784	0
R	-0.315	-0.78	-0.334	-1.673	-0.865	-0.705	-0.17	-0.295	-0.138	-0.635	-1.194	-1.146	-0.153	-1.734	-1.094	-0.292	0.217	-0.967	1.088	-0.291	-0.812	-0.966	-0.942	0	-0.186	-0.52	0.056
S	-1.174	-2.012	0	-2.834	0	-1.019	-0.638	-0.814	-0.341	-0.971	-2.722	-1.687	0	-3.119	-1.934	-0.583	-0.267	-2.481	0.303	-1.217	-1.816	-1.767	-2.231	-1.405	-1.389	-1.426	-0.717
T	-0.809	-1.764	-0.037	-2.345	-0.519	-0.856	-0.591	-0.7	-0.298	-0.641	-2.35	-1.552	-0.153	-2.462	-1.332	-0.556	-0.296	-2.106	0.506	-0.692	-1.452	-1.026	-1.76	-0.851	-1.063	-0.755	-0.238
V	0	-1.441	0.525	-1.198	1.023	0.064	0.139	0.199	0.319	0.777	-1.454	-0.257	-0.112	-1.423	-0.607	0.163	0.008	-1.194	0.41	0.841	0	0.882	0	-0.032	-0.111	0.629	0.637
W	-0.862	-0.349	-0.863	0.378	-0.604	-1.226	-0.814	-0.952	-0.865	-1.034	-1.669	-1.662	-1.123	-1.855	-1.252	-0.814	-0.591	-0.211	0.009	-0.531	-1.392	-0.486	-1.248	-0.434	-0.578	-0.174	0
Y	-0.527	0	-0.347	0.304	-0.024	-0.866	-0.537	-0.609		-0.485	-1.506	-1.279	-0.816	-1.655	-1.18	-0.449	-0.399	0	0.446	-0.106	-1.158	-0.052	-0.983	-0.129	-0.085	0.323	0.349

Table 3.3 Predicted stability change ($\Delta\Delta G$) of Ok4CL15 variants. Replacement of amino acids at selected with all other 19 standard amino acids leads to change in folding energy. Replacements showing positive $\Delta\Delta G$ chosen for further study (highlighted in red circle).

Amino acids	V238	V239	V242	P279	V306	G308	A309	G332	G334	I339	V340	I341	S189	H237	S243	G307	A310	P311	Q331	Y333	M335	T336	M343	G359	T417	D419	I431	R434	K525
A	-1.757	-2.882	-2.569	-0.991	-2.443	-0.619	0	-0.782	-0.552	-2.948	-2.21	-3.05	-0.426	-1.101	-0.58	-1.94	0	-0.253	-0.018	-2.83	-1.899	-0.904	-2.042	-0.632	-0.578	0.265	-2.206	-1.113	-0.088
C	-1.189	-1.66	-2.239	-0.277	-1.19	-0.591	-0.354	-1.24	-0.75	-2.148	-1.307	-1.676	-0.03	-0.21	-0.504	0	-0.475	-0.277	0.192	-1.184	-1.304	-0.5	-2.113	0	-0.207	0.174	-0.796	-0.978	-0.309
D	-2.033	-2.355	-3.454	-2.47	-3.1	-1.466	-0.659	-1.619	-0.536	-2.322	-2.056	-3.262	0.269	-0.716	-0.874	-1.419	-1.349	-0.183	-1.666	-1.882	-0.268	-1.225	-2.038	-0.481	-1.446	0	-1.806	-0.549	0.164
E	-1.844	-1.908	-3.308	-2.029	-2.773	-1.45	-0.665	-1.555	-0.75	-1.97	-2.037	-2.887	0.393	-0.506	-0.836	-1.269	-1.446	-0.068	-1.105	-1.572	0.035	-1.048	-1.676	-0.434	-1.512	-0.453	-1.472	-0.19	0.365
F	-1.534	-0.936	-1.981	-0.23	-1.601	-0.938	-0.891	-1.464	-1.366	-1.783	-1.518	-1.675	-0.466	0.463	-0.818	-1.111	-1.252	-0.873	-0.309	-0.987	-0.41	-0.89	-0.488	-1.053	-0.68	-0.377	-1.263	-1.013	-0.025
G	-2.646	-3.51	-3.353	-1.759	-2.99	0	-0.02	0	0	-2.342	-2.608	-3.459	-0.501	-1.379	-1.245	-2.36	-0.713	-0.41	-0.232	-2.988	-2.289	-1.805	-2.341	-1.14	-1.194	-0.459	-2.394	-1.205	-0.463
H	-2.048	-1.608	-2.563	-1.602	-2.611	-1.352	-1.104	-1.236	-1.315	-2.014	-1.942	-2.507	-0.97	0	-1.504	-2.226	-0.997	-0.271	-0.208	-1.254	-0.972	-1.727	-1.543	-1.455	-1.536	-0.639	-1.559	-1.596	-0.846
I	-0.397	-1.57	-0.878	0.857	-0.249	0.189	0.052	-0.263	-0.03	-0.976	-0.736	-0.675	0.357	0.179	0.343	-0.516	-0.205	-0.036	-1.469	-1.386	-0.404	0.137	-0.258	0.134	0.709	1.108	0.026	0.899	0
K	-1.6	-1.448	-2.412	-1.047	-1.945	-1.025	-0.762	-1.067	-1.004	-1.8	-1.625	-2.135	-0.305	-0.466	-0.546	-1.371	-1.269	-0.238	0.346	-1.463	-0.25	-1.144	-0.869	0.113	-1.109	-0.474	-1.224	-0.969	0
L	-0.301	-1.502	-0.823	0.789	-0.459	0.078	0.041	-0.382	-0.276	0	-0.793	0	0.324	0.106	0.477	-0.662	-0.38	-0.038	-1.442	-1.453	-0.494	0.199	-0.336	-0.046	0.648	1.025	-0.759	-0.057	0.876
M	-0.581	-1.408	-1.28	0.467	-0.682	-0.088	-0.169	-0.752	-0.545	-1.435	-0.931	-0.795	0.332	-0.202	0.002	0.201	-0.511	-0.145	-1.179	-1.131	0	0.024	0	-0.165	0.423	1.255	-0.675	-0.533	0.485
N	-1.448	-2.276	-2.47	-1.64	-2.151	0.83	-0.291	-0.953	-0.601	-1.846	-1.443	-2.485	-0.196	-0.936	-0.984	-1.659	-0.71	0.081	-0.533	-2.141	-0.506	-1.311	-1.177	-0.096	-1.422	-0.828	-1.238	-0.61	-0.208
P	-1.471	-2.4	-1.595	0	-1.316	-0.719	-0.606	-1.018	-0.927	-2.06	-1.798	-1.814	-0.273	-0.506	-0.31	-1.294	-1.092	0	0.375	-2.181	-1.423	-0.543	-1.6	-0.386	-0.388	0.423	-1.712	-1.046	0.483
Q	-1.529	-1.899	-2.597	-1.35	-2.086	-0.942	-0.526	-1.207	-0.866	-1.83	-1.605	-2.295	-0.382	-0.547	-1.059	-1.497	-1.052	-0.098	0	-1.898	-0.256	-1.292	-1.028	-0.173	-1.367	-0.506	-1.217	-0.75	-0.131
R	-1.118	-1.088	-1.862	-1.109	-1.577	-0.739	-0.46	-0.662	-0.677	-1.165	-1.124	-1.677	-0.161	-0.869	-0.902	-1.477	-0.755	0.181	0.416	-0.966	-0.045	-1.16	-0.759	-0.263	-0.9	-0.639	-0.572	0	-0.249
S	-2.227	-2.81	-3.242	-2.17	-2.877	-1.037	-0.75	-1.256	-0.877	-2.758	-2.25	-3.249	0	-0.972	0	-2.378	-0.838	-0.159	-0.58	-2.491	-1.101	-1.729	-1.826	-0.314	-1.449	-1.179	-1.866	-1.503	-0.763
T	-1.896	-2.511	-2.933	-1.484	-2.346	-0.88	-0.679	-1.123	-0.784	-2.429	-1.908	-2.826	-0.117	-0.669	-0.705	-1.814	-0.814	-0.206	-0.141	-2.26	-0.749	0	-1.394	-0.058	0	-0.487	-1.502	-0.948	-0.423
V	0	-2.089	0	0.48	0	0.064	-0.058	-0.19	-0.153	-1.888	0	-1.737	0.259	-0.169	0.171	-0.759	-0.215	-0.092	0.912	-1.648	-0.817	0.103	-0.678	-0.208	0.568	1.167	-0.913	-0.104	0.533
W	-1.65	-0.383	-2.101	-0.926	-2.047	-1.307	-1.028	-1.64	-1.556	-1.752	-1.805	-1.899	-0.572	0.634	-1.065	-1.334	-1.524	-0.588	-0.288	-0.554	-0.755	-1.039	-1.013	-1.313	-0.986	-0.557	-1.15	-0.563	0.042
Y	-1.594	0	-2.103	-0.524	-1.769	-0.89	-0.774	-1.271	-1.204	-1.729	-1.644	-1.946	-0.242	0.303	-0.701	-1.328	-1.153	-0.268	0.355	0	-0.47	-0.633	-0.747	-0.989	-0.767	-0.076	-0.976	-0.259	0.368

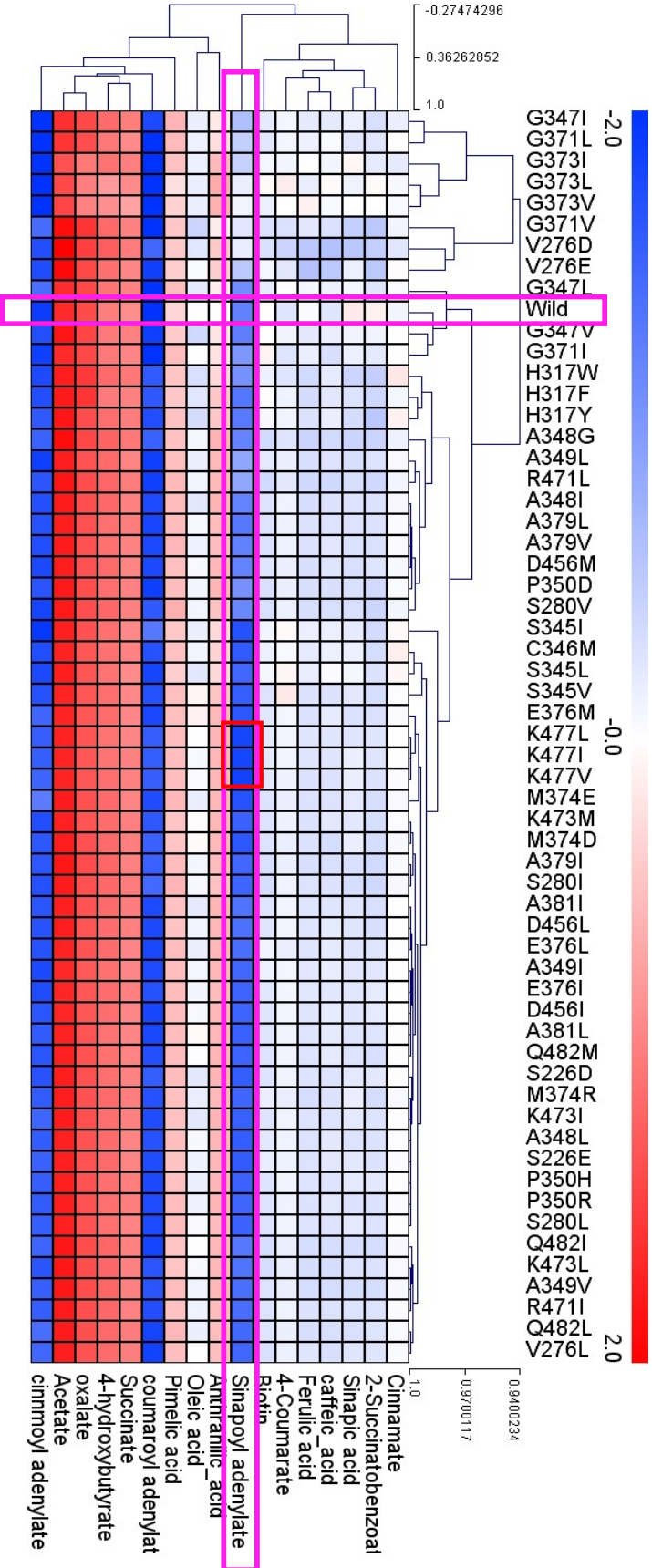


Figure 3.5 Binding energy of Ok4CL7 variants with different substrates. Significant change binding energy is highlighted in the red box.

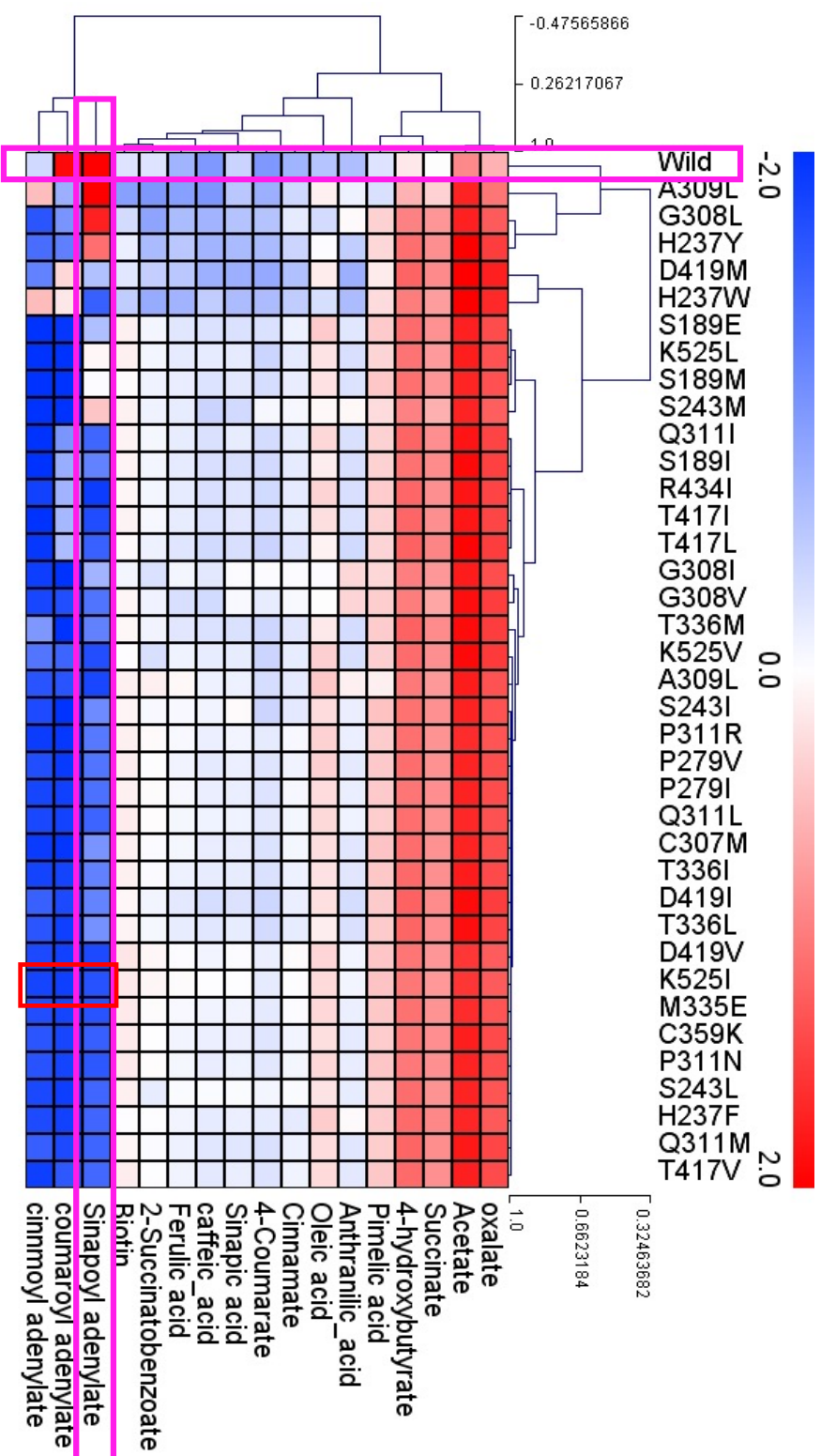


Figure 3.6 Binding energy of Ok4CL15 variants with different substrates. Significant change binding energy is highlighted in the red box.

3.4 Summary

The sequence of amino acids (primary structure) in protein is crucial for three-dimensional structure or conformation and ultimately for enzyme activity. Multiple sequence alignment of reported plant 4CLs and *O. kilimandscharicum* 4CLs suggested that box I, box II and other residues required for enzyme activity are conserved. Moreover, Ok4CL7 and -15 showed differential substrate specificity towards sinapic acid, which is different from reported 4CLs from *G. max* and *N. tabacum*. The comparison of 4CL sequences suggests that other than the position of V and L, some other residues are responsible for substrate selectivity.

In silico analysis of 4CL interaction with preferred and non-preferred substrates leads to identifying amino acid residues interacting with the substrates. Models were developed by replacing interacting amino acids with other standard amino acids to check their importance in the interaction with the substrates. In Ok4CL7, replacing lysine at 477 positions with either leucine, isoleucine, or valine leads to an increase in affinity for sinapic acid. In the case of Ok4CL15, replacing lysine at 525 with isoleucine leads to an increase in affinity for sinapic acid. Developing 4CL mutants and testing their activity with different substrate is in progress. A preliminary experiment shows that replacing leucine at 477 with valine in Ok4CL7 leads to loss of enzyme activity.

Chapter 4

In planta characterization of three 4CL isoforms from *Ocimum kilimandscharicum*



4.1 Introduction

Plants produce a variety of specialized metabolites to cope with various biotic and abiotic stress conditions and modulate their growth, development and health accordingly. Many of these specialized metabolites and their derivatives have similar functions with less or more efficiency (like enzyme isoforms). Having several alternate combinations for the similar function is an advantage to plants, which helps in modifying the response or the strength of response to a specific stimuli, leading to the achievement of the desired goal. Several reports suggested that 4-Coumarate-CoA Ligases (4CLs) are involved in the protection of plants from biotic and abiotic stress conditions (reviewed in Lavhale et al. 2018). 4CL isoforms from several plant species have been characterized by their overexpression in model plants, *Nicotiana tabacum* or *Nicotiana benthamiana*. For example, Df4CL2 isoform from *Dryopteris fragrans* (commonly known as the fragrant woodfern) was overexpressed in *N. tabacum* to study its *in planta* function. In this study, the authors found that Df4CL2 was involved in the biosynthesis of both lignin and flavonoids (Li et al. 2020). Similarly, *Fm4CL-like-1* gene from Manchurian ash (*Fraxinus mandshurica*) was overexpressed in *N. tabacum* for its functional characterization (Chen et al. 2019). Considering these recent reports, it is apparent that the *in planta* characterization of the enzymes of plant origin gives better idea about their physiological functions at the metabolite and morphological levels. Another advantage of *in planta* characterization of such enzymes with multiple isoforms is that it provides the information on their unique as well as common functions at the whole-plant level.

Earlier, we have investigated the biochemical and biophysical characteristics of two 4CL isoforms (Ok4CL7 and -15) from the plant species *O. kilimandscharicum* (Chapter 2; Lavhale et al. 2021). Both Ok4CL7 and -15 have common as well as unique differential substrate preferences. For example, *p*-coumaric acid and ferulic were used by both of them, whereas caffeic acid is exclusively used by Ok4CL7 and sinapic acid by Ok4CL15. Ok4CL11 was also tested for the *in vitro* activity, but it was found to be inactive with any of the substrates tested under given experimental conditions (cinnamic acid, *p*-coumaric acid, caffeic acid, ferulic acid, and sinapic acid). Thus, the current chapter deals with the *in planta* functional characterization of Ok4CL isoforms (Ok4CL7, -11 and -15). Owing to a poor regeneration efficiency coupled with limited success of *Agrobacterium*-mediated stable transformation methods for several members of *Ocimum* species, in this work, we have overexpressed three Ok4CL isoforms (Ok4CL7, -11 and -15) in *N. benthamiana* and investigated their effects on the plant phenotypes and metabolite profiling. Interestingly, we observed a ‘rootless or

‘retarded root growth’ phenotype in the *Ok4CL11* overexpression lines of *N. benthamiana*, whereas no such phenotype was observed in *Ok4CL7* and *-15* overexpression lines. Considering the role of 4CLs in the phenylpropanoid biosynthesis pathway, we also carried out metabolite profiling of the overexpression lines and wild-type *N. benthamiana* plants. To dissect the ‘rootless’ or ‘retarded root growth’ phenotype observed in the *Ok4CL11* overexpression lines, a number of approaches, such as transcriptomics, auxin and kaempferol quantification, grafting (*in vitro*), chemical treatment, phenotype rescue, etc. were undertaken. The findings of these experiments and ongoing work are described in this chapter.

4.2 Materials and methods

4.2.1 Generation of *Ok4CL7*, *-11* and *-15* overexpression gene constructs

pRI101AN binary vector is used for *in planta* gene expression. This vector contains a 5' untranslated region (UTR) of the *alcohol dehydrogenase* gene from *Arabidopsis thaliana*, which helps to enhance the stability of the transcript. It is suitable for gene expression in dicotyledonous plants because pRI101AN contains a 35S promoter from the Cauliflower mosaic virus (CaMV) and the *nopaline synthase* terminator (**Figure 4.1A**). The longest open frame (ORF) of *Ok4CL7* (1731 bp), *-11* (1731 bp) and *-15* (1530 bp) were amplified by PCR using high-fidelity Phusion DNA Polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, United States) using cDNA synthesized from the leaf tissue of *O. kilimandscharicum* as a template. Subsequently, these ORFs were cloned in to a sub-cloning vector pGEM-T Easy (Promega Corporation, Madison, Wisconsin, United States) and confirmed by Sanger sequencing. Further, these ORF sequences were mobilized in to a binary vector pRI101AN. *Ok4CL7* and *-11* were cloned using *NdeI* and *EcoRI* restriction enzyme sites in forward and reverse primers, respectively. Similarly, *Ok4CL15* was cloned using *NdeI* and *SacI* restriction enzyme sites in forward and reverse primers, respectively. Cloning of *Ok4CL* genes in pRI101AN vector was confirmed using restriction digestion (**Figure 4.1B**), followed by Sanger sequencing.

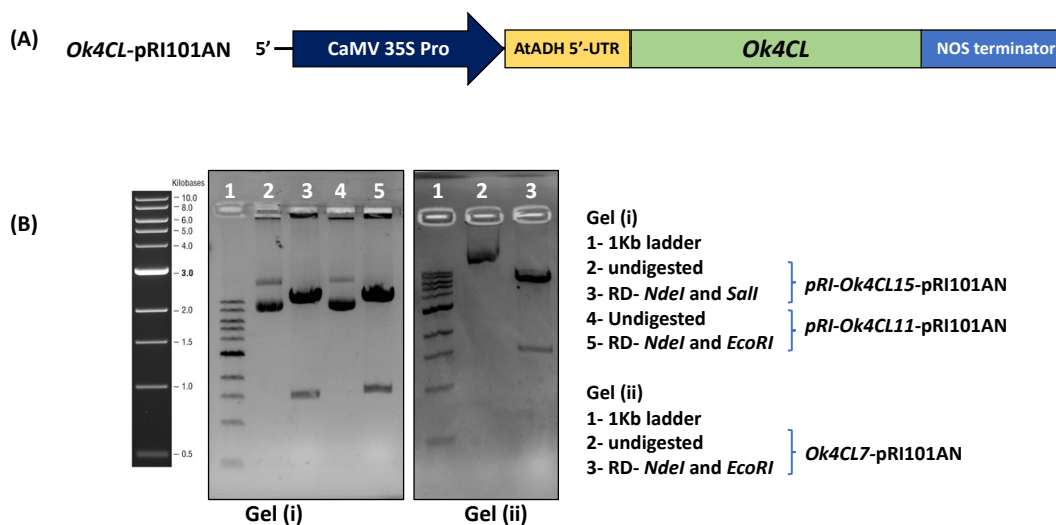


Figure 4.1 *Ok4CLs* overexpression constructs in a binary vector pRI101AN. (A) Schematic representation of the overexpression constructs of *Ok4CL* genes. (B) Confirmation of *Ok4CL7*, *-11* and *-15* (Gel i-ii) gene constructs in pRI101AN vector by *EcoRI* and *NdeI/Sall* restriction enzyme digestions.

4.2.2 Development of *Ok4CL* overexpression lines of *N. benthamiana*

The *Ok4CL11OE*-pRI101AN construct was transformed in *A. tumefaciens* (GV3101 strain) using the freeze and thaw transformation method. Overexpression lines of *Ok4CL7*, *-11* and *-15* in *N. benthamiana* were generated through *Agrobacterium*-mediated leaf transformation method as described by Horsch et al. (1985) (**Figure 4.2**). After 3 to 4 weeks on regeneration medium, shoots emerged from the leaf-derived calli. Healthy shoots (~1 to 2 cm in length) were excised and transferred on to the rooting media containing the selection antibiotic kanamycin (50 µg/mL) (**Figure 4.2**). These overexpression lines were screened for the presence of the respective transgene. Confirmed overexpression lines were used to study the function of *Ok4CL* isoforms by analysing their phenotypic traits, metabolite content, and gene expression analysis.

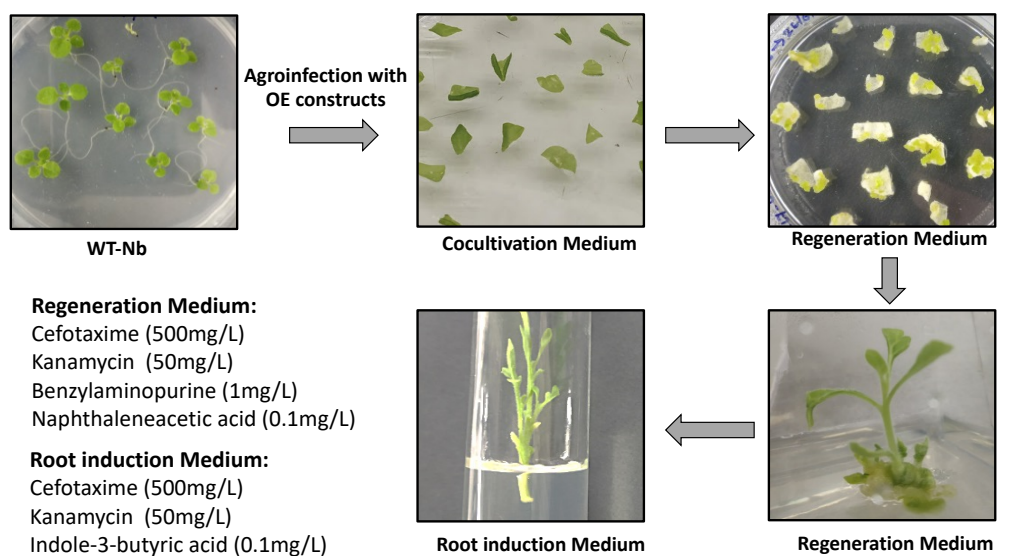


Figure 4.2 Flow chart depicting the procedure for generation of transgenic *N. benthamiana* plants using *Agrobacterium*-mediated leaf transformation method (as described in the Methods section).

4.2.3 Analysis of root growth in *Ok4CL* overexpression lines and wild-type *N. benthamiana* plants

Wild-type *N. benthamiana* plants and overexpression (OE) lines of Ok4CL isoforms (Ok4CL7/11/15) were sub-cultured using stem as explants and grown *in vitro* for 3 to 4 weeks under long-day conditions (16 h light and 8 h dark) in a Plant tissue culture room. After 25 days of stem sub-culture, root growth parameters like number of roots per plant, main root length (cm) and lateral root formation (numbers and lengths) were scored for *Ok4CL* overexpression lines compared to wild-type (WT) plants.

Further, root apices (approximately 2 cm in length from the tip) of *OK4CL11-OE* lines and WT *N. benthamiana* plants were excised and cultured on Murashige and Skoog's (MS) medium (Murashige and Skoog, 1962) having 2% sucrose (w/v) under long-day conditions. Root apex cultures were observed over a period of ten days, and root growth measurements (root length, number of lateral roots and their lengths) were taken daily from 5th day onwards. On the 10th day, root tissues (1 cm apices and rest of the roots) were harvested separately in liquid nitrogen for metabolite analysis. All samples were stored at -80 °C until further use. Metabolites were extracted as described previously (Methods section).

4.2.4 Metabolite profiling and transcriptome of *Ok4CL* overexpression lines and wild-type *N. benthamiana* plants

Wild-type *N. benthamiana* plants and overexpression (OE) lines of Ok4CL isoforms (Ok4CL7, -11 and -15) were grown *in vitro* for 25 days. Whole plants were harvested in liquid nitrogen. Six plants of the individual line were selected for metabolite profiling and RNA-sequencing (RNA-seq). Three biological replicates were prepared by pooling tissues from two plants for each biological replicate (n=3). The tissues were crushed to a fine powder in liquid nitrogen using mortar and pestle. Aliquots of crushed tissue were prepared and stored at -80 °C for further use.

Metabolites were extracted in 300 µL of 80% LCMS grade methanol (0.1 % formic acid) using 100 mg of finely crushed tissue. Tubes were kept for 2 minutes on the vortex to mix solvent and tissue properly. Subsequently, tube contents were sonicated for 15 minutes, and the cell lysate was separated by centrifugation at 15000 rpm for 15 minutes at 4 °C. The supernatant was collected and filtered using a 0.22 µm syringe filter. All samples were stored at -80 °C for further use.

For RNA-seq, the total RNA from the respective aliquots of the samples used for metabolite analysis was isolated using RNAiso Plus (TaKaRa, Japan) as per the manufacturer's instructions. The total RNA was extracted from 100 mg of fine crushed tissue. RNA was solubilized in 50 µL of DEPC treated water. RNA quality was checked on 1% agarose gel (**Figure 4.3**). Genomic DNA contamination was removed by DNase treatment. Approximately, 3 µg of the total RNA was processed for RNA-seq analysis.

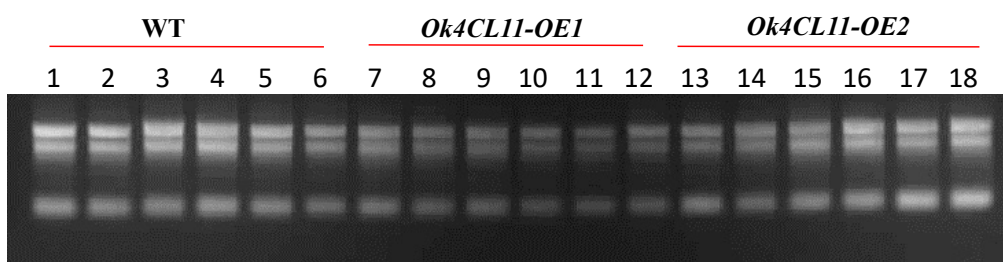


Figure 4.3 The total RNA separation on 1% agarose gel. Well numbers (1 to 6) represent RNA isolated from wild-type *N. benthamiana* plants, well numbers 7 to 12 indicate RNA from *Ok4CL11-OE1 N. benthamiana* line, whereas well numbers 13 to 18 represents RNA for *Ok4CL11-OE2 N. benthamiana* line.

The concentrations of purified mRNA from each sample were quantified on Qubit. Further, using bioanalyzer, it was observed that all nine samples had RNA integrity (RIN) values greater than 7 confirming the good quality of our RNA samples. cDNA libraries were run on Illumina platform (NovaSeq 6000) using paired end sequencing with read length of 150 bp.

The fastq files were trimmed using fastp tools based on the quality and trim 2 bases from front and 3 bases from tail. All reads are trimmed to avoid specific sequence bias. Trimmed samples are mapped using the Hisat 2 genome assembler against the reference *N. benthamiana* draft genome sequence v1.0.1 (https://solgenomics.net/organism/Nicotiana_benthamiana/genome). Read counts data normalization, visualization and differential gene expression analysis was performed using the DESeq2 package (version 1.22.2). The criteria used for differential gene expression analysis comparisons amongst the samples include $\log_2FC > 1$ and padj value < 0.05 .

Gene expression analysis using real time PCR

The total RNA (2 μg) was used for cDNA synthesis using SSIV-RT (Invitrogen) and Oligo dT primers. RT-qPCR reactions were run on CFX96 Real-Time System (BIO-RAD) with gene-specific primers (**Table 4.1**). The reactions were carried out using TAKARA SYBR® green master mix (Takara-Clontech) and incubated at 95 °C for 30 s, followed by 40 cycles at 95 °C for 5 s, gene-specific annealing temperature for 15 s and extension for 72 °C for 15 s. PCR specificity was checked by melting curve analysis, and data were analyzed using the $2^{-\Delta\Delta C_t}$ method.

4.2.5 Root growth phenotype rescue by gene silencing and media supplements

Root growth was affected in *Ok4CL11* overexpressing *N. benthamiana* plants. There are several ways to reverse this phenotype. Phenotypes could be reverted by gene silencing or providing media supplements to overcome the depletion of particular metabolites responsible for the phenotype. These two strategies (silencing of *Ok4CL11* in the *Ok4CLOE* background and media supplements to improve lignin biosynthesis in the *Ok4CL11*-OE lines) were used to complement the root growth phenotype. Silencing construct of *Ok4CL11* in pRI101AN was prepared by cloning an unique 350 bp sequence of its sense and antisense strand with 500 bp intron in between the two complementary sequences, which will form double stranded RNA (dsRNA) after transcription (**Figure 4.4A**). The gene silencing mechanism (RNA interference;

RNAi) involves the formation of dsRNA with the help of DICER complex, which is then cleaved by the RISC complex (RNA-induced gene silencing), leading to the formation of 21 to 22 nt short-interfering RNAs (siRNAs.) These siRNAs bind to the target mRNAs, causing degradation of target mRNAs (**Figure 4.4A**). The *Ok4CL11-pRII01AN* RNAi construct was transformed into *A. tumefaciens* (GV3101 strain), confirmed by restriction digestion and sequence verified. The leaves of the two *Ok4CL11-OE* lines (OE1 and OE2) were transformed with the *Ok4CL11* RNAi construct using *Agrobacterium*-based plant transformation method (described earlier) and complementation lines of *N. benthamiana* were generated to observe the rescue of root growth phenotype.

To check whether the root growth is affected by decreased lignin content or increased kaempferol level, wild-type plants will be sub-cultured using stem as an explant on MS media supplemented with kaempferol (1 μ M). Also, *Ok4CL11* overexpressing *N. benthamiana* lines (OE1 and OE2) will be grown on a lignin precursor, coniferyl aldehyde. We are also planning to sub-culture wild-type *N. benthamiana* plants on MS medium containing two auxin inhibitors (2,3,5-triiodobenzoic acid [TIBA; 10 μ M] and Naphthylphthalamic acid [NPA; 10 μ M]) as well as kaempferol (100 μ M) to study if the application of these auxin inhibitors can lead to the ‘rootless or ‘retarded root growth’ phenotype.

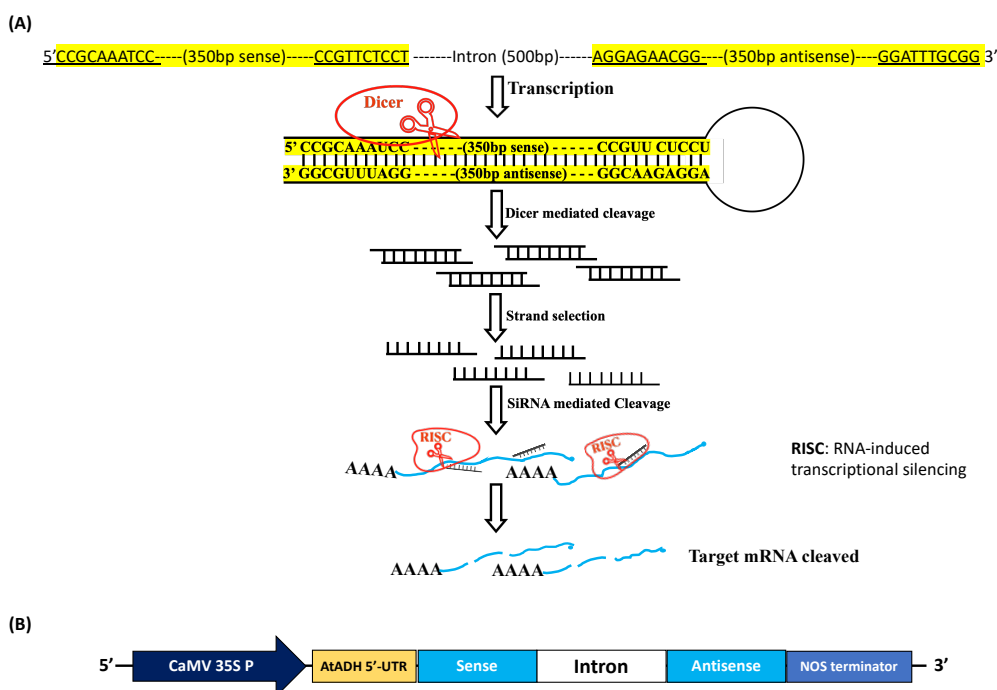


Figure 4.4 (A) Mechanism of siRNA-mediated post-transcriptional gene silencing. (B) Schematic representation of the *Ok4CL11-pRII01AN* RNAi construct. Sense or antisense is

the partial sequence of *Ok4CL11* (350 bp) and the intron in between both sequences is 500 bp.

4.2.6 Quantification of GUS activity from the *Ok4CL11* overexpression lines and wild-type *N. benthamiana* plants transformed with the *DR5pro:GUS-pBI101* construct

DR5 is an artificially synthesized and auxin-responsive promoter that contains seven direct tandem repeats of 11 bp, including the auxin-responsive factor gene-binding site (TGT CTC) (Ulmasov et al., 1997). *DR5* is widely used to monitor auxin accumulation throughout the plant. Auxin is synthesized mainly in the shoot apex, and young leaf. Additionally, root tip is also known to synthesize auxin to support its growth (Ljung et al. 2005). Plants establish polar auxin transport (PAT) throughout the stem from the stem apex to its base, which is crucial for overall root growths, including primary root length, root hair formation, and the emergence of lateral roots (Manzano et al. 2014; Orman-Ligeza et al. 2016). To investigate the contribution of auxin accumulation towards retarded root growth in *Ok4CL11-OE* lines, the *DR5* promoter sequence (359 bp) driving the expression of *GUS* gene (*DR5pro:GUS*) was cloned in the pBI101 binary vector, using *BamHI* restriction enzyme site on both primers and the correct orientation of the promoter sequence was confirmed by sanger sequencing (Figure 4.5).



Figure 4.5 Schematic representation of *DR5pro:GUS-pBI101*- construct. *GUS* gene sequence is 1.8 kb in length.

GUS gene encodes a β -glucuronidase, which hydrolyses X-Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid), and gives an insoluble and indigo-coloured compound. The *DR5pro:GUS-pBI101* binary construct was transformed into *A. tumefaciens* (GV2260) and subsequently used to develop its stable transgenic plants in the background of WT and *Ok4CL11-OE* lines (OE1 and OE2) of *N. benthamiana* (as per the protocol described earlier). WT as well as *Ok4CL11-OE* lines containing *DR5pro:GUS-pBI101* construct are generated. These lines will be confirmed and used for GUS assay as well as for quantification of MUG activity in a tissue-specific manner.

4.2.7 Grafting (homo- and hetero-grafts) to study its effect on root growth

Grafting is known to stimulate various molecular, biochemical and phenotypic changes in both scion (apex) and stock (base). This tool is routinely used to identify numerous long-distance macromolecules signals in plants. As we previously observed a root growth retardation in *Ok4CL11-OE* lines (OE1 and OE2) compared to WT *N. benthamiana* plants, we developed both homo- (WT/WT and OE2/OE2) and hetero-grafts (WT/OE2 and OE2/WT) of wild-type and *Ok4CL11-OE2* line. Homo-grafts of both plants were developed using stock and scion from the same plant. Hetero-grafts were developed using stock and scion from two different plants i.e. stock either from WT or OE2 line, and scion from either WT or OE2 line. *In vitro* grafts were developed using the ‘wedge’ grafting technique and kept horizontally on wet blotting paper containing 4 ml of liquid MS media (2% sucrose w/v) in a sterile petri dish. Plates were kept in the dark for 48 hours and then incubated under long-day conditions. After 7 days, grafts were transferred to magenta boxes containing MS medium (containing 2% sucrose w/v) in a vertical position. Carefully, any new branches grown on the stock tissues were removed before transferring to the solid MS medium. Two weeks after transfer to solid MS medium, root growth parameters were scored from all four types of grafts. Currently, the grafting experiment is being repeated using both *Ok4CL11-OE* lines (OE1 and OE2) along with WT.

4.3 Results and discussion

4.3.1 Overexpression lines were confirmed by PCR amplification of *Ok4CLs* using genomic DNA and cDNA from putative transformants of *N. benthamiana*

Putative transformants of *Ok4CLs* in *N. benthamiana* were confirmed by PCR amplification of respective transgenes using cDNA (synthesized from mRNA) as well as genomic DNA isolated from leaves of putative transformants and wild-type plants of *N. benthamiana* as negative control (**Figure 4.6**). We have generated five independent lines for *Ok4CL11* overexpression (*Ok4CL11-OE1* to *Ok4CL11-OE5*), two lines for *Ok4CL7* (*Ok4CL7-OE1* and *Ok4CL7-OE2*), and *Ok4CL15* gene (*Ok4CL15-OE1* and *Ok4CL15-OE2*) (**Figure 4.6**).

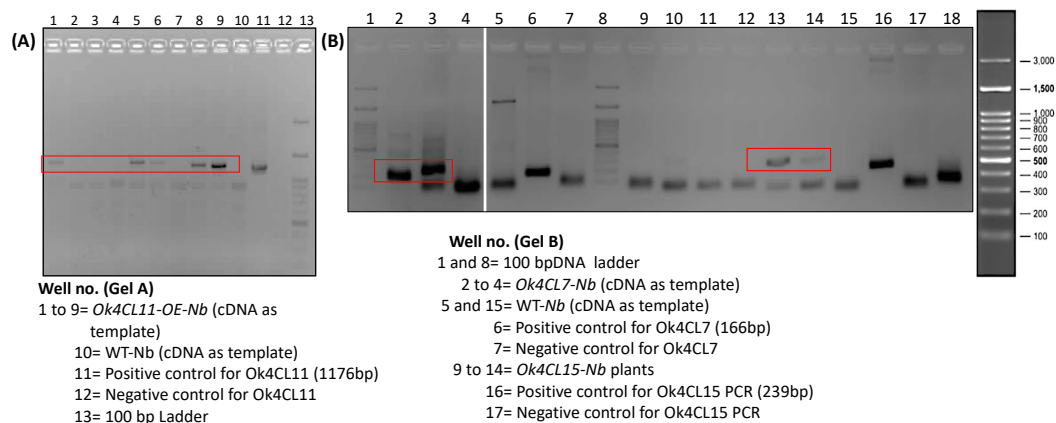


Figure 4.6 Screening of *Ok4CLs* overexpression lines using RT-PCR. RT-PCR was performed using cDNA isolated from leaves of (A) *Ok4CL11*, (B) *Ok4CL7*, and *Ok4CL15* overexpression lines and WT *N. benthamiana* plants (negative control). Plasmid (35S:*Ok4CL11*-pRI101AN) construct was used as a template for PCR positive control. Binary construct plasmids (35S:*Ok4CL11*-pRI101AN, 35S:*Ok4CL7*-pRI101AN and 35S:*Ok4CL15*-pRI101AN) served as a template for PCR of respective *Ok4CL* gene (positive control). PCR master mix was used as a template for PCR negative control.

4.3.2 Root growth was affected in *N. benthamiana* plants overexpressing *Ok4CL11* gene

An interesting phenotype of ‘rootless’ or ‘retarded root growth’ was observed in *N. benthamiana* lines overexpressing *Ok4CL11* gene (OE1 and OE2). Such phenotype was also observed for other three *Ok4CL11*-OE lines (OE1-OE5). Further, the aerial (aboveground) morphology of *Ok4CL11* overexpression lines was similar to WT. Supplementing auxin in the media stimulates rooting (An et al. 2020). However, all five *Ok4CL11*-OE lines (*Ok4CL11*-OE1 to -5) when grown on MS medium supplemented with auxin IBA (Indole butyric acid, 0.1 mg/L), we observed no positive effect on root formation. Root growth analysis of *Ok4CL11* overexpression lines (OE1 and OE2) suggests that >60% plants were rootless and few plants had significantly affected root growth characteristics. Root number, root length and lateral root number (LR) were scored in *Ok4CL11* overexpression lines and wild-type *N. benthamiana* plants. *Ok4CL11* overexpression lines (OE1 and OE2) have 80% less root numbers, 50% less number of lateral roots, and root length was reduced by 50% compared to WT plants (Figure 4.7B-D). Root growth and morphology of *Ok4CL15* overexpression line

(*Ok4CL15-OE1* and *OE2*) was seen to be comparable to WT plants (**Figure 4.7A-D**). Root numbers were significantly higher in the *Ok4CL7* overexpression line (*Ok4CL7-OE1* and *OE2*) compared to wild-type *N. benthamiana* (**Figure 4.7A-D**).

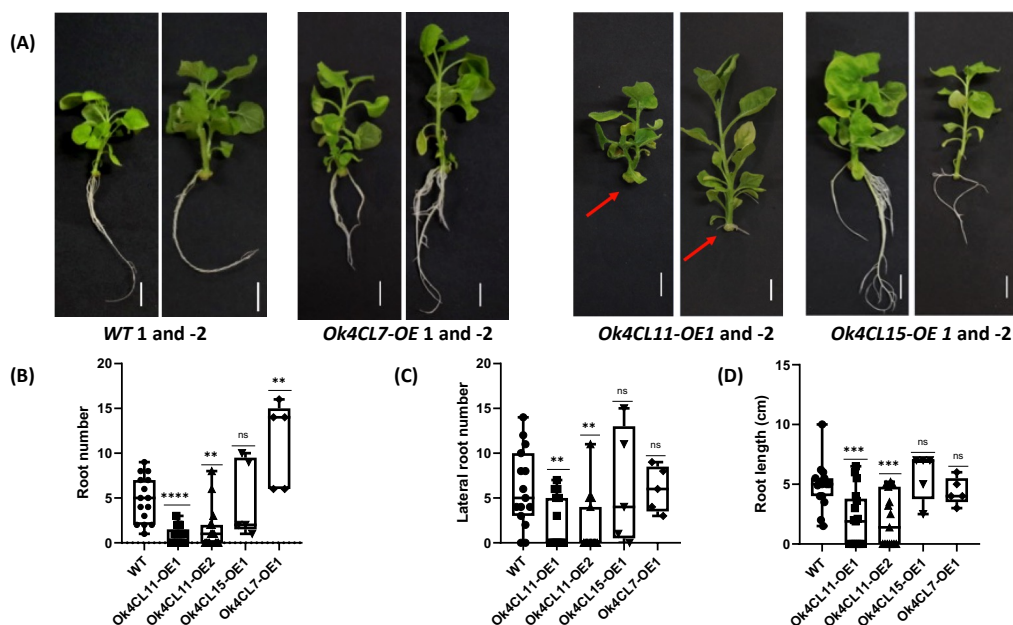


Figure 4.7 (A) Representative phenotypes of *N. benthamiana* plants overexpressing *Ok4CL7*, *-11* (OE1 and OE2), and *-15* compared to wild-type (WT) plants. Analysis of root phenotypes from WT and *Ok4CL*-OE *N. benthamiana* lines (*Ok4CL11*-OE1 and *Ok4CL11*-OE2). (B) Number of roots per plant, (C) number of lateral roots on the root; (D), and the overall root length (cm). WT is a non-transformed plant of *N. benthamiana*. N= number of biological replicates. At least fifteen plants per line or WT (i.e. n=15) were used for measurements of root growth parameters.

Further, we investigated the effect of *Ok4CL11* overexpression on root apex growth. Root apices (approximately 2 cm in length) were excised from the main roots of two *Ok4CL11* overexpression lines (OE1 and OE2) and WT, and grown *in vitro* for 10 days on MS medium. On an average, the overexpression lines (OE1 and OE2) show ~3 mm faster root growth compared to WT *N. benthamiana* plants. Moreover, root growth was found to be significantly higher in *Ok4CL11* overexpression lines compared to WT. The average length of the root for overexpression lines was nearly 20 mm at the end of experiment (10 days), whereas the root length was approximately 8 mm in WT. LR numbers and length was also higher in overexpression lines (OE1 and OE2) compared to WT (**Figure 4.8**).

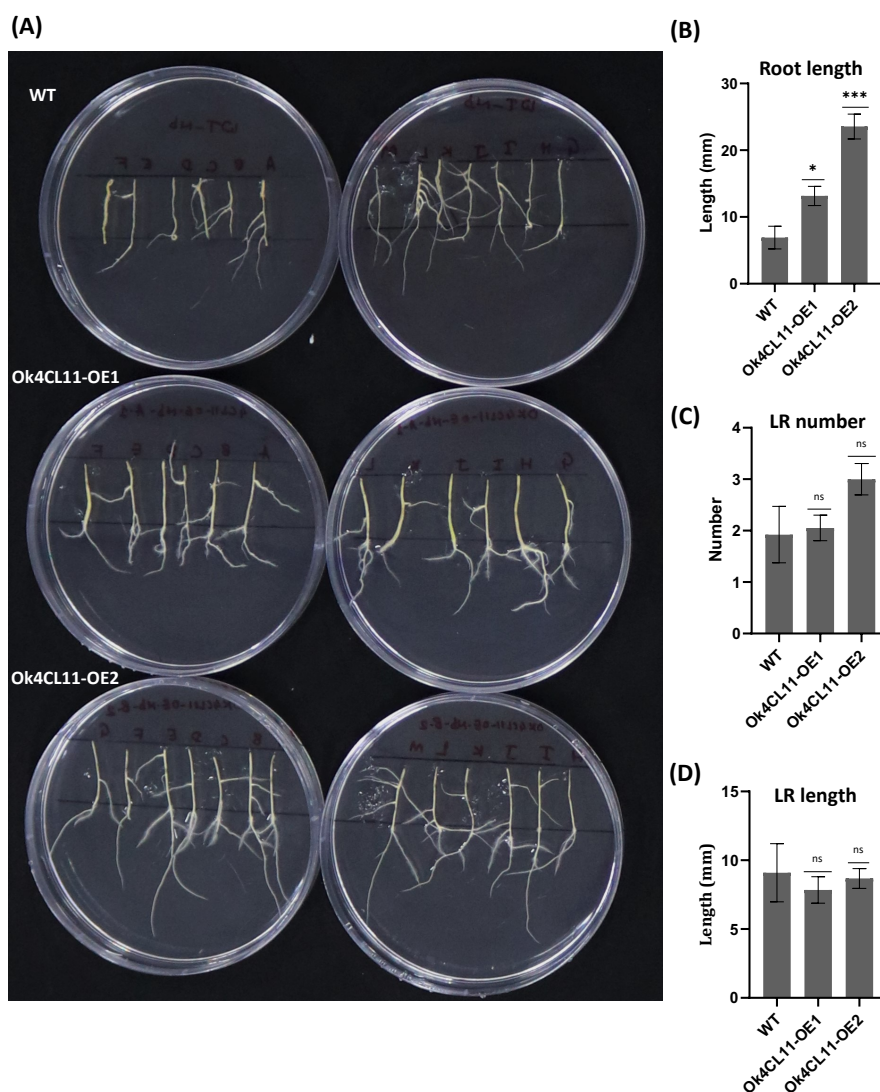


Figure 4.8 (A) Analysis of root apex growth. (B) Root length, (C) lateral root number, and (D) lateral root length. N= number of biological replicates. For WT, n=13; for OE1 line - 1, n= 9; and OE2 line, n=27) were used for measurements of root growth parameters.

4.3.3 Flavonoid and lignin content was altered in leaf and stem tissues of *Ok4CL11* overexpressing *N. benthamiana* lines

The effect of *Ok4CL11* overexpression on the metabolite content was studied using the leaf extracts (in 80% methanol) of transgenic and wild-type *N. benthamiana* plants. As auxin, a phytohormone, is essential for root growth in plants, and we proposed that auxin could be the

potential reason for a reduced root growth phenotype of *Ok4CL11-OE* lines. Earlier reports have suggested that the flavonoids, such as quercetin and kaempferol, can inhibit the polar auxin transport from the shoot-apex to root (Teale et al., 2021). Our preliminary metabolite analysis using LC-MS method showed that the level of flavonoids and their glycosides is increased in *Ok4CL11* OE lines compared to WT-*Nb* plants. (e.g. Apigenein, rutin, 3-Rha-7-Rha-quercetin, etc Figure 4.9A). Level of kaempferol also found to be increased in *4CL11-OE* lines compared to WT plants (**Figure 4.9B**). Interestingly, inactivated form of IAA also found to be accumulated in *Ok4CL11-OE* lines (Indole-3-acetyl-L-aspartic acid). Further, phloroglucinol HCL specifically stains lignin and forms pink colour. Staining the transverse hand sections of WT and *Ok4CL11* overexpressing *N. benthamiana* plants suggested that the lignin content is reduced in overexpression lines compared to wild-type (**Figure 4.9C**). Overall, these results hypothesize that reduced polar auxin transport due to enhanced accumulation of flavonoids could be one of the reasons for the reduced root growth phenotype of *Ok4CL11-OE* lines. Further experiments were focused on studying this hypothesis.

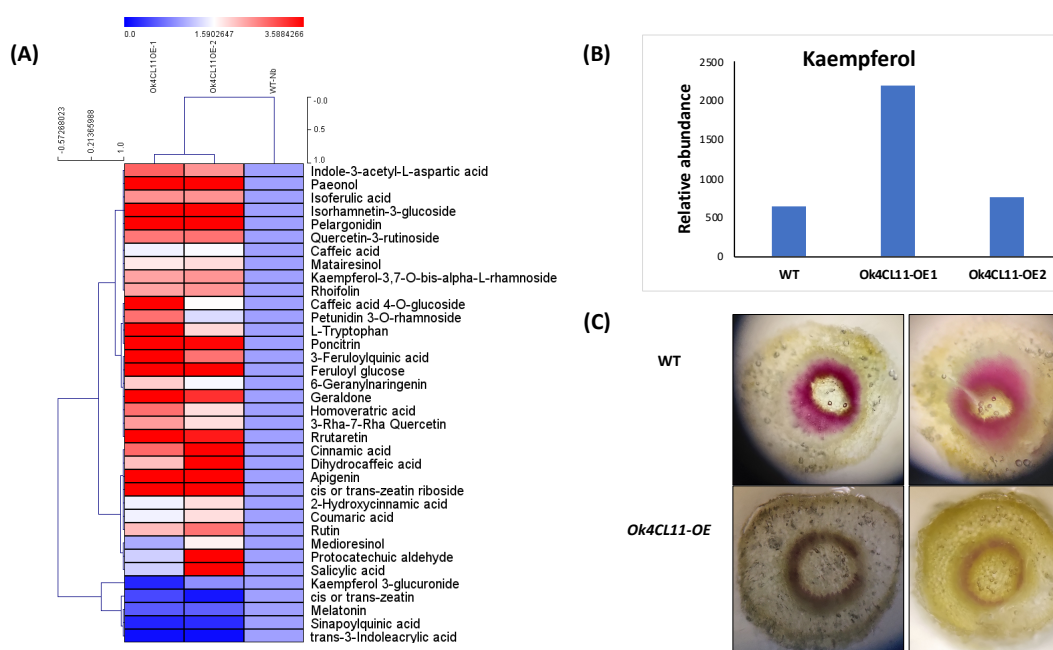


Figure 4.9 (A) Relative fold change in metabolite content in WT and *Ok4CL11* OE lines (B) Relative abundance of kaempferol in *Ok4CL11*-OE lines (OE1 and OE2) and wild-type *N. benthamiana* plants. (C) Lignin accumulation in *Ok4CL11*-OE lines (OE1 and OE2) and wild-type (WT) *N. benthamiana* plants.

4.3.4 ‘Rootless phenotype’ in *Ok4CL11* overexpressing *N. benthamiana* plants was rescued by transforming *Ok4CL11* RNAi construct

Two *Ok4CL11* overexpression *N. benthamiana* lines (OE1 and OE2) were used for transforming the silencing construct of *Ok4CL11* (*Ok4CL11-pRI101* RNAi) using *Agrobacterium*-mediated plant transformation method. Silencing lines are being developed. Currently, few shoots that have emerged from the overexpression *N. benthamiana* lines (OE1 and OE2), following the transformation of *Ok4CL11* RNAi construct, are transferred to rooting medium, and have started to show root formation in *Ok4CL11-OE* lines (**Figure 4.10**). These results suggest that *Ok4CL11-pRI101* RNAi construct could partially rescue the ‘rootless’ phenotype of the *Ok4CL11* overexpression lines (detailed analysis is still pending).

Further, grafting was conducted to understand the mobile nature of the signal (auxin) that is contributing to the ‘rootless’ phenotype. Different grafts (homo-grafts; WT/WT, OE2/OE2) and hetero-grafts; WT/OE2, OE2/WT) were performed using WT and *Ok4CL11-OE2* line. Minimum ten grafts of each type were developed, and all survived on MS medium. These grafts were allowed to grow on solid MS medium for 2 weeks. Our observations suggested that the ‘retarded root growth’ phenotype cannot be rescued by grafting WT on OE2 line (WT/OE2); however, the roots emerged from the WT stock in opposite hetero-grafts (OE2/WT). The root growth in OE2/WT hetero-grafts was not as vigorous as seen for WT/WT homo-grafts (**Figure 4.10**). These results suggest that the local overexpression of *Ok4CL11* might contribute to the affected root growth. Based on these findings, we hypothesize that no or negligible level of auxin in the stem of *Ok4CL11-OE1/2* lines (due to excessive accumulation of kaempferol) and the already present minimum auxin level in the WT stock could be the reasons for the above observed root growth phenotypes.

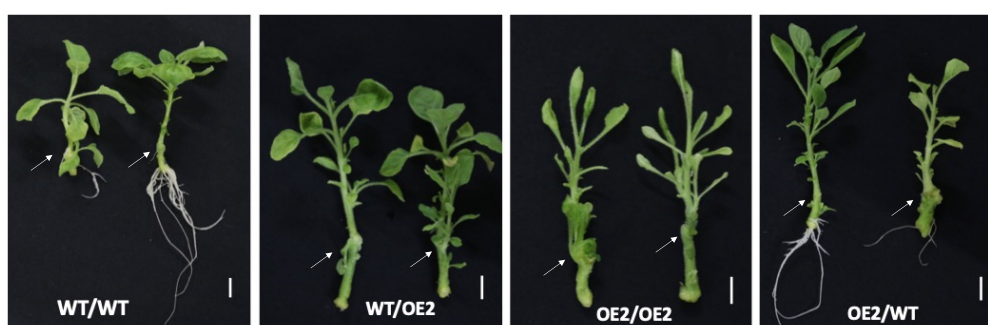


Figure 4.10 Grafting experiment using *Ok4CL11-OE2* line and wild-type (WT) *N. benthamiana* plants. (A) Stock and scion from WT (WT homo-graft). (B) Stock from the

OE2 line and scion from WT *N. benthamiana* plants (WT and OE2 hetero-graft). **(C)** Stock and scion from the OE2 line (OE2 homo-graft). **(D)** Stock from WT and scion from the OE2 line (OE2 and WT hetero-graft). Scale= 1 cm. Arrows indicate graft junctions.

4.3.5 RNA-seq analysis to identify the cause behind ‘retarded root growth’ phenotype in *Ok4CL11-OE* lines

Following three weeks of stem sub-culture, whole plants of the two *Ok4CL11* overexpression lines (OE1 and OE2) and WT were harvested for RNA-seq and metabolite profiling. The main aim of RNA-sequencing analysis is to identify auxin related genes that could be responsible for reduced growth phenotype of overexpression lines. RNA-seq was carried out in triplicates; wherein 2 plants (off 6) were pooled together to form three biological replicates. The concentrations of purified mRNA from each sample were quantified on Qubit. Further, using bioanalyzer, it was observed that all nine samples had RNA integrity (RIN) values greater than 7 confirming the good quality of our RNA samples (**Table 4.2**).

Table 4.2 QC reports for RNA-sequencing samples.

Sr.	Sample Name	Qubit conc. (ng/μl)	Nanodrop conc. (ng/μl)	260/280	RIN Value
1	Wild-type Rep 1	812	751.2	1.98	7.9
2	Wild-type Rep 2	926	961.8	2.00	7.5
3	Wild-type Rep 3	638	650.4	1.99	8.0
4	Ok4CL11-OE2 Rep 1	338	308.6	1.94	8.3
5	Ok4CL11-OE2 Rep 2	576	644.3	1.95	8.0
6	Ok4CL11-OE2 Rep 3	628	583.4	2.00	8.0
7	Ok4CL11-OE1 Rep 1	758	831.3	1.88	7.8
8	Ok4CL11-OE1 Rep 2	298	285.5	1.97	8.2
9	Ok4CL11-OE1 Rep 3	854	932.1	2.03	8.2

The total RNA was quantified on Qubit and RNA integrity was further checked using Bioanalyzer. OE1 and OE2 are overexpression lines. Rep= Biological replicate; RIN= RNA integrity number.

The total number of differentially expressed (DE) genes in comparison to wild-type plants were 11784 and 8269 in Ok4CL11-OE1 and -OE2 lines, respectively (**Fig. 4.11A**). In the Ok4CL11-OE1 line with a rootless phenotype, 5792 DE genes were upregulated, whereas 5992 DE genes were downregulated (**Fig.4.11A**). However, in the Ok4CL11-OE1 line having reduced root growth phenotype 4321 DE genes were upregulated and 3948 DE genes were downregulated (**Fig.4.11A**). To identify the set of common as well as unique genes that were

differentially expressed in both and/or either line, a comparative analysis was performed and represented by Venn diagrams. Analysis revealed about 2746 and 3135 genes were commonly upregulated and downregulated, respectively. Amongst the Ok4CL11-OE lines, OE1 showed 3246 and 2657 DE genes that were specifically up- and downregulated, whereas 1202 and 1186 were uniquely up- and downregulated in the OE2 line, respectively (Fig. 4.11B).

RT-qPCR analysis showed that the majority of the DE genes involved in the phenylpropanoid pathway were downregulated compared to WT (Fig. 4.11C). These genes include *4CL*, *Chalcone synthase (CHS)*, *Chalcone isomerase (CHI)*, *Flavonoid 3'-monooxygenase (F3H)*, *Phenylalanine ammonia-lyase (PAL)*, *Cinnamyl alcohol dehydrogenase (CAD)*, *Cinnamoyl-CoA reductase (CCR)*, *Cinnamate 4-hydroxylase (C4H)*, *Shikimate O-hydroxycinnamoyl transferase (HCT)* (Fig. 4.11C). Moreover, two *glycosyltransferase (GT)* genes encoding the enzymes that possibly convert flavonoids, such as kaempferol and quercetin to their respective glycosylic derivatives were upregulated in both *Ok4CL11-OE* lines (Fig. 4.11C). Additionally, another *GT* and one of the *4CLs* in *N. benthamiana* were also downregulated in both OE lines compared to WT (Fig. 4.11C).

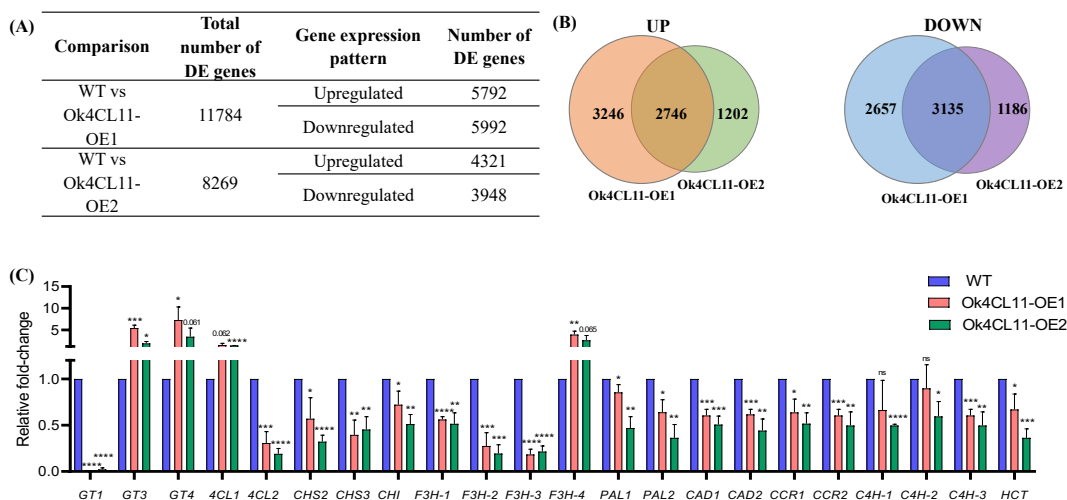


Figure 4.11 RNA-seq and gene expression analysis : (A) Table for DE genes and (B) Venn diagrams for DE genes (C)RT-qPCR for of phenylpropanoid pathway genes. *GT*, *Glycosyl transferase*; *4CL*, *4-Coumarate-CoA ligase*; *CHS*, *Chalcone synthase*; *CHI*, *Chalcone isomerase*; *F3H*, *Flavonoid 3'-monooxygenase*; *PAL*, *Phenylalanine ammonia-lyase*; *CAD*, *Cinnamyl alcohol dehydrogenase*; *CCR*, *Cinnamoyl-CoA reductase*; *C4H*, *Cinnamate 4-hydroxylase*; *HCT*, *Shikimate O-hydroxycinnamoyl transferase*

4.3.6 Preliminary analysis suggest that Ok4CL7 and -15 localize to peroxisomes

The vector control (mCherry:pRI101AN) and the C-terminal fusion constructs of mCherry with Ok4CL7 or -15 in pRI101AN vector were generated individually, and transformed in *A. tumefaciens* (GV3101 strain).

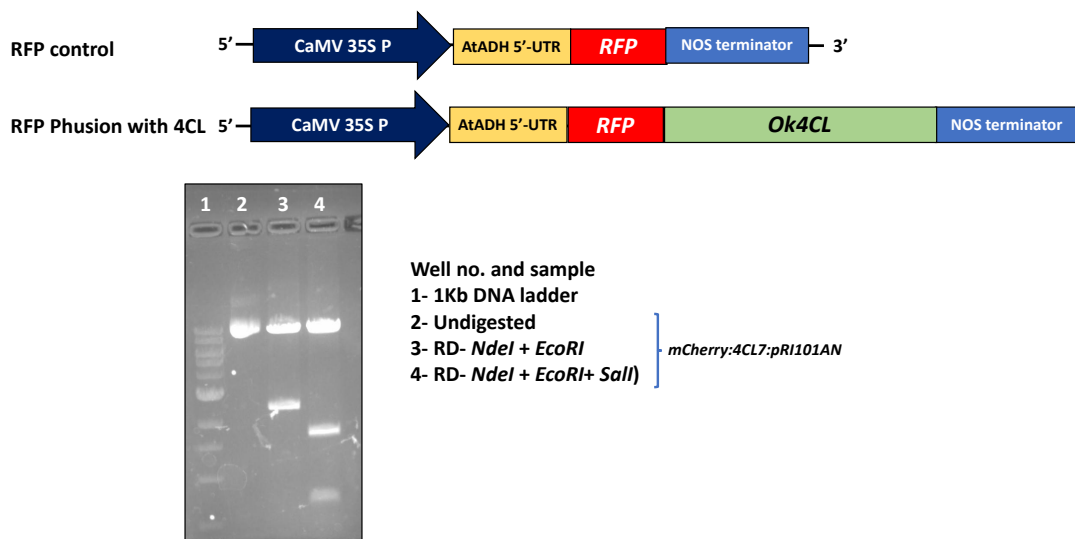


Figure 4.12 (A) Schematic representation of *mCherry*-pRI101AN and *mCherry:Ok4CL*-pRI101AN construct prepared for the localization study. **(B)** Confirmation of *mCherry* and *Ok4CL7* genes in pRI101AN vector by *EcoRI* and *NdeI/Sall* restriction enzyme digestion.

These constructs were transiently expressed by agroinfiltration in leaves of WT *N. benthamiana* (4 week old soil grown plants). Local leaves were taken for microscopic imaging after 3 to 4 days of agroinfiltration. Confocal microscope imaging showed that mCherry localizes to the nucleus and cell membrane, while mCherry fused with Ok4CL7 or -15 showed localization to peroxisomes (**Figure 4.13**). This is a preliminary analysis, and experiments are underway to study the localization of Ok4CL11 as well as the N-terminal mCherry fusion constructs of Ok4CL7, -11 and -15 in leaves of *N. benthamiana* following agroinfiltration.

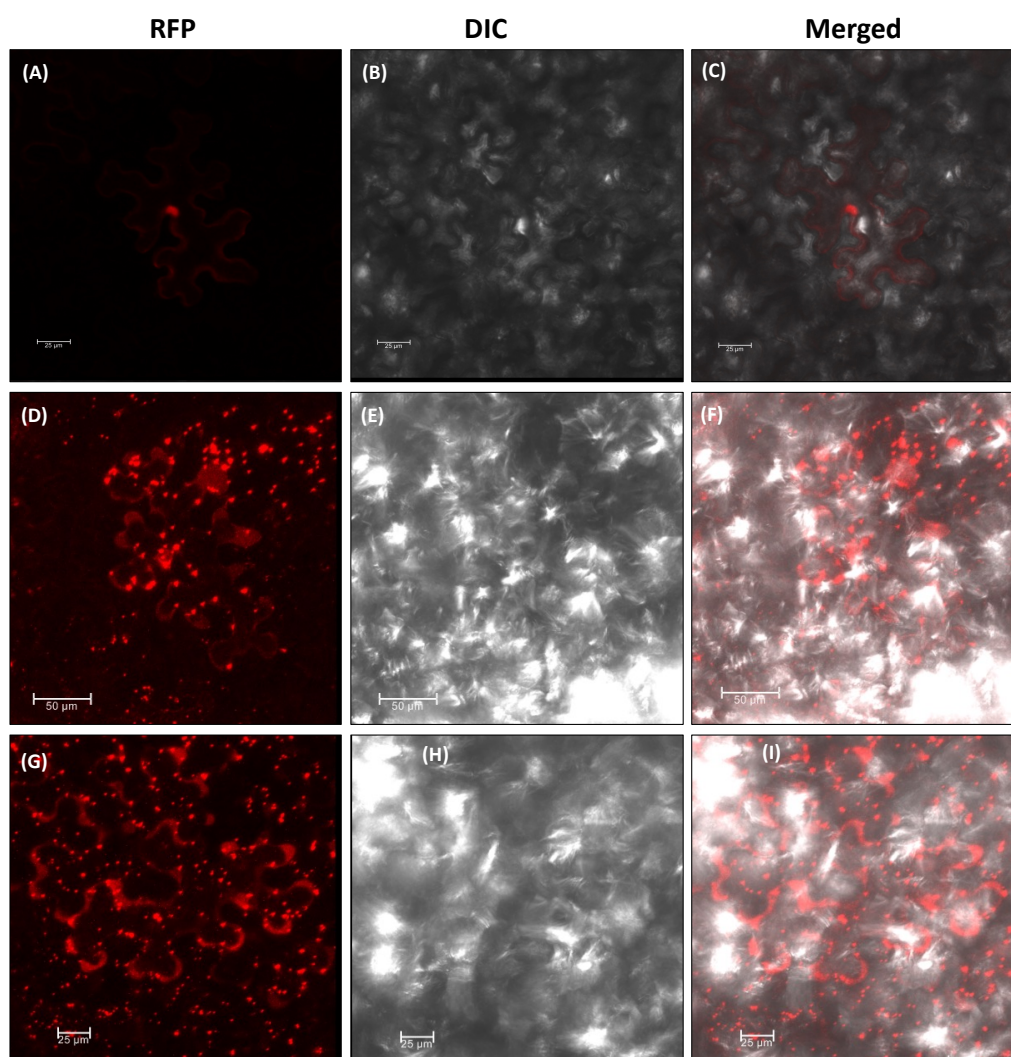


Figure 4.13 Ok4CL7 and -15 protein localization. Localization of mCherry to the nucleus and cell membrane (A-C); Ok4CL7 (D-F); and Ok4CL15 (G-I).

4.3.7 List of other ongoing experiments to validate the function of *Ok4CL11* in producing rootless plants

- Mimicking the ‘retarded root growth’ phenotype of *Ok4CL11-OE* lines in WT *N. benthamiana* plants by independent application of kaempferol (100 μ M) or auxin inhibitors (TIBA [10 μ M] or NPA1 [10 μ M]) under *in vitro* conditions.
- Quantifying the levels of natural auxins (IAA and/or IBA) and kaempferol from three regions of the stem (apical, middle and basal) from *Ok4CL11-OE* lines in comparison to WT *N. benthamiana* plants.

- Auxin accumulation would be assessed by transforming the *DR5prom:GUS* fusion construct in WT and *Ok4CL11-OE* lines, followed by GUS assay and MUG activity measurements using whole *in vitro* grown plants. We hypothesize that these assays would demonstrate a reduced GUS expression in the stem and the highest expression in the shoot apex, young leaves and root tips of *Ok4CL11-OE* lines compared to the respective WT plants (control).
- Histology of leaf and stem tissues to observe if the vascular bundles as well as the lignin content in the tissues of *Ok4CL11-OE* lines compared to WT.
- ‘Retarded root growth’ phenotype of *Ok4CL11* is also being tested in Arabidopsis (*Col-0*) and potato (*Solanum tuberosum* cv. Désirée).

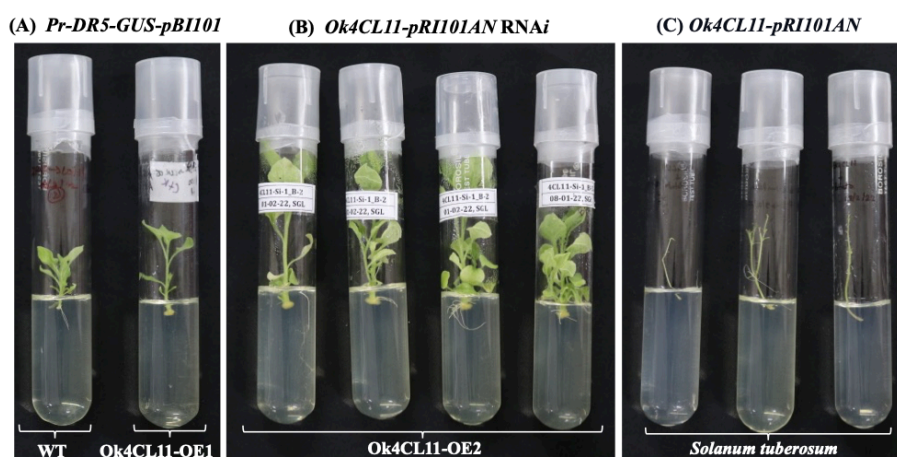


Figure 4.14 Phenotypes observed in ongoing experiments. **(A)** *DR5prom:GUS-pBI101* expressed in wild *N. benthamiana* and *Ok4CL11* overexpressing *N. benthamiana* (*Ok4CL11-OE1-pRI101AN*). **(B)** Root growth was observed in *Ok4CL11-OE2-pRI101AN* upon expression of *Ok4CL11-pRI101AN* RNAi construct. **(C)** Potato shoots regenerated from agro-infected leaf tissues represent putative transformants of *Ok4CL11-OE* in *Solanum tuberosum* cv. Désirée.

4.4 Summary

Products of 4CL enzyme are utilized for biosynthesis of structural and non-structural phenylpropanoid. Several reports suggest that specific 4CL isoforms are involved in the diversion of flux toward structural phenylpropanoid lignin or other non-structural phenylpropanoids like flavonoids (Lee Diana et al. 1997; Gui et al. 2011; Chen et al. 2014; Liu et al. 2017; Baldi et al. 2017). In this chapter, we have overexpressed three *Ocimum Ok4CL* isoforms (*Ok4CL7*, *-11* and *-15*) in *N. benthamiana* and investigated their effects on the plant phenotypes, gene expression and metabolite analyses.

The overall plant phenotypes of *N. benthamiana* lines overexpressing *Ok4CL7* and *Ok4CL15* were comparable to WT plants, except *Ok4CL15* overexpression line having increased number of roots. Interestingly, *Ok4CL11* overexpression *N. benthamiana* lines (representative lines OE1 and OE2) showed ‘rootless’ or ‘retarded root growth’ phenotype. The root growth is influenced by numerous endogenous and exogenous factors. Several molecules, including hormones (auxin), flavonoids, lignin, etc. also play a crucial role in regulating primary and lateral root growth. Plant specialized metabolites like flavonoids are crucial not only for overall plant growth and development, but they also provide protection against various biotic and abiotic stresses (Shah and Smith 2020). *Ok4CL11* overexpression representative lines (OE1 and OE2) have elevated levels of kaempferol compared to WT *N. benthamiana* plants, suggesting that *Ok4CL11* is involved in the biosynthesis of flavonoids. Flavonoids, such as kaempferol, apigenin and quercetin are known for their ability to inhibit polar auxin transport by inhibiting PIN transporters (Jacobs and Rubery 1988; Peer et al. 2004). Kaempferol interact with PIN transporters and produces dimer, same as known for auxin transport inhibitor 1-naphthylphthalamic acid (NPA) (Teale et al. 2021). Overexpression lines (OE1 and OE2) have 80% fewer root numbers, 50% fewer lateral roots, and root length was reduced by 50%. These results are consistent with the recent report by Chapman and Muday (2021). The mutants of the genes involved in the biosynthesis pathway of flavonoids (kaempferol and quercetin), such as *chalcone synthase (CHS)*, *Flavonoid 3'-Hydroxylase (F3'H)* and *flavonol synthase1 (FLS1)* produced negligible levels of flavonoids and displayed enhancement in number of lateral root primordia and also the emergence of lateral roots in *Arabidopsis*. Further, authors demonstrated that kaempferol or downstream derivatives function as a negative regulator of lateral root emergence (Chapman and Muday 2021). ROS plays a vital role in the plant development process, including root hair elongation, root gravitropism, root tip elongation, elongation in the root tip and lateral root emergence

(Manzano et al. 2014; Krieger et al. 2016; Orman-Ligeza et al. 2016; Gayomba and Muday 2020). Accumulation of kaempferol leads to reduced ROS production, which can lead to inhibition of lateral root emergence (Chapman and Muday 2021). Based on these points, it appears that the 'retarded root growth' phenotype in *Ok4CL11* overexpression lines could be because of elevated levels of kaempferol and reduced polar auxin transport. The ongoing experiments will further ascertain these claims.

Chapter 5

Summary and future prospects

5.1 Summary

4CL is one of the important enzymes from the phenylpropanoid biosynthesis pathway. The 4CL is essentially involved in channelizing the precursors for different pathways of phenylpropanoid biosynthesis. 4CLs are explored from several plants for different aspects, including their functional roles in growth and development, industrial applications in biofuel production, metabolic pathway flux understanding and the management of stress response of plants. The focus of this work was to functionally characterize three isoforms of 4CL from *O. kilimandscharicum* (Ok4CL7, -11 and -15). We have characterized the functions of these Ok4CLs using *in vitro*, *in silico* and *in vivo* approaches.

In Chapter 2, two Ok4CL isoforms (Ok4CL7 and -15) were functionally characterized *in vitro* from *O. kilimandscharicum* for the first time. My several attempts failed obtain recombinant protein of Ok4CL11 in *E. coli*. Nevertheless, all three isoforms of Ok4CLs were taken forward for further experiments. Gene expression profiles of *Ok4CL7* and *-15* were high in trichomes and roots compared to other tissue types tested. Optimum pH required for recombinant Ok4CL7 and Ok4CL15 activities was found to be 8 and 7, respectively. Both 4CLs showed maximum activities at 40 °C. Ok4CL7 catalyzed the conversion of *p*-coumaric acid, ferulic acid and caffeic acid into their corresponding CoA esters. Furthermore, Ok4CL15 could catalyze the conversion of *p*-coumaric acid, ferulic acid and sinapic acid into respective CoA esters. However, both 4CLs (Ok4CL7 and -15) were unable to utilize cinnamic acid. *In silico* data indicated that, specific non-covalent interactions between the substrate and binding pockets of Ok4CLs are essential for their catalytic activities. It appears that Ok4CL15 could be involved in lignin biosynthesis due to its high expression in root tissues and catalytic activity to convert sinapic acid to sinapoyl CoA. The results of this study allow us to understand the function of 4CL isoforms of *O. kilimandscharicum* and its potential role in the diversification of the phenylpropanoid pathway.

In Chapter 3, *in silico* analysis was performed to decipher the reason behind the differential substrate selectivity. Molecular modeling and docking, followed by *in silico* study of enzyme-substrates interaction, provided information about the residue contacts with the substrate in the binding pocket. Feruloyl adenylate showed a higher number of contacts and lower binding energy with Ok4CL7 and -15 than cinnamoyl adenylate. Differential substrate selection or affinity towards a particular substrate by Ok4CLs may contribute to regulation flux in the phenylpropanoid pathway. *p*-coumaric acid and ferulic were used by both of them, whereas caffeic acid is exclusively used by Ok4CL7 and sinapic acid by Ok4CL15. Important

finding of this study is that Ok4CL7 and -15 have differential selectivity for sinapic acid, different from reported 4CLs from *G. max* and *N. tabacum*. The comparison of 4CL sequences suggests that other than the position of V and L, some other residues are also responsible for the substrate selectivity. Models were developed by replacing interacting amino acids with other standard amino acids to check their importance in interaction with the substrate. In Ok4CL7, the replacement of lysine at 477 position with either leucine, isoleucine, or valine leads to an increase in affinity for sinapic acid. In the case of Ok4CL15, replacing lysine at 525 with isoleucine leads to an increase in affinity for sinapic acid. Experimental validation of these results is in progress by developing mutant 4CL enzymes and their kinetics study with different substrates.

In Chapter 4, we studied the *in planta* functional characterization of three Ok4CL isoforms (Ok4CL7, -11 and -15). Due to a poor regeneration efficiency and limited success of *Agrobacterium*-mediated stable transformation methods for several members of *Ocimum* species, we have overexpressed three Ok4CL isoforms (Ok4CL7, -11 and -15) in *N. benthamiana* and studied their effects on the plant phenotypes and metabolite profiling. Interestingly, we observed a ‘rootless or ‘retarded root growth’ phenotype in the *Ok4CL11* overexpression lines of *N. benthamiana*, whereas no such phenotype was observed in *Ok4CL7* and *-15* overexpression lines. Plants overexpressing *Ok4CL7* have normal root growth phenotype, whereas those overexpressing *Ok4CL15* have higher number of roots. Considering the role of 4CLs in the phenylpropanoid biosynthesis pathway, we also carried out metabolite profiling of the *Ok4CL11* overexpression lines and wild-type *N. benthamiana* plants. To dissect the ‘rootless’ or ‘retarded root growth’ phenotype observed in the *Ok4CL11* overexpression lines, a number of approaches, such as transcriptomics, auxin and kaempferol quantification, grafting (*in vitro*), chemical treatment, phenotype rescue by RNAi, etc. were undertaken. Some of these objectives are near completion, whereas majority of them are currently ongoing. Metabolite analysis suggested that overexpression of *Ok4CL11* leads to increase in kaempferol accumulation. Kaempferol is known for inhibiting polar auxin transport (PAT) by interaction with PIN auxin transporters. Plant specialized metabolites like flavonoids are crucial not only for overall plant growth and development, but they also provide protection against various biotic and abiotic stresses (Shah and Smith 2020). Preliminary analysis of lignin staining by phloroglucinol suggests lignin content increased in *Ok4CL11*-OE lines compared to WT. Based on these points, it appears that the ‘retarded root growth’ phenotype in *Ok4CL11*

overexpression lines could be because of elevated levels of kaempferol and a reduced polar auxin transport. The ongoing experiments will further ascertain these claims.

Overall, in this work, we have functionally characterized two 4CL isoforms (Ok4CL7 and -15) using *in-vitro* approaches, while three isoforms of Ok4CLs (Ok4CL7, -11 and -15) were using *in planta* characterization. This work will provide insights for exploring the potential use of 4CL for metabolic pathway engineering and various other applications.

5.2 Future prospects

The metabolic pathways are regulated at entry point, branch point and/or in some cases by intermediate enzymes (rate limiting enzymes) for the final product biosynthesis. However, terminal enzymes are well studied for most of the plant specialized metabolite biosynthesis. The properties of specific Ok4CL isoforms could be further explored for application in different fields. In biofuel industries, high saccharification of biomass is preferred to increase the yield. Lignin is the complex biopolymer, which affects the saccharification. Using specific 4CL isoform having potential to decrease lignin content, it is possible to improve biofuel production efficiency. Overexpression of *Ok4CL11* leads to increase in flavonoid content, which means it is diverting phenolics flux towards the flavonoid biosynthesis and leads to decline in the lignin content of plants. Ok4CL11 isoform could be explored for biofuel production by overexpressing it in other important crops.

4CLs are also successfully used for synthesis of commercially important compounds like curcumin, resveratrol, trihydroxycinnamoyl spermidines, etc. In most cases 4CL was expressed in fusion with other enzymes. It is also possible to enhance the level of important metabolites by overexpressing a particular Ok4CL isoform. In future, 4CL isoforms from *O. kilimandscharicum* could be explored for plant natural product biosynthesis.

4CL isoforms are also known for their role in protection from various biotic and abiotic stresses (reviewed by Lavhale et al., 2018). Further, Ok4CL isoforms could be explored for their application in development of crops with resistant against biotic or abiotic stress. The site-directed mutagenesis approach used in this work could also be used to engineer 4CL with specific substrate preferences. Using this approach, the efficiency of 4CL could also be improved for the above mentioned applications.

5.3 Open Questions

As a consequence of this study there several important open question are need to be solved.

Few of the important questions listed below-

- i) Products of 4CL enzyme could be utilized for both lignin and flavonoid biosynthesis. How overexpression of *Ok4CL11* in *N. benthamiana* contributed to increased accumulation of flavonoid content?
- ii) What is the effect of phenylpropanoid diversion towards flavonoid by 4CL overexpression on detiled morphology, defense, stress response and ecological the consequences?
- iii) Why multiple isoforms present for single enzyme? What is the role of evolution and how its evolved? (Detailed analysis of factors contributing to origin of isoforms such as gene duplication, neofunctionalization, alternative splicing, transposable elements, etc)
- iv) Which are the regulatory enzyme of the flavonoid and lignin biosynthesis? (rate limiting step, committed step, and other mechanisms of pathway regulation)
- v) How the expression of 4CL isoforms is regulated at tissue specific level?
- vi) How plant maintain optimum level phenylpropanoids and how it modulates in response to environmental cues or biotic and abiotic stresses (storage, activation inactivation, degradion, etc)?
- vii) What are the potential application of rootles phenotype?
- viii) What are the active site residues or domains responsible for specific substrate preference? This enzyme could be used or improved for efficient production of important compounds such as penicillin V, resveratrol, curcumin, etc.

Bibliography

- An CH, Lee KW, Lee SH, et al (2015) Heterologous expression of *IbMYB1a* by different promoters exhibits different patterns of anthocyanin accumulation in tobacco. *Plant Physiology and Biochemistry* 89:1–10. <https://doi.org/10.1016/j.plaphy.2015.02.002>
- An H, Zhang J, Xu F, et al (2020) Transcriptomic profiling and discovery of key genes involved in adventitious root formation from green cuttings of highbush blueberry (*Vaccinium corymbosum* L.). *BMC Plant Biology* 20:1–14. <https://doi.org/10.1186/S12870-020-02398-0/FIGURES/6>
- Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Abdollahi Mandoulakani B, Eyvazpour E, Ghadimzadeh M (2017) The effect of drought stress on the expression of key genes involved in the biosynthesis of phenylpropanoids and essential oil components in basil (*Ocimum basilicum* L.). *Phytochemistry* 139:1–7. <https://doi.org/10.1016/j.phytochem.2017.03.006>
- Allina SM, Pri-Hadash A, Theilmann DA, et al (1998) 4-Coumarate-coenzyme A ligase in hybrid poplar properties of native enzyme, cDNA cloning, and analysis of recombinant enzyme. *Plant Physiol* 116:743–754. <https://doi.org/10.1104/pp.116.2.743>
- Amalraj A, Pius A, Gopi S, Gopi S (2017) Biological activities of curcuminoids, other biomolecules from turmeric and their derivatives—a review. *J Tradit Complement Med* 7:205–233. <https://doi.org/10.1016/j.jtcme.2016.05.005>
- An CH, Lee KW, Lee SH, et al (2015) Heterologous expression of *IbMYB1a* by different promoters exhibits different patterns of anthocyanin accumulation in tobacco. *Plant Physiol Biochem* 89:1–10. <https://doi.org/10.1016/j.plaphy.2015.02.002>
- Anand A, Jayaramaiah RH, Beedkar SD, et al (2016) Comparative functional characterization of eugenol synthase from four different *Ocimum* species: implications on eugenol accumulation. *Biochim Biophys Acta Proteins Proteom* 1864:1539–1547. <https://doi.org/10.1016/j.bbapap.2016.08.004>
- Baldi P, Moser M, Brillì M, et al (2017) Fine-tuning of the flavonoid and monolignol pathways during apple early fruit development. *Planta* 245:1021–1035. <https://doi.org/10.1007/s00425-017-2660-5>

- Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *Journal of Biological Chemistry* 266:8551–8559. [https://doi.org/10.1016/S0021-9258\(18\)93010-3](https://doi.org/10.1016/S0021-9258(18)93010-3)
- Bennett RN, Wallsgrove RM (1994) Secondary metabolites in plant defence mechanisms. *New Phytol.* 1994 Aug;127(4):617-633. doi: 10.1111/j.1469-8137.1994.tb02968.x.
- BERG P (1956) Acyl adenylates; the synthesis and properties of adenylyl acetate. *The Journal of biological chemistry* 222:1015–1023. [https://doi.org/10.1016/s0021-9258\(20\)89958-x](https://doi.org/10.1016/s0021-9258(20)89958-x)
- Blanco-Ulate B, Hopfer H, Figueroa-Balderas R, et al (2017) Red blotch disease alters grape berry development and metabolism by interfering with the transcriptional and hormonal regulation of ripening. *J Exp Bot* 68:1225–1238. <https://doi.org/10.1093/jxb/erw506>
- Boerjan W, Ralph J, Baucher M (2003) Lignin Biosynthesis. *Annu Rev Plant Biol.* 2003;54:519-46. doi: 10.1146/annurev.arplant.54.031902.134938.
- Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *Proc Natl Acad Sci USA* 96:12929–12934. <https://doi.org/10.1073/pnas.96.22.12929>
- Cao Y, Han Y, Li D, et al (2016) Systematic analysis of the 4-Coumarate: coenzyme A ligase (4CL) related genes and expression profiling during Fruit development in the Chinese Pear. *Genes* 7:89. <https://doi.org/10.3390/genes7100089>
- CH Wang JYYC (2016) Characterization and functional analysis of 4-Coumarate-CoA Ligase genes in Mulberry. *PLoS One* 11:e0155814. <https://doi.org/10.1371/journal.pone.0155814.g001>
- Chandran D, Sharopova N, Ivashuta S, et al (2008) Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in *Medicago truncatula*. *Planta* 228:151–166. <https://doi.org/10.1007/s00425-008-0726-0>
- Chapman JM, Muday GK (2021) Flavonols modulate lateral root emergence by scavenging reactive oxygen species in *Arabidopsis thaliana*. *J Biol Chem.* 2021 Jan-Jun;296:100222. doi: 10.1074/jbc.RA120.014543. Epub 2020 Dec 25.
- Chemler JA, Koffas MA (2008) Metabolic engineering for plant natural product biosynthesis in microbes. *Curr Opin Biotechnol* 19:597–605. <https://doi.org/10.1016/j.copbio.2008.10.011>

- Chen HC, Song J, Wang JP, et al (2014a) Systems biology of lignin biosynthesis in *Populus trichocarpa*: heteromeric 4-coumaric acid: Coenzyme A ligase protein complex formation, regulation, and numerical modeling. *Plant Cell* 26:876–893. <https://doi.org/10.1105/tpc.113.119685>
- Chen HY, Babst BA, Nyamdari B, et al (2014b) Ectopic expression of a Loblolly Pine class II 4-Coumarate:CoA ligase alters soluble phenylpropanoid metabolism but not lignin biosynthesis in *Populus*. *Plant Cell Physiol* 55:1669–1678. <https://doi.org/10.1093/pcp/pcu098>
- Chen X, Wang H, Li X, et al (2019) Molecular cloning and functional analysis of 4-Coumarate:CoA ligase 4(4CL-like 1) from *Fraxinus mandshurica* and its role in abiotic stress tolerance and cell wall synthesis. *BMC Plant Biol.* 2019 Jun 3;19(1):231. doi: 10.1186/s12870-019-1812-0
- Citti C, Battisti UM, Braghiroli D, et al (2017) A metabolomic approach applied to a liquid chromatography coupled to high-resolution tandem mass spectrometry method (HPLC–ESI–HRMS/MS): towards the comprehensive evaluation of the chemical composition of cannabis medicinal extracts. *Phytochem Anal* 29:144–155. <https://doi.org/10.1002/pca.2722>
- Costa MA, Bedgar DL, Moinuddin SGA, et al (2005) Characterization in vitro and in vivo of the putative multigene 4-coumarate: CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* 66:2072–2091. <https://doi.org/10.1016/j.phytochem.2005.06.022>
- Cukovic D, Ehrling J, VanZiffle JA, Douglas CJ (2001) Structure and evolution of 4-Coumarate: coenzyme A ligase (4CL) gene families. *Biol Chem* 382:645–654. <https://doi.org/10.1515/bc.2001.076>
- Dastmalchi M, Dhaubhadel S (2015) Proteomic insights into synthesis of isoflavonoids in soybean seeds. *PROTEOMICS* 15:1646–1657. <https://doi.org/10.1002/pmic.201400444>
- Deng Y, Lu S (2017) Biosynthesis and Regulation of Phenylpropanoids in Plants. *Critical Reviews in Plant Sciences* 36:257–290. <https://doi.org/10.1080/07352689.2017.1402852>
- Di P, Hu Y, Xuan H, et al (2012) Characterization and the expression profile of 4-coumarate: CoA ligase (*Ii4CL*) from hairy roots of *Isatis indigotica*. *African Journal of Pharmacy and Pharmacology* 6:2166–2175. <https://doi.org/10.5897/AJPP12.852>
- Dixon RA, Paiva NL (1995) Stress-Induced Phenylpropanoid Metabolism. *Plant Cell*. 1995 Jul;7(7):1085-1097. doi: 10.1105/tpc.7.7.1085.

- Douglas C, Hoffmann H, Schulz W, Hahlbrock K (1987) Structure and elicitor or uv-light-stimulated expression of two 4-coumarate-CoA ligase genes in parsley. *EMBO J* 6:1189–1195. <https://doi.org/10.1002/j.1460-2075.1987.tb02353.x>
- Douglas CJ (1996) Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. *Trends Plant Sci* 1:171–178. [https://doi.org/10.1016/1360-1385\(96\)10019-4](https://doi.org/10.1016/1360-1385(96)10019-4)
- Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Ehltling J, Büttner D, Wang Q, et al (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J* 19:9–20. <https://doi.org/10.1046/j.1365-313x.1999.00491.x>
- Emran M, Chowdhury K, Choi B, et al (2013) Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. 6(4):254-262
- Ehltling J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *Plant J* 25:455–465. <https://doi.org/10.1046/j.1365-313x.2001.01122.x>
- Ferrieri AP, Arce CCM, Machado RAR, et al (2015) A *Nicotiana attenuata* cell wall invertase inhibitor (NaCWII) reduces growth and increases secondary metabolite biosynthesis in herbivore-attacked plants. *New Phytologist* 208:519–530. <https://doi.org/10.1111/nph.13475>
- Floros DJ, Petras D, Kapon CA, et al (2017) Mass spectrometry based molecular 3D-cartography of plant metabolites. *Front Plant Sci.* 2017 Mar 29;8:429. <https://doi.org/10.3389/fpls.2017.00429>
- Fraser CM, Chapple C (2011) The Phenylpropanoid Pathway in Arabidopsis. *The Arabidopsis Book* 9:e0152. <https://doi.org/10.1199/tab.0152>
- Freund DM, Hegeman AD (2017) Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. *Current Opinion in Biotechnology* 43:41–48
- Fritzemeier K-H, Cretin C, Kombrink E, et al (1987) Transient induction of phenylalanine ammonia-lyase and 4-Coumarate:CoA ligase mRNAs in Potato leaves infected with virulent or avirulent races of *Phytophthora infestans*. *Plant Physiol* 85:34–41. <https://doi.org/10.1104/pp.85.1.34>

- Fujino N, Tenma N, Waki T, et al (2018) Physical interactions among flavonoid enzymes in snapdragon and torenia reveal the diversity in the flavonoid metabolon organization of different plant species. *Plant J* 94:372–392. <https://doi.org/10.1111/tpj.13864>
- Fukushima A, Kusano M, Redestig H, et al (2009) Integrated omics approaches in plant systems biology. *Current Opinion in Chemical Biology* 13:532–538. <https://doi.org/10.1016/J.CBPA.2009.09.022>
- Fulda M, Heinz E, Wolter FP (1994) The fadD gene of Escherichia coli K12 is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family. *Mol Gen Genet* 242:241–249. <https://doi.org/10.1007/bf00280412>
- Gaid MM, Scharnhop H, Ramadan H, et al (2011) 4-Coumarate:CoA ligase family members from elicitor-treated *Sorbus aucuparia* cell cultures. *J Plant Physiol* 168:944–951. <https://doi.org/10.1016/j.jplph.2010.11.021>
- Gao S, Yu HN, Xu RX, et al (2015) Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. *Phytochemistry* 111:48–58. <https://doi.org/10.1016/j.phytochem.2014.12.017>
- Gayomba SR, Muday GK (2020) Flavonols regulate root hair development by modulating accumulation of reactive oxygen species in the root epidermis. *Development*. 2020 Apr 27;147(8):dev185819. doi: 10.1242/dev.185819
- Giovinazzo G, Ingrosso I, Paradiso A, et al (2012) Resveratrol biosynthesis: plant metabolic engineering for nutritional improvement of food. *Plant Foods Hum Nutr* 67:191–199. <https://doi.org/10.1007/s11130-012-0299-8>
- Goufo P, Moutinho-Pereira JM, Jorge TF, et al (2017) Cowpea (*Vigna unguiculata* L. Walp.) metabolomics: Osmoprotection as a physiological strategy for drought stress resistance and improved yield. *Frontiers in Plant Science* 8:. <https://doi.org/10.3389/fpls.2017.00586>
- Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652. <https://doi.org/10.1038/nbt.1883>

- Gui J, Shen J, Li L (2011) Functional characterization of evolutionarily divergent 4-Coumarate: coenzyme A ligases in Rice. *Plant Physiol* 157:574–586. <https://doi.org/10.1104/pp.111.178301>
- Gurav TP, Dholakia BB, Giri AP (2021) A glance at the chemodiversity of *Ocimum* species: Trends, implications, and strategies for the quality and yield improvement of essential oil. *Phytochemistry Reviews* 2021 1–35. <https://doi.org/10.1007/S11101-021-09767-Z>
- H Jin ECPB (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J* 19:6150–6161. <https://doi.org/10.1093/emboj/19.22.6150>
- Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. *Ann Rev Plant Physiol* 40:347–369. <https://doi.org/10.1146/annurev.pp.40.060189.002023>
- Hamada K, Nishida T, Yamauchi K, et al (2004) 4-Coumarate: coenzyme A ligase in black locust (*Robinia pseudoacacia*) catalyses the conversion of sinapate to sinapoyl-CoA. *J Plant Res* 117:303–310. <https://doi.org/10.1007/s10265-004-0159-1>
- Hamberger B, Hahlbrock K (2004) The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proceedings of the National Academy of Sciences of the United States of America* 101:2209–2214. <https://doi.org/10.1073/pnas.0307307101>
- Hamberger B, Hahlbrock K (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate: coenzyme A ligase (4CL) genes in Aspen (*Populus tremuloides*). *Proc Natl Acad Sci USA* 95:5407–5412. <https://doi.org/10.1073/pnas.0307307101>
- Hassan S, Mathesius U (2012) The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *J Exp Bot.* 2012 May;63(9):3429-44. doi: 10.1093/jxb/err430
- Herrmann KM, Weaver LM (1999) The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol* 50:473–503. <https://doi.org/10.1146/annurev.arplant.50.1.473>
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7:1071–1083. <https://doi.org/10.1105/tpc.7.7.1071>

- Hrazdina G, Wagner GJ (1985) Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Arch Biochem Biophys* 237:88–100. [https://doi.org/10.1016/0003-9861\(85\)90257-7](https://doi.org/10.1016/0003-9861(85)90257-7)
- Hu WJ, Harding SA, Lung J, et al (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* 17:808–812. <https://doi.org/10.1038/11758>
- Hu W-J, Kawaoka A, Tsai C-J, et al (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*) *Proc Natl Acad Sci U S A*. 1998 Apr 28;95(9):5407-12. doi: 10.1073/pnas.95.9.5407.
- Hu Y, Wang DC, Gai Y, et al (2010) Crystal structures of a *Populus tomentosa* 4-coumarate: CoA ligase shed light on its enzymatic mechanisms. *Plant Cell* 22:3093–3104. <https://doi.org/10.1105/tpc.109.072652>
- Hue HTT, Ha DTT, Hai N van, Hien LTT (2016) Isolation and characterization of the 4-coumarate:coenzyme A ligase (4CL1) promoter from *Eucalyptus camaldulensis*. *Physiol Mol Biol Plants* 22:399–405. <https://doi.org/10.1007/s12298-016-0369-8>
- Jacobs M, Rubery PH (1988) Naturally occurring auxin transport regulators. *Science (New York, NY)* 241:346–349. <https://doi.org/10.1126/SCIENCE.241.4863.346>
- Jin Z, Wungsintaweekul J, Kim SH, et al (2020) 4-Coumarate:coenzyme A ligase isoform 3 from *Piper nigrum* (Pn4CL3) catalyzes the CoA thioester formation of 3,4-methylenedioxycinnamic and piperic acids. *Biochemical Journal* 477:61–74. <https://doi.org/10.1042/BCJ20190527>
- Jung JH, Kannan B, Dermawan H, et al (2016) Precision breeding for RNAi suppression of a major 4-coumarate:coenzyme A ligase gene improve cell wall saccharification from field grown sugarcane. *Plant Mol Biol* 92:505–517. <https://doi.org/10.1007/s11103-016-0527-y>
- Katsuyama Y, Matsuzawa M, Funa N, Horinouchi S (2008) Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway. *Microbiology* 154:2620–2628. <https://doi.org/10.1099/mic.0.2008/018721-0>

- Kim YB, Kim JK, Uddin MR, et al (2013) Metabolomics analysis and biosynthesis of Rosmarinic acid in *Agastache rugosa* Kuntze treated with Methyl Jasmonate. PLoS One 8:e64199. <https://doi.org/10.1371/journal.pone.0064199>
- Kollmann FFP, Côté WA, Côté WA (1968) The Structure of Wood and the Wood Cell Wall. In: Principles of Wood Science and Technology. Springer Berlin Heidelberg, pp 1–54
- Knobloch KH, Hahlbrock K (1977) 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Partial purification, substrate specificity, and further properties. Archives of Biochemistry and Biophysics 184:237–248. [https://doi.org/10.1016/0003-9861\(77\)90347-2](https://doi.org/10.1016/0003-9861(77)90347-2)
- Korkina LG (2007) Phenylpropanoids as naturally occurring antioxidants: From plant defense to human health. Cell Mol Biol (Noisy-le-grand). 2007 Apr 15;53(1):15-25
- Krieger G, Shkolnik D, Miller G, Fromm H (2016) Reactive oxygen species tune root tropic responses. Plant Physiol. 172(2):1209-1220. doi: 10.1104/pp.16.00660
- Kueger S, Steinhauser D, Willmitzer L, Giavalisco P (2012) High-resolution plant metabolomics: From mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. Plant J. 2012 Apr;70(1):39-50. doi: 10.1111/j.1365-313X.2012.04902.x
- Kumar A, Ellis BE (2003) 4-Coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. Plant Molecular Biology 51:327–340. <https://doi.org/10.1023/A:1022004923982>
- Lavhale SG, Joshi RS, Kumar Y, Giri AP (2021) Functional insights into two *Ocimum kilimandscharicum* 4-coumarate-CoA ligases involved in phenylpropanoid biosynthesis. International journal of biological macromolecules 181:202–210. <https://doi.org/10.1016/J.IJBIOMAC.2021.03.129>
- Lavhale SG, Kalunke RM, Giri AP (2018) Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. Planta 248:1063–1078. doi: 10.1007/s00425-018-2965-z.
- Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiology 112:193–205. <https://doi.org/10.1104/pp.112.1.193>

- Lee Diana, Meyer K, Chapple G, Douglas CJ (1997) Antisense suppression of 4-coumarate:coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. *Plant Cell* 9:1985–1998. <https://doi.org/10.1105/tpc.9.11.1985>
- Li SS, Chang Y, Li B, et al (2020) Functional analysis of 4-coumarate: CoA ligase from *Dryopteris fragrans* in transgenic tobacco enhances lignin and flavonoids. *Genetics and Molecular Biology* 43:. <https://doi.org/10.1590/1678-4685-GMB-2018-0355>
- Li Y, Kim JI, Pysh L, Chapple C (2015) Four isoforms of arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiology* 169:2409–2421. <https://doi.org/10.1104/pp.15.00838>
- Li Z, Nair SK (2015) Structural basis for specificity and flexibility in a plant 4-coumarate:CoA ligase. *Structure* 23:2032–2042. <https://doi.org/10.1016/j.str.2015.08.012>
- Li ZB, Li CF, Li J, Zhang YS (2014) Molecular cloning and functional characterization of two divergent 4-Coumarate: coenzyme A ligases from Kudzu (*Pueraria lobata*). *Biol Pharm Bull* 37:113–122. <https://doi.org/10.1248/bpb.b13-00633>
- Liu T, Yao R, Zhao Y, et al (2017) Cloning, functional characterization and site-directed mutagenesis of 4-Coumarate: Coenzyme a ligase (4CL) involved in coumarin biosynthesis in *Peucedanum praeruptorum* Dunn. *Front Plant Sci.* 17;8:4. <https://doi.org/10.3389/fpls.2017.00004>
- Ljung K, Hull AK, Celenza J, et al (2005) Sites and Regulation of Auxin Biosynthesis in Arabidopsis Roots. *The Plant Cell* 17:1090. <https://doi.org/10.1105/TPC.104.029272>
- Lindermayr C, Fliegmann J, Ebel J (2003) Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. *Journal of Biological Chemistry* 278:2781–2786. <https://doi.org/10.1074/jbc.M202632200>
- Lindermayr C, Möllers B, Fliegmann J, et al (2002) Divergent members of a soybean (*Glycine max* L.) 4-Coumarate:coenzyme A ligase gene family: Primary structures, catalytic properties, and differential expression. *European Journal of Biochemistry* 269:1304–1315. <https://doi.org/10.1046/j.1432-1033.2002.02775.x>
- Lu S, Wang J, Chitsaz F, et al (2020) CDD/SPARCLE: The conserved domain database in 2020. *Nucleic Acids Research* 48:D265–D268. <https://doi.org/10.1093/nar/gkz991>

- Madeira F, Park YM, Lee J, et al (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47:W636–W641. <https://doi.org/10.1093/NAR/GKZ268>
- Manzano C, Pallero-Baena M, Casimiro I, et al (2014) The emerging role of reactive oxygen species signaling during lateral root development. *Plant Physiol* 165:1105–1119
- Md. Emran Khan Chowdhury;, Bosung Choi;, Byoung-Kwan Cho;, et al (2013) Regulation of 4CL, encoding 4-coumarate: Coenzyme A ligase, expression in kenaf under diverse stress conditions | *Plant Omics*. 6:254–262
- Mierziak J, Kostyn K, Kulma A (2014) Flavonoids as important molecules of plant interactions with the environment. *Molecules* 19:16240–16265. <https://doi.org/10.3390/molecules191016240>
- Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 30:2785–2791. <https://doi.org/10.1002/jcc.21256>
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plantarum* 15: 473-497
- Naik P, Wang JP, Sederoff R, et al (2018) Assessing the impact of the 4CL enzyme complex on the robustness of monolignol biosynthesis using metabolic pathway analysis. *PLoS One* 13:e0193896. <https://doi.org/10.1371/journal.pone.0193896>
- Oliveira MB, de Andrade R v., Grossi-de-Sá MF, Petrofeza S (2015) Analysis of genes that are differentially expressed during the *Sclerotinia sclerotiorum*–*Phaseolus vulgaris* interaction. *Front Microbiol* 6:1162. <https://doi.org/10.3389/fmicb.2015.01162>
- Orman-Ligeza B, Parizot B, Rycke R de, et al (2016) RBOH-mediated ROS production facilitates lateral root emergence in *Arabidopsis*. *Development* 143:3328–3339
- Orlova I, Shasany AK, Taguchi G, et al (2012) Contribution of CoA Ligases to Benzenoid Biosynthesis in *Petunia* Flowers. *The Plant Cell* 24:2015–2030. <https://doi.org/10.1105/TPC.112.097519>
- Panche AN, Diwan AD, Chandra SR (2016) Flavonoids: an overview. *J Nutr Sci* 5:e47. <https://doi.org/10.1017/jns.2016.41>
- Peer WA, Bandyopadhyay A, Blakeslee JJ, et al (2004) Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with

- altered auxin transport in *Arabidopsis thaliana*. *The Plant cell* 16:1898–1911. <https://doi.org/10.1105/TPC.021501>
- Peng XQ, Ke SW, Liu JQ, et al (2016) Deletion and hormone induction analyses of the 4-coumarate: CoA ligase gene promoter from *Pennisetum purpureum* in transgenic tobacco plants. *Plant Cell Tiss Organ Cult* 126:439–448. <https://doi.org/10.1007/s11240-016-1012-7>
- Petersen M, Simmonds MSJ (2003) Rosmarinic acid. *Phytochemistry* 62:121–125. [https://doi.org/10.1016/s0031-9422\(02\)00513-7](https://doi.org/10.1016/s0031-9422(02)00513-7)
- Raes J, Rohde A, Christensen JH, et al (2003) Genome-Wide Characterization of the Lignification Toolbox in *Arabidopsis*. *Plant Physiology* 133:1051–1071. <https://doi.org/10.1104/pp.103.026484>
- Rani A, Singh K, Sood P, et al (2009) p-coumarate:CoA ligase as a key gene in the yield of catechins in tea [*Camellia sinensis* (L.) O. Kuntze]. *Funct Integr Genomics* 9:271–275. <https://doi.org/10.1007/s10142-008-0098-3>
- Rao G, Pan X, Xu F, et al (2015) Divergent and Overlapping Function of Five 4-Coumarate/Coenzyme A Ligases from *Populus tomentosa*. *Plant Molecular Biology Reporter* 33:841–854. <https://doi.org/10.1007/s11105-014-0803-4>
- Saewan N, Jimtaisong A (2013) Photoprotection of natural flavonoids. *J Appl Pharm Sci* 3:129–141. <https://doi.org/10.7324/japs.2013.3923>
- Schatz TLG, Grisebach H (1982) Enzymic Synthesis of Lignin Precursors Purification and Properties of 4-Coumarate : CoA Ligase from Cambial Sap of Spruce (*Picea dies* L.) *Eur J Biochem.* 123(3):583-6
- Schmelz S, Naismith JH (2009) Adenylate-forming enzymes. *Curr Opin Struct Biol* 19:666–671. <https://doi.org/10.1016/j.sbi.2009.09.004>
- Schmelzer E, Kruger-Lebus S, Hahlbrock K (1989) Temporal and spatial patterns of gene expression around sites of attempted fungal infection in Parsley Leaves. *Plant Cell* 1:993–1001. <https://doi.org/10.1105/tpc.1.10.993>
- Schmid J, Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. *Phytochemistry* 39:737–749. [https://doi.org/10.1016/0031-9422\(94\)00962-s](https://doi.org/10.1016/0031-9422(94)00962-s)

- Schneider K, Hövel K, Witzel K, et al (2003) The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proc Natl Acad Sci USA* 100:8601–8606. <https://doi.org/10.1073/pnas.1430550100>
- Shah A, Smith DL (2020) Flavonoids in Agriculture: Chemistry and Roles in, Biotic and Abiotic Stress Responses, and Microbial Associations. *Agronomy* 2020, Vol 10, Page 1209 10:1209. <https://doi.org/10.3390/AGRONOMY10081209>
- Shi M-Z, Xie D-Y (2014) Biosynthesis and Metabolic Engineering of Anthocyanins in *Arabidopsis thaliana*. *Recent Patents on Biotechnology* 8:47–60. <https://doi.org/10.2174/1872208307666131218123538>
- Shigeto J, Ueda Y, Sasaki S, et al (2017) Enzymatic activities for lignin monomer intermediates highlight the biosynthetic pathway of syringyl monomers in *Robinia pseudoacacia*. *J Plant Res* 130:203–210. <https://doi.org/10.1007/s10265-016-0882-4>
- Shinde BA, Dholakia BB, Hussain K, et al (2017) Dynamic metabolic reprogramming of steroidal glycol-alkaloid and phenylpropanoid biosynthesis may impart early blight resistance in wild tomato (*Solanum arcanum* Peralta). *Plant Mol Biol* 95:411–423. <https://doi.org/10.1007/s11103-017-0660-2>
- Shockey JM, Fulda MS, Browse J (2003) Arabidopsis contains a large superfamily of Acyl-Activating enzymes. Phylogenetic and biochemical analysis reveals a new class of Acyl-Coenzyme A synthetases. *Plant Physiol* 132:1065–1076. <https://doi.org/10.1104/pp.103.020552>
- Silber M v., Meimberg H, Ebel J (2008) Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. *Phytochemistry* 69:2449–2456. <https://doi.org/10.1016/j.phytochem.2008.06.014>
- Singh P, Jayaramaiah RH, Agawane SB, et al (2016) Potential dual role of eugenol in inhibiting advanced glycation end products in diabetes: proteomic and mechanistic insights. *Sci Rep* 6:18798. <https://doi.org/10.1038/srep18798>
- Singh P, Kalunke RM, Giri AP (2015) Towards comprehension of complex chemical evolution and diversification of terpene and phenylpropanoid pathways in *Ocimum* species. *RSC Advances* 5:106886–106904. <https://doi.org/10.1039/C5RA16637C>

- Singh P, Kalunke RM, Shukla A, et al (2020) Biosynthesis and tissue-specific partitioning of camphor and eugenol in *Ocimum kilimandscharicum*. *Phytochemistry* 177:112451. <https://doi.org/10.1016/j.phytochem.2020.112451>
- Soltani BM, Ehltng J, Hamberger B, Douglas CJ (2006) Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of *Arabidopsis thaliana* 4CL gene family members. *Planta* 224:1226–1238. <https://doi.org/10.1007/s00425-006-0296-y>
- Stöckigt J, Zenk MH (1975) Chemical Syntheses and Properties of Hydroxycinnamoyl-Coenzyme A Derivatives. *Zeitschrift für Naturforschung - Section C Journal of Biosciences* 30:352–358. <https://doi.org/10.1515/znc-1975-5-609>
- Stuible HP, Büttner D, Ehltng J, et al (2000) Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett* 467:117–122. [https://doi.org/10.1016/s0014-5793\(00\)01133-9](https://doi.org/10.1016/s0014-5793(00)01133-9)
- Sun H, Li Y, Feng S, et al (2013) Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem Biophys Res Commun* 430:1151–1156. <https://doi.org/10.1016/j.bbrc.2012.12.019>
- Surjadinata BB, Jacobo-Velázquez DA, Cisneros-Zevallos L (2017) UVA, UVB and UVC light enhances the biosynthesis of phenolic antioxidants in fresh-cut carrot through a synergistic effect with wounding. *Molecules* 22:668. <https://doi.org/10.3390/molecules22040668>
- Sutela S, Hahl T, Tiimonen H, et al (2014) Phenolic compounds and expression of 4CL genes in silver birch clones and Pt4CL1a lines. *PLoS ONE* 9:. <https://doi.org/10.1371/journal.pone.0114434>
- Szklarczyk D, Franceschini A, Wyder S, et al (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43:D447–D452. <https://doi.org/10.1093/nar/gku1003>
- Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>

- Tatsis EC, O'Connor SE (2016) New developments in engineering plant metabolic pathways. *Current Opinion in Biotechnology* 42:126–132. <https://doi.org/10.1016/J.COPBIO.2016.04.012>
- Teale WD, Pasternak T, Bosco CD, et al (2021) Flavonol-mediated stabilization of PIN efflux complexes regulates polar auxin transport. *The EMBO Journal* 40:e104416. <https://doi.org/10.15252/EMBJ.2020104416>
- Tohge T, Watanabe M, Hoefgen R, Fernie AR (2013) Shikimate and phenylalanine biosynthesis in the green lineage. *Front Plant Sci* 4:62. <https://doi.org/10.3389/fpls.2013.00062>
- Trott O, Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 31:NA-NA. <https://doi.org/10.1002/jcc.21334>
- Tzin V, Galili G (2010) The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book Am Soc Plant Biol* 8:e0132. <https://doi.org/10.1199/tab.0132>
- Uhlmann A, Ebel J (1993) Molecular cloning and expression of 4-Coumarate-Coenzyme A Ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. *Plant Physiol* 102:1147–1156. <https://doi.org/10.1104/pp.102.4.1147>
- Vanholme R, Demedts B, Morreel K, et al (2010) Lignin biosynthesis and structure. *Plant Physiology* 153:895–905. <https://doi.org/10.1104/pp.110.155119>
- Vogt T (2010) Phenylpropanoid biosynthesis. *Mol Plant* 3:2–20. <https://doi.org/10.1093/mp/ssp106>
- Wang B, Sun W, Li Q, et al (2015) Genome wide analysis phenylpropanoid biosynthesis genes in *Salvia miltiorrhiza*. *Planta* 241:711–725. <https://doi.org/10.1007/s00425-014-2212-1>
- Wang Y, Yi H, Wang M, et al (2011) Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-CoA Ligase: stilbene synthase. *J Am Chem Soc* 133:20684–20687. <https://doi.org/10.1021/ja2085993>
- Waterhouse A, Bertoni M, Bienert S, et al (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* 46:W296–W303. <https://doi.org/10.1093/NAR/GKY427>

- Wei XX, Wang XQ (2004) Evolution of 4-coumarate:coenzyme A ligase (4CL) gene and divergence of *Larix* (Pinaceae). *Mol Phylogenet Evol* 31:542–553. <https://doi.org/10.1016/j.ympev.2003.08.015>
- Whetten R, Sederoff R (1995) Lignin Biosynthesis. *Plant Cell*. 7(7):1001-1013. doi: 10.1105/tpc.7.7.1001.
- Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiol Plant* 107:142–149. <https://doi.org/10.1034/j.1399-3054.1999.100119.x>
- Zhang CH, Ma T, Luo WC, et al (2015) Identification of 4CL genes in desert Poplars and their changes in expression in response to salt stress. *Genes* 6:901–917. <https://doi.org/10.3390/genes6030901>
- Zhao Y, Dong W, Wang K, et al (2017) Differential sensitivity of fruit pigmentation to ultraviolet light between two peach cultivars. *Front Plant Sci* 8:1552. <https://doi.org/10.3389/fpls.2017.01552>

ABSTRACT

Name of the Student: Lavhale Santosh Govind**Registration No. :** 10BB15A26044**Faculty of Study:** Biological Sciences**Year of Submission:** 2022**AcSIR academic centre/CSIR Lab:** CSIR NCL, Pune**Name of the Supervisor:** Dr. Ashok P. Giri**Title of the thesis:** Functional characterization of 4-Coumarate-CoA ligases involved in phenylpropanoid biosynthesis from *Ocimum kilimandscharicum*

In plants, specialized metabolites are indispensable for growth and development and various biotic and abiotic stress responses. 4-Coumarate-CoA Ligase (4CL) is a diverse group of enzymes mainly involved in the biosynthesis of several phenolic specialized metabolites. 4CL catalyzes the ligation of CoA to cinnamic acid and its derivatives. Activated CoA esters are utilized for the biosynthesis of phenolic metabolites and lignin. *Ocimum* species produce a variety of such specialized metabolites having multiple medicinal properties. Here, we characterize the diversity of *Ocimum kilimandscharicum* 4CLs (Ok4CLs).

Gene expression analysis suggested that *Ok4CL7* is highly expressed in leaf trichomes, whereas *Ok4CL15* is abundant in the roots. The recombinant Ok4CL7 and -15 had optimal enzyme activities at 40 °C in pH 8 and 7, respectively. Ok4CL7 showed substrate preference towards p-coumaric acid, ferulic acid and caffeic acid. While, Ok4CL15 preferred p-coumaric acid, ferulic acid and sinapic acid.

In silico analysis was performed to decipher the reason behind the differential substrate selectivity. Molecular modeling and docking, followed by *in silico* study of enzyme-substrates interaction, provided information about residues contacts with the substrate in the binding pocket. Feruloyl adenylate showed a higher number of contacts and lower binding energy with Ok4CL7 and -15 than cinnamoyl adenylate. Differential substrate selection or affinity towards a particular substrate by Ok4CLs may contribute to regulation flux in the phenylpropanoid pathway. Further, mutant Ok4CL proteins were expressed to validate the *in-silico* results.

Owing to a poor regeneration efficiency coupled with limited success of *Agrobacterium*-mediated stable transformation methods for several members of *Ocimum* species, we have overexpressed three Ok4CL isoforms (Ok4CL7, -11 and -15) in *N. benthamiana* and investigated their effects on the plant phenotypes and metabolite profiling. Our experiments showed that overexpression (OE) of *Ok4CL11* in *N. benthamiana* leads to rootless/reduced root growth phenotype compared to wild-type (WT), but the other two isoforms of Ok4CLs (Ok4CL7 and -15) do not show such root growth phenotype. Targeted metabolite profiling of whole *in vitro* plants showed an excessive accumulation of a flavonoid, kaempferol. Silencing of *Ok4CL11* in the *Ok4CL11-OE* background partially complemented the rootless phenotype and these RNAi lines exhibited improved root growth compared to the background OE lines. Grafting suggested that the reduced root growth phenotype cannot be rescued by grafting WT on OE line (WT/OE), whereas the opposite hetero-grafts (OE/WT) produced roots from the WT stock. In summary, our results suggest the unique function of Ok4CL11 as a negative regulator of root growth development possibly by inhibiting polar auxin transport *via* excessive accumulation of kaempferol.

Details of the publications emanating from the thesis work

List of publications

1. **Lavhale, S.G.**, Joshi, R.S., Kumar, Y., Giri, A.P., Functional insights into two *Ocimum kilimandscharicum* 4-coumarate-CoA ligases involved in phenylpropanoid biosynthesis. Int J Biol Macromol. (2021) 30;181:202-210
2. **Lavhale, S.G.**, Kalunke, R.M. & Giri, A.P., Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. Planta (2018), 248: 1063-1078
3. **Lavhale, et al.**, Identification of amino acid residues in 4-Coumarate CoA ligase isoforms responsible for substrate specificity and transmembrane localization **(2022) (under preparation)**
4. **Lavhale, et al.**, Overexpression of *Ocimum 4-Coumarate CoA ligase 11* negatively regulate root growth development in *N. benthamiana* by inhibiting polar auxin transport via excessive accumulation of a flavonoid, k'aempferol **(2022) (under preparation)**



Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: <http://www.elsevier.com/locate/ijbiomac>

Functional insights into two *Ocimum kilimandscharicum* 4-coumarate-CoA ligases involved in phenylpropanoid biosynthesis

Santosh G. Lavhale^{a,b}, Rakesh S. Joshi^{a,b}, Yashwant Kumar^c, Ashok P. Giri^{a,b,*}^a Plant Molecular Biology Unit, Biochemical Sciences Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India^b Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh 201002, India^c Translational Health Science and Technology Institute (THSTI), Faridabad, Haryana 121001, India

ARTICLE INFO

Article history:

Received 25 December 2020

Received in revised form 25 February 2021

Accepted 22 March 2021

Available online 24 March 2021

Keywords:

Ocimum kilimandscharicum

4-Coumarate-CoA ligase

Flavonoids

Lignin

Phenylpropanoids

ABSTRACT

Plant 4-coumarate-CoA ligase (4CL) catalyzes the ligation of CoA to cinnamic acid and its derivatives. Activated CoA esters are utilized for the biosynthesis of phenolic metabolites and lignin that play essential function in plants. Here, we characterize the diversity of *Ocimum kilimandscharicum* 4CLs (Ok4CLs). Phylogenetic analysis suggest that Ok4CLs could be grouped into three classes, class I - enzymes mostly involved in lignin biosynthesis, class II - non-structural phenylpropanoid biosynthesis and class III - yet to be characterized for specific role(s). We selected two Ok4CLs namely Ok4CL7 and Ok4CL15 for further characterization. Gene expression analysis suggested that Ok4CL7 is highly expressed in leaf trichomes, whereas Ok4CL15 is abundant in the roots. The recombinant Ok4CL7 and Ok4CL15 had optimal enzyme activities at 40 °C in pH 8 and 7, respectively. Ok4CL7 showed substrate preference towards *p*-coumaric acid, ferulic acid and caffeic acid. While, Ok4CL15 preferred *p*-coumaric acid, ferulic acid and sinapic acid. Feruloyl adenylate showed higher number of contacts and lowers binding energy with Ok4CL7 and 15 compared to cinnamoyl adenylate. Based on root-specific expression and preference for sinapic acid, Ok4CL15 might be involved in lignin biosynthesis. Further exploration is needed to unravel the role of diverse Ok4CLs in *O. kilimandscharicum*.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Phenylpropanoids protect plants from various biotic or abiotic stress conditions. Many of them are known to have or being explored for medicinal properties. 4-coumarate CoA ligase (EC 6.2.1.12) (4CL) is a crucial enzyme involved in the phenylpropanoid pathway as well as lignin biosynthesis. It serves as the main branch point in the pathway [1]. They are encoded by a multigene family of adenylate forming enzymes. 4CLs convert hydroxy or methoxy cinnamic acid derivatives to the corresponding activated thioesters. Products of 4CL are utilized by various oxygenase, reductases, and transferases for the biosynthesis of lignin, flavonoids, anthocyanins, tannins, aurones, stilbenes, coumarins, suberin, cutin, sporopollenin, etc. [2]. It has been reported that 4CLs might play crucial role in lignin biosynthesis [3–5]. Lignin is the polymer of H-lignin, G-lignin and S-lignin monomers. Monolignols of H-lignin, G-lignin and S-lignin are synthesized from *p*-coumaroyl CoA, feruloyl CoA and sinapoyl CoA, respectively [4–6]. Also, several non-structural phenylpropanoids synthesized via 4CLs exhibit diverse functions in plant physiology. Among these, flavonols and isoflavones

negatively regulate the transport of auxin hormone [7]. They are also involved in the attraction of symbiotic bacteria for nitrogen fixation [8,9]. Moreover, flavonoids protect the plant against UV-B irradiation, sugar stress, nutrition deficient (low phosphate/iron/nitrogen), low temperature, drought, pathogen infection, and herbivores attack [10]. Anthocyanins, another metabolite from this class, shows various colors ranging from orange to pink, red, and purple. They supply pigmentations to flowers, fruits and seeds to attract pollinators and seed dispersal [11].

4CL enzymes contain two conserved peptide motifs: box I (SSGTTGLPKGV) and box II (GEICIRG) [12,13]. Members of adenylate forming enzymes family have conserved adenosine monophosphate binding domain (box I). Crystal structure analysis suggested that during the catalysis process, enzyme undergo two conformations: adenylate forming and thioester forming [14,15]. 4CL isoforms with differential substrate affinities can be used to manipulate metabolite flux. Expression of 4CL isoforms can be modulated in response to a specific trigger [16–18]. In *Arabidopsis thaliana*, 4CL1 was upregulated in leaf in response to the wound, while 4CL3 was upregulated in light-exposed seedlings [12,19]. Similarly, *Oryza sativa* 4CL2 was upregulated by UV irradiation, suggesting its role in flux diversion towards flavonoid biosynthesis [17]. *Ocimum basilicum* 4CL (*Ob4CL*) involved in the non-structural phenolic biosynthesis was downregulated in response to drought stress [20]. 4CLs expression transcriptionally regulated by the

* Corresponding author at: Plant Molecular Biology Unit, Biochemical Sciences Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411 008, MS, India.
E-mail address: ap.giri@ncl.res.in (A.P. Giri).

differential methylation patterns of their promoters [21]. Furthermore, they are controlled by MYB transcription factors in response to various plant hormones like abscisic acid, methyl jasmonate and salicylic acid [22–24].

Biocatalytic properties of 4CLs have been explored for commercial applications in fuel, flavour and natural product synthesis [16,25–27]. The silencing of lignin biosynthesis specific 4CL isoform causes reduction in lignin content. This de-lignified precursor serves as a suitable source for producing chemicals, fiber, energy, or fuel [25,28]. In another study, resveratrol biosynthesis was improved in yeast by fusion of *At4CL1* from *A. thaliana* and *STILBENE SYNTHASE* from *Vitis vinifera* (VvSTS) [29]. Moreover, curcuminoids biosynthesis in *Escherichia coli* was achieved by expressing 4CL from *Lithospermum erythrorhizon* and *CURCUMINOID SYNTHASE* from rice [30]. These potential applications of 4CL urge the importance of molecular investigation in the functionality of 4CLs in phenylpropanoid rich medicinal plants like species of *Ocimum*.

Ocimum has remarkable diversity in metabolite contents with a variety of medicinal properties [31]. Functions of *Ocimum* 4CLs beyond phenylpropanoid biosynthesis are still enigmatic. Gene expression profiles and quantification of metabolite contents showed that eugenol and camphor biosynthesis is tissue-specific in *O. kilimandscharicum* [32]. A particular 4CL isoform was co-expressed with *EUGENOL SYNTHASE 1* in leaf tissue [32]. In present report, we have analyzed the diversity of 4CL using the phylogenetic and similarity search network. Following this, we have performed the structural-functional characterization of two recombinant 4CL isoforms from *O. kilimandscharicum* (Ok4CL7 and -15). Molecular details of their substrate specificities were analyzed using molecular docking. Characterization of two recombinant Ok4CL7 and -15 might shed light on their putative functions in the *O. kilimandscharicum* phenylpropanoid and lignin pathway.

2. Materials and methods

2.1. Plant material and chemicals

O. kilimandscharicum plants were grown in the field at CSIR-National Chemical Laboratory, Pune, India. After flowering, tissues (Flowers, young leaves, roots, and trichomes) were harvested and stored in liquid nitrogen for further analysis. Substrates of 4CL, such as p-coumaric acid, cinnamic acid, ferulic acid, caffeic acid and sinapic acid (Sigma-Aldrich, St. Louis, USA), Coenzyme A (Sisco Research Laboratories, Mumbai, India), complementary DNA (cDNA) synthesis kit (Applied Biosystems, Waltham, USA), RNA isolation total plant RNA extraction-Spectrum kit (Sigma Aldrich) were purchased. The plasmids like pGEM-T, pGEX 4T, pET28a and pRI 101-AN were purchased from Takara (Takara, Kyoto, Japan) and the solvents for High-Performance Liquid Chromatography (HPLC) and Mass Spectrometry (MS) from Merck (Kenilworth, NJ, USA).

2.2. Identification of 4-Coumarate CoA ligases from *O. kilimandscharicum* transcriptome datasets

Total RNA extracted from 100 mg tissue of flowers, young leaves, roots, and trichome using Spectrum Plant RNA Isolation kit (Sigma-Aldrich). The quality of RNA was checked on the NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA). The transcriptome was sequenced using the Illumina *NextSeq500* next-generation sequencing (NGS) (San Diego, USA) platform. De novo assembly was performed using Trinity software without a reference genome [33] and transcript annotation using NCBI-BLAST-2.2.29+ tool [34]. Transcriptome Sequencing and analysis is outsourced at Genotypic Technology, Bangalore, India. Annotated 4-Coumarate CoA Ligase (4CL) encoding genes were selected from the de novo assembled *O. kilimandscharicum* transcriptome. Open Reading Frame (ORF) analysis was performed using the ORF Finder web tool from NCBI (<https://www.ncbi.nlm.nih.gov/orffinder/>).

Sequences having start and stop codons as well as the presence of two conserved signature motifs (Box-I and Box-II) selected for further study. All those enzymes having Box-I (SSGTTGLPKGV) are grouped into a superfamily of adenylate forming enzymes. Fifteen ORF sequences had these two conserved motifs, and they were selected for further analyses.

2.3. Phylogenetic, sequence similarity network and domain analysis

Phylogenetic analysis of 4CLs was performed using MEGA 6 software [35]. Deduced amino acid sequences of all putative 4CLs from *O. kilimandscharicum* and reported 4CLs were aligned using the MUSCLE alignment tool [36]. Reported 946 plant 4CL sequences were retrieved from UniProt protein database (partial, hypothetical and putative sequences were excluded). Aligned sequences in fasta file format used for phylogenetic tree construction using the neighbour-joining method with 1000 bootstrap value and by keeping the rest of the parameters default. The relation among 4CL sequences was also studied by Sequence Similarity Network (SSN) using SSNpipe with default protocol (<https://github.com/ahvdk/SSNpipe>, <https://github.com/ahvdk/SSNpipe/wiki/SSNpipe-Usage-Examples>). BLAST e-value used for edge distance calculations. $1e140$ cut-off value was used for network building as it is the highest to achieve confident clustering. The output network was visualized using Cytoscape 3.7.2 (<https://cytoscape.org/>). Domain architecture of selected 4CLs studied using Conserved Domains Database (CDD) [37]. The schematic diagram of the proteins domain shown in Fig. 1C is prepared using DOG (Domain illustrator) [38].

2.4. Gene expression analysis

Gene expression of putative 4CLs studied in different tissues from *O. kilimandscharicum*. Total RNA was isolated from 100 mg of freshly collected tissue, using Spectrum Plant RNA extraction kit as described [39]. Genomic DNA contamination in isolated RNA was removed using RNase free DNase treatment. Two μg RNA was used for cDNA synthesis using SuperScript III reverse transcriptase system (Invitrogen, Carlsbad, USA). cDNA synthesis was confirmed by PCR using a set of primers that amplifies a short sequence elongation factor 1 α (EF1 α). qRT-PCR primers were designed for all putative 15 *Ok4CL* genes using Primer 3.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and Oligo Analysis online tool (<http://www.operon.com/tools/oligo-analysis-tool.aspx>). qRT-PCR primers were synthesized (Eurofins, Bangalore, India) and used for analysis. qRT-PCR was performed on an Applied Biosystems 7500 Fast Real-Time PCR System by using the SYBR green protocol. Three biological and technical replicates were used for each sample. The reaction mixture contains 5 μL of SYBR green master-mix, 0.5 μL of 10 μM forward and reverse gene-specific primers (Supplementary Table S1) and 1 μL of diluted cDNA (1:2) with nuclease-free water added to make up a volume to 10 μL . Standard plots of putative 4CLs were generated using a gene-specific set of primer-pairs. Different cDNA dilutions (1:2, 1:3, 1:8, 1:16 and 1:32) were used to generate standard plots. The qRT-PCR reaction for sample cDNA was optimized using different concentrations of primer and cDNA. Absolute gene expression was performed using the standard plot method. Heatmap for expression of *Ok4CL* genes expression across various tissues of *O. kilimandscharicum* was illustrated using normalized z-score. Multiple Experimental Viewer (MeV) (<http://mev.tm4.org/#/welcome>) software was used to plot the heatmap.

2.5. Cloning, recombinant protein expression and purification of *Ok4CL7* and 15

As the representative of different clades, two candidate *Ok4CLs*, *Ok4CL7* and 15, were selected for heterologous expression and in vitro characterization. Putative full-length coding sequences of *Ok4CL7* (1730 bp) and *Ok4CL15* (1608 bp) were amplified from cDNA by PCR

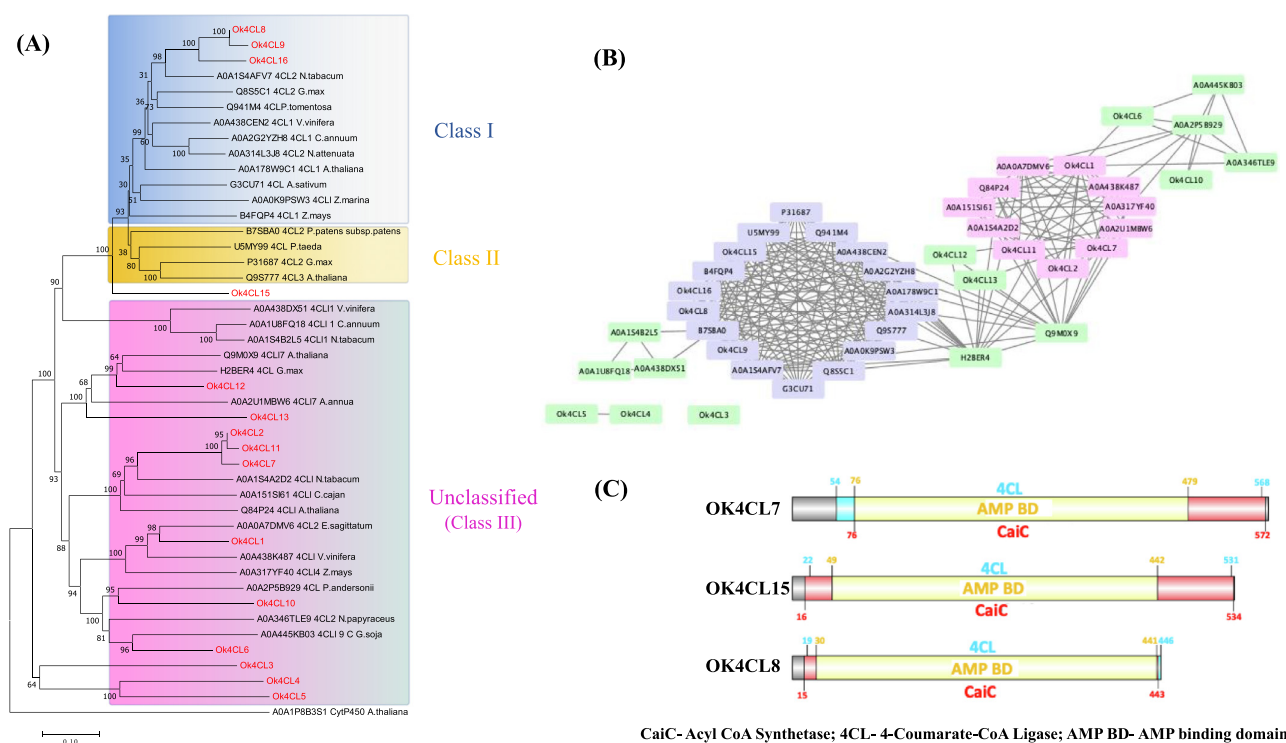


Fig. 1. Diversity and domain analysis of Ok4CLs. A. Phylogenetic analysis of 4CLs. Class-I 4CLs are involved in the biosynthesis of lignin and Class-II-4CLs are involved in the biosynthesis of non-structural phenylpropanoids. B. The sequence similarity network analysis of various plant 4CLs, including that from *O. kilimandscharicum* showed five clusters at the E -value of $1e^{-40}$. Three clusters were observed in the SSN network in corroboration with a phylogenetic tree. Ok4CL7 and -11 are grouped in one cluster similar to a single clade. Ok4CL4, -5 and -6 are found to be disconnected from the total network, indicating their individuality similar to the phylogenetic tree. The remaining sequences were grouped in two major clusters. C. Domain analysis. Ok4CL7, 15 and 8 from *O. kilimandscharicum* have several domains, including CaIC (Acyl CoA Synthetase); 4CL (4-Coumarate-CoA Ligase); AMP BD- AMP binding domain

using a high-fidelity Q5 polymerase (New England Biolabs, Ipswich, USA). Amplified sequences were cloned into expression vectors (*Ok4CL7* in pGEX 4 T and *Ok4CL15* in pET 28a) using T4 DNA Ligase (New England Biolabs). *NdeI* and *EcoRI/SalI* restriction sites used for *Ok4CL7* cloning, while *NdeI* and *SalI* used for *Ok4CL15*. These restriction sites were included in forward and reverse primers, respectively (Supplementary Table S1).

The constructs having either of the 4CL genes were transformed in ArcticExpress (DE3) competent cells (Agilent Technologies, Santa Clara, USA). A single colony of ArcticExpress cells having expression construct/plasmid was inoculated in primary culture and grown in a shaker incubator at 37 °C overnight (~14 h). 10 mL of inoculum used from primary culture for 1 l of secondary culture. Secondary culture allowed growing up to log phase ($OD_{600} \sim 0.45$). Once the culture reaches to log-phase, then it is kept at 4 °C for chilling. Gene expression was induced by adding 0.5 M IPTG to 1 L culture ($OD_{600} \sim 0.45$) and cultures grown at 12 °C in shaker incubator for overnight (~14 h). The bacterial cell pellet was obtained by centrifugation of culture at 3500 $\times g$ for 10 min. Pellet was resuspended in 10 mL lysis buffer (Tris [50 mM], NaCl [200 mM], glycerol [5% v/v] and lysozyme [0.2 mg/mL] pH 8.0), followed by sonication using VibraCell (Sonics, Newtown, USA) (10 s on and off cycle with an amplitude of 45%) for cell lysis. For both recombinant proteins, buffer containing Tris (50 mM), NaCl (200 mM), glycerol (5% v/v) was used throughout the purification process. Cell lysate/cell debris was separated using centrifugation at 11,200g for 60 min at 4 °C. The supernatant was collected in a fresh tube. Supernatant further filtered using 0.2-micron syringe filters for removal of any solid matter. Clear supernatant along with slurry was kept for binding on a shaker for 3 h at 4 °C. *Ok4CL7* was purified using Glutathione Sepharose 4B (GE Healthcare, Chicago, USA), whereas *Ok4CL15* with Ni-NTE slurry (Roche, Basel, Switzerland). For

elution of *Ok4CL7*, reduced glutathione (20 μ M; pH 8.0) was used, whereas imidazole (200 μ M; pH 8.0) used for *Ok4CL15*. Size of purified proteins confirmed by separating on 12% SDS-PAGE gel. Salts were removed using dialysis against Milli Q water. Purified recombinant *Ok4CL7* and *Ok4CL15* protein aliquots were stored at -80 °C after the snap freeze in the liquid nitrogen until further use.

2.6. Enzymatic assays of recombinant *Ok4CL7* and *Ok4CL15*

All enzyme assays were performed in three replicates. The activity of recombinant purified *Ok4CL* enzymes was studied with five substrates: *p*-coumaric acid, caffeic acid, cinnamic acid, ferulic acid and sinapic acid as described previously [1,23]. The reaction mixture of 200 μ L contains substrate (0.2 mM), adenosine triphosphate (2.5 M), coenzyme A (0.2 mM) and enzyme (5 μ g). All assays were performed in Tris-HCl buffer (200 mM; pH 8.0) having MgCl₂ (25 mM). Blanks used for all assays contain heat-inactivated enzymes and all other components. The absorbance change was monitored with a UV-visible spectrophotometer (Labindia Pvt. Ltd., Mumbai, India). Optimum pH and temperature required for recombinant *Ok4CL7* and *Ok4CL15* activity were determined using ferulic acid and sinapic acid, respectively. Range of pH 5 to pH 9 used for determination optimum pH, while for finding out optimum temperature range from 20 °C to 80 °C were used. The absorbance was measured at absorption maxima at 311, 333, 345, 346, and 352 nm for the corresponding cinnamoyl-CoA, 4-coumaroyl-CoA, feruloyl-CoA, caffeoyl-CoA, and sinapoyl-CoA products, respectively [40,41]. Substrate concentration used for enzyme kinetics ranged from 10 to 400 μ M for coumaric acid and caffeic acid, whereas 25 to 800 μ M for ferulic acid. All other components of the reaction were kept constant (ATP [2.5 mM]; CoA [0.2 mM], Enzyme [5 μ g]). The average value of three replicates used to plot the

graph. Standard deviation calculated using three replicate values and plotted as an error bar of respective value in the graph.

2.7. Characterization of recombinant Ok4CL products on HPLC and LC-MS/MS

Ok4CL assay products for substrates 4-coumaric acid, caffeic acid, ferulic acid and sinapic acid were analyzed on Waters HPLC using reversed-phase C18 column at 24°C with 0.7 mL/min flow rate. Photodiode array (PDA) detector used to analyze the change in absorption. 20 µL of sample injected from each reaction. Three runs of each enzyme with each substrate. i) standard substrate; ii) heat-inactivated enzyme and all other components; iii) reaction mixture with the active enzyme. Product peak was observed at absorbance 333 nm (p-coumaroyl CoA), 345 nm (caffeoyl CoA), 346 nm (feruloyl CoA) and 352 nm (sinapoyl CoA) [22,42]. Acetonitrile and ammonium acetate (0.1 M; pH 4.5) was used as the mobile phase. Previous protocol with modifications was used for HPLC analysis [43]. The gradient used ranges from 0 to 100% ACN. ACN reaches 65% in approximately 25 min and 100% in about 28 min. 100% ACN wash was given for 30 min. For equilibration of the system before the next run, 100% ammonium acetate (pH 4.5) for 8 min. The product peak was collected and dried using lyophilizer. Purified metabolites were dissolved in 50% acetonitrile and 50% water and used for characterization by LCMS. Orbitrap Fusion mass spectrometer (Thermo Scientific) coupled with the heated electrospray ion source was used for data acquisition. For MS1 mode, the mass resolution was kept at 120,000 and for MS2 acquisition, the mass resolution was 30,000. Mass range of data acquisition was 60–900 da. Extracted metabolites were separated on UPLC ultimate 3000 installed with the Xbridge Amide column. Accurate mass and fragmentation pattern were acquired for the mixture by separating them on the HILIC column and positive and negative ionization mode both. The purified mixture was separated by solvent A was 20 mM ammonium acetate in the water of pH 9.0 and mobile phase B was 100% acetonitrile. The elution gradient starts from 85% B to 10% B over 14 min, with a flow rate of 0.35 mL/min. LC/MS acquired data has been processed using the Xcalibur software (Thermo Scientific) using the default setting.

2.8. Molecular docking analysis

To understand the interaction and substrate preferences of Ok4CLs molecular docking was performed. Three-dimensional structures of Ok4CL7 and Ok4CL15 were predicted using *Populus tomentosa* 4CL (PDB ID: 3NI2) and *Nicotiana tabacum* 4CL (PDB ID: 5BST) crystal structures as templates, respectively. Structures were energy minimized using Maestro 10.1 Tools (Maestro, Schrödinger, LLC, New York, NY, 2020) and substrate binding pocket residues were predicted by superimposing models with templates. Ok4CL structures were then prepared for docking by adding Kollman and Gasteiger charges in AutoDock Tools [44]. After adding polar hydrogens to protein structure, the docking grid was set around substrate-binding residues using the AutoGrid Tool. The prepared structure was saved in *.pdbqt format for further analysis. Structures of the feruloyl adenylate and cinnamoyl adenylate were generated in Marvin sketch software (<http://www.chemaxon.com>) and further converted in *.pdbqt format using AutoDock Tool. Docking of substrates with Ok4CLs was performed using AutoDock Vina [45]. After docking simulations, ten docking poses were generated with each substrate. The binding pose with the lowest binding score was selected for further analysis. Substrate binding poses were analyzed using the BIOVIA Discovery Studio 4.5 software (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016).

2.9. Statistical analysis

At least three replicates used for all statistical analysis. Significance of Ok4CL7 and Ok4CL15 gene expression calculated using multiple unparallel

t-tests in GraphPad Prism version 9.0.0 for macOS, GraphPad Software, CA, USA, www.graphpad.com. (ns = $P > 0.05$; * = $P \leq 0.05$; ** = ≤ 0.01 ; *** = $P \leq 0.001$; **** = $P \leq 0.0001$). Significant differences between data were determined using the One-way Analysis of Variance (ANOVA: single factor) tests. Error bars represent the mean \pm standard deviation. One-way ANOVA test suggested a significant difference between data $p < 0.005$ for both the enzymes substrate preference experiments.

3. Results and discussion

3.1. Ok4CLs diverged depending on their biological roles

Fifteen annotated 4CL sequences were retrieved from *O. kilimandscharicum* transcriptome datasets. Domain analysis depicts that most of Ok4CLs contain AMP-BD, CaiC, and 4CL domains, indicating their homology and possible functional conservation (Table 1). The open reading frame of these Ok4CLs had Box-I and Box-II signature motifs. The phylogenetic tree was constructed using 946 plant 4CLs protein sequences, including 15 putative Ok4CLs. These 4CLs were grouped into nine major clades (Supplementary Fig. S1). Representative 4CLs selected from each clade and putative Ok4CLs were further used for phylogenetic tree construction (Fig. 1A). Selected 4CLs were grouped into three major clades belonging to known three classes (CLs involved in lignin biosynthesis, non-structural phenolics biosynthesis and uncharacterized) as mentioned earlier. Ok4CL8, 9 and 16 grouped with other 4CLs in class-I, while all other Ok4CLs appeared in the uncharacterized clade (Class III) along with other 4CLs. Ok4CL3, 4 and 5 form a separate clade due to the presence of the fatty acyl CoA synthetase domain. Interestingly, Ok4CL15 is not clustered with any of the class. Comparatively, It is close to class I and class II than unclassified 4CLs.

A similar classification was reported for the 4CL isoforms in several plants [12,16,46]. *Piper nigrum* 4CLs distributed in class-I and class-II. Class-I 4CLs were active towards the lignin biosynthesis substrates like coumaric and ferulic. In addition to this, Class-II 4CLs are also active towards piperonylic acid, 3,4-(Methylenedioxy) cinnamic acid (MDCA), and piperic acid [47]. 4CLs from monocot and dicot separated into two clades due to their parallel evolution after speciation [17]. *Pinus taeda* 4CL (Pinta 4CL3, gymnosperm) isoform was grouped with Class-II angiosperm 4CLs, suggesting their evolutionary conservation [48]. Some 4CLs contain conserved motifs as 4CLs and they are homologous with 4CL sequences [49]. These are also classified as 4CL-like and their functions are not yet clear [50].

Further, similarity among the 4CLs was studied using the sequence similarity network (SSN). Our analysis showed three major clusters with an E value of $1e^{140}$ (Fig. 1B). In the first cluster, Ok4CL7 and Ok4CL15 were grouped with 4CL from *Epimedium sagittatum* (A0A0A7DMV6), *Prunus andersonii* (A0A2P5B929) and *Narcissus papyraceus* (A0A346TLE9). In the second cluster, Ok4CL2, Ok4CL3, Ok4CL8, Ok4CL12, Ok4CL13 and Ok4CL15 were clustered with others 4CLs from *Nicotiana tabacum* (A0A1S4A2D2), (A0A1S4AFV7), *Zea mays* (A0A317YF40), *Vitis vinifera* (A0A338K487), *Arabidopsis thaliana* (Q9M0X9) (Q82P24), *Cajanus cajan* (A0A151S161), *Glycine max* (H2BER4), *Artemisia annua* (A0A2U1MBW6). In the third cluster, OkCL1, OkCL9, OkCL10 and OkCL16 were clustered with 4CLs from *Nicotiana attenuata* (A0A314L3J8), *G. max* (P31687), *Physcomitrella patens* subsp. *patens* (B7SBA0), *V. vinifera* (A0A438CEN2), *G. max* (Q8S5C1) *A. thaliana* (Q9S777), *C. annuum* (A0A1U8FQ18), *Allium sativum* (G3CU71G3CU71) *Zostera marina* (A0A0K9PSW3), *Populus tomentosa* (Q941M4), *A. thaliana* (A0A178W9C1), *Capsicum annuum* (A0A2G2YZH8) and *Pinus taeda* (U5MY99). Corroboration in phylogenetic and SSN clustering affirms the 4CLs clustering according to their evolutionary conservation.

The 4CL isoforms are characterized from several plants. These isoforms were grouped into two broad classes. Class-I mainly involved in lignin biosynthesis and contains 4CLs from dicotyledonous plants, such as At4CL1 and At4CL2 [2,36]. Many of candidates from Class-II

Table 1
Classified features of Ok4CL enzyme (clade and domain arrangement) along with selected other plant 4CLs.

Name and accession number	Length (aa)	Clade	Domain(s)
Ok4CL1_ MW413287	555	Clade III	AMP-BD, CaiC, 4CL
Ok4CL2_ MW413288	477	Clade III	AMP-BD, CaiC, 4CL
Ok4CL3_ MW413289	523	Clade III	AMP-BD, CaiC, FACL_fum10p_like
Ok4CL4_ MW413290	484	Clade III	AMP-BD, CaiC, ttLC FACS AEE21 like
Ok4CL5_ MW413291	614	Clade III	AMP-BD, CaiC, PLN02479, ttLC FACS AEE21 like
Ok4CL6_ MW413292	380	Clade III	AMP-BD, CaiC
Ok4CL7_ MW413293	576	Clade III	AMP-BD, CaiC, 4CL
Ok4CL8_ MW413294	446	Clade I	AMP-BD, CaiC, 4CL
Ok4CL9_ MW413295	425	Clade I	AMP-BD, CaiC
Ok4CL10_ MW413296	454	Clade III	AMP-BD, CaiC
Ok4CL11_ MW413297	576	Clade III	AMP-BD, CaiC, 4CL
Ok4CL12_ MW413298	424	Clade III	AMP-BD, CaiC, 4CL
Ok4CL13_ MW413299	454	Clade III	AMP-BD, CaiC
Ok4CL15_ MW413300	535	Clade I/II/III	AMP-BD, CaiC, 4CL
Ok4CL16_ MW413301	412	Clade I	AMP-BD, CaiC
<i>N. tabacum</i> 4CL2_A0A1S4AFV7	542	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>G. max</i> 4CL2_Q8S5C1	547	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>P. tomentosa</i> 4CL_Q941M4	536	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>V. vinifera</i> 4CL1_A0A438CEN2	548	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>C. annuum</i> 4CL1_A0A2G2YZH8	553	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>N. attenuata</i> 4CL2_A0A314L3J8	558	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>A. thaliana</i> 4CL1_A0A178W9C1	561	Clade I	AMP-BD, CaiC, 4CL, PLN02246, PLN02574
<i>A. sativum</i> 4CL_G3CU71	545	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>Z. marina</i> 4CL1_A0A0K9PSW3	552	Clade I	AMP-BD, CaiC, 4CL
<i>Z. mays</i> 4CL1_B4FQP4	555	Clade I	AMP-BD, CaiC, 4CL, PLN02246
<i>P. patens</i> subsp. <i>patens</i> 4CL2_B7SBA0	585	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>P. taeda</i> 4CL_U5MY99	575	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>G. max</i> 4CL2_P31687	562	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>A. thaliana</i> 4CL3_Q9S777	561	Clade II	AMP-BD, CaiC, 4CL, PLN02246
<i>V. vinifera</i> 4CL1_A0A438DX51	547	Clade III	AMP-BD, CaiC
<i>C. annuum</i> 4CL1_A0A1U8FQ18	554	Clade III	AMP-BD, CaiC, 4CL
<i>N. tabacum</i> 4CL1_A0A1S4B2L5	551	Clade III	AMP-BD, CaiC, 4CL
<i>A. thaliana</i> 4CL17_Q9M0X9	544	Clade III	AMP-BD, CaiC, 4CL
<i>G. max</i> 4CL_H2BER4	540	Clade III	AMP-BD, CaiC, 4CL
<i>A. annua</i> 4CL17_A0A2U1MBW6	535	Clade III	AMP-BD, CaiC, 4CL
<i>N. tabacum</i> 4CL1_A0A1S4A2D2	565	Clade III	AMP-BD, CaiC, 4CL
<i>C. cajan</i> 4CL1_A0A151SI61	608	Clade III	AMP-BD, CaiC, 4CL
<i>A. thaliana</i> 4CL1_Q84P24	566	Clade III	AMP-BD, CaiC, 4CL, PLN02574
<i>E. sagittatum</i> 4CL2_A0A0A7DMV6	550	Clade III	AMP-BD, CaiC, 4CL
<i>V. vinifera</i> 4CL1_A0A438K487	549	Clade III	AMP-BD, CaiC, 4CL
<i>Z. mays</i> 4CL14_A0A317YF40	551	Clade III	AMP-BD, CaiC, 4CL
<i>P. andersonii</i> 4CL_A0A2P5B929	556	Clade III	AMP-BD, CaiC, 4CL
<i>N. papyraceus</i> 4CL2_A0A346TLE9	560	Clade III	AMP-BD, CaiC, 4CL
<i>G. soja</i> 4CL19C_A0A445KB03	472	Clade III	AMP-BD, CaiC, 4CL

AMP-BD:AMP binding domain; CaiC:Acyl-CoA synthetase; 4CL:4 Coumarate CoA Ligase PLN02246:4-coumarate–CoA ligase; PLN02574:4-coumarate–CoA ligase- like; ttLC FACS AEE21 like: Fatty acyl-CoA synthetases similar to LC-FACS from *Thermus thermophilus* and *Arabidopsis*; FACL_fum10p_like: Subfamily of fatty acid CoA ligase (FACL) similar to Fum10p of *Gibberella moniliformis*.

are possibly involved in the biosynthesis of non-structural phenolics, such as flavonoids and antitoxins and possess 4CLs from monocotyledonous, dicotyledonous and gymnosperm plants, e.g. At4CL3 [1,51,52].

Selected Ok4CLs have conserved adenosine monophosphate binding sites and acyl CoA synthetase (CaiC) domain. 4CL domain was absent in Ok4CL3, Ok4CL4, Ok4CL5, Ok4CL9 and Ok4CL16. Out of these, Ok4CL3,

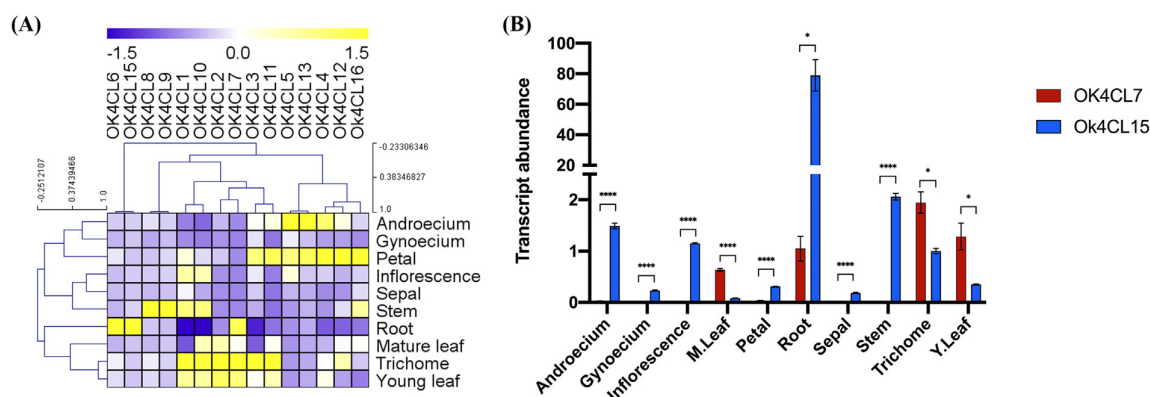


Fig. 2. 4CL gene expression analysis in different tissues of *O. kilimandscharicum*. A. Gene expression analysis of putative fifteen *Ok4CLs* in ten different tissues of *O. kilimandscharicum*, such as androecium, gynoecium, inflorescence, mature leaf, petal, root, sepal, stem, trichome, and young leaf. B. *Ok4CL7* and *Ok4CL15* expression analyses in all ten tissue types of *O. kilimandscharicum*.

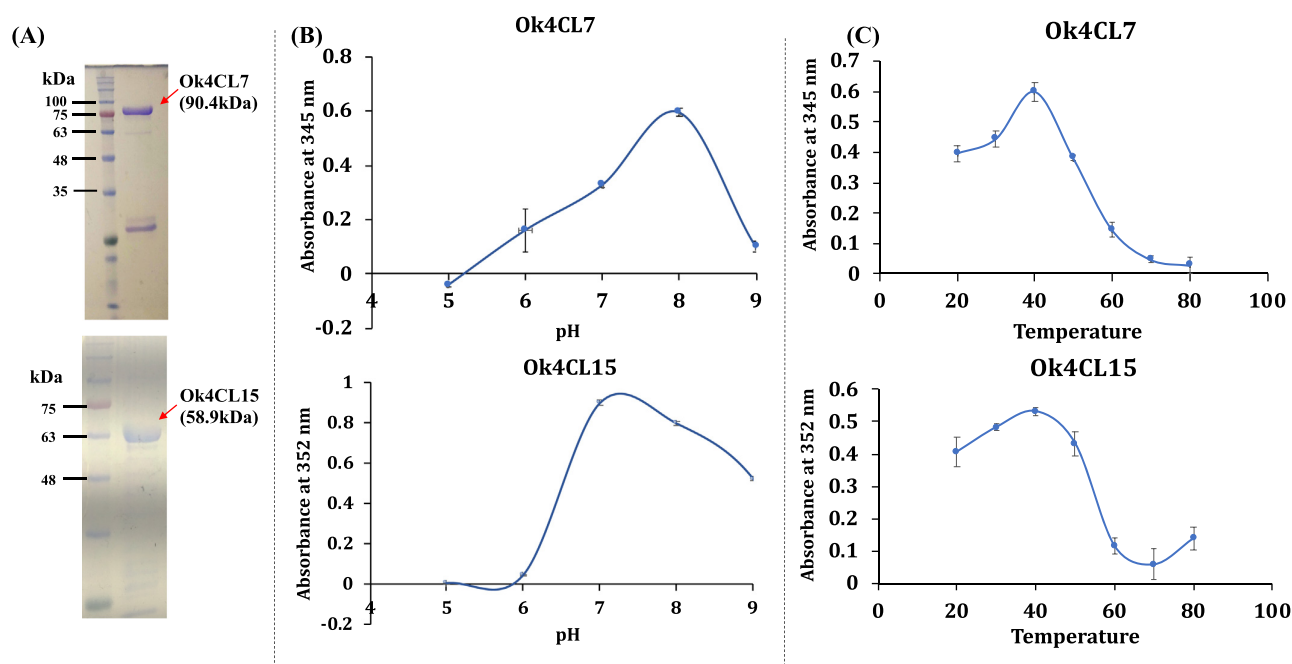


Fig. 3. Determination of optimal conditions for 4CL7 and 15 activities. A. Recombinantly expressed Ok4CL7 and -15 showed the predominant band on 12% SDS PAGE. Ok4CL7 and -15 proteins were observed at expected size 90.4 kDa and 58.9 kDa, respectively. B. Determination of optimum pH for recombinantly expressed Ok4CL7 and -15. Optimum activity of OK4CL7 was observed at pH 8, while Ok4CL15 showed optimum activity at pH 7. C. Determination of optimal temperature for recombinantly expressed Ok4CL7 and -15. Optimal activities of OK4CL7 and -15 were observed at 40 °C.

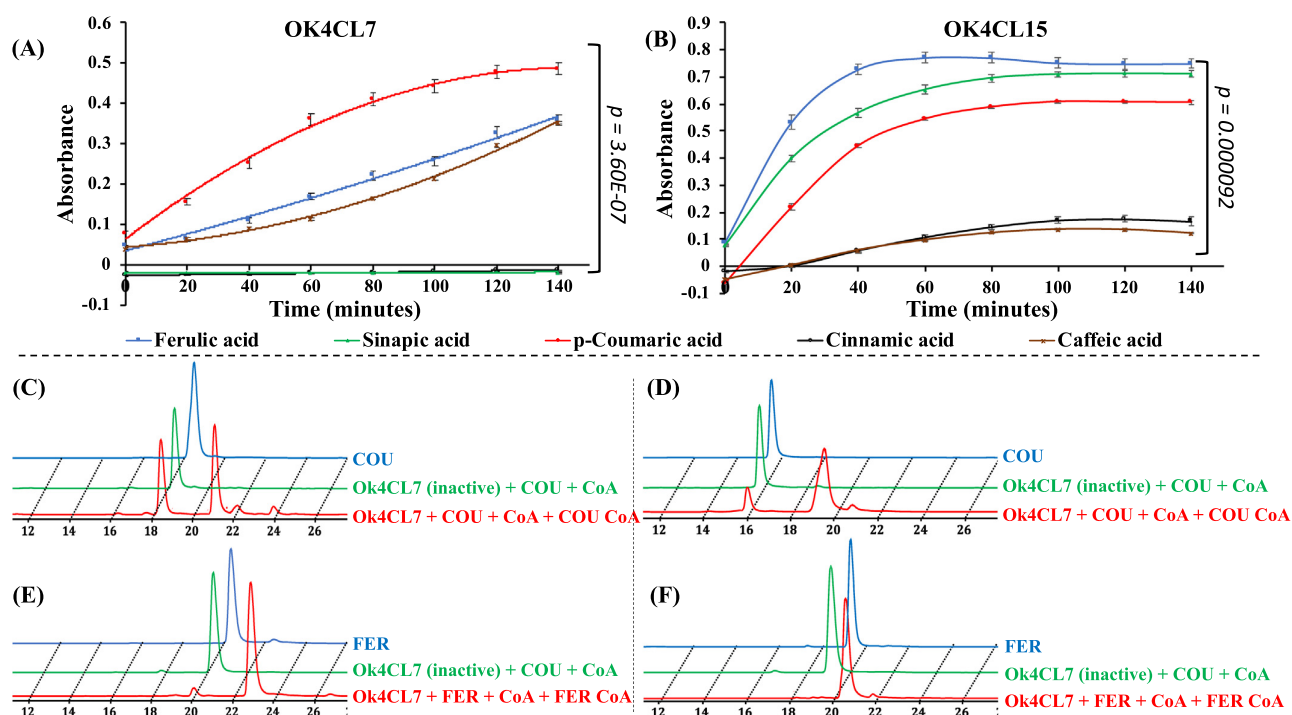


Fig. 4. The activity of 4CL7 and 15 with different substrates. Activity assay of recombinant Ok4CL7 and -15 studied with five substrates (cinnamic acid, caffeic acid, *p*-coumaric acid, ferulic acid, and sinapic acid). Product formation was analyzed by measuring the reaction mixture absorption at the corresponding λ max of CoA product. A. Ok4CL7 showed activity with *p*-coumaric acid (*p*-coumaroyl CoA: 333 nm), ferulic acid (feruloyl CoA 345 nm), and caffeic acid (caffeoyl CoA 345 nm). B. Ok4CL15 showed activity with *p*-coumaric acid, ferulic acid, and sinapic acid (sinapoyl CoA 352 nm). Analysis of enzyme assay reaction on HPLC. Three runs of each enzyme with each substrate. i) standard substrate; ii) heat-inactivated enzyme and all other components; iii) reaction mixture with the active enzyme. C. Enzyme activity with *p*-coumaric acid (COU), *p*-coumaroyl CoA product peak observed at RT 21.05. D. Enzyme activity with caffeic acid (CAF), caffeoyl CoA product peak observed at RT 19.02; E. Enzyme activity with ferulic acid (FER), feruloyl CoA product peak observed at RT 21.38. F. Enzyme activity with sinapic acid (SIN), sinapoyl CoA product peak observed at RT 20.7.

Ok4CL4 and Ok4CL5 contain fatty acyl CoA synthetase (FACS) domain, proposing their potential role in fatty acid metabolism (Table 1). Ok4CL7, 8 and 15 have 4CL, CaiC and AMP binding domain (Fig. 1C).

3.2. Ok4CL7 and 15 exhibits differential expression in various tissues

Comparative analysis of gene expression values indicates differential and tissue-specific expression of 4CLs in *O. kilimandscharicum* (Fig. 2A). Most of the 4CLs were expressed in the trichome, while the least expression was observed in floral organs. Ok4CL11 was highly expressed in young and mature leaves, trichomes, androecium and petal. Whereas, Ok4CL8 had high expression in stem tissue. Ok4CL5 and Ok4CL15 were abundant in inflorescence and root tissues (Fig. 2A). Ok4CL7 is highly expressed in trichome, young leaf, mature leaf and root, while Ok4CL15 shows expression in most of the *O. kilimandscharicum* tissues, with the highest levels in the root (Fig. 2B).

The expression of 4CL isoforms studied in several plants [16,51,53]. In some plants, 4CL isoforms are expressed across all the tissues, while others have tissue-specific expressions, e.g. *Hibiscus cannabiuns*, *Isatis indigotica*, and *Rubus idaeus* [53–55]. Expression of 4CL isoforms also varies during plant development [43,54]. In *H. cannabiuns*, 4CL expression in stem increases from 2 weeks to 20 weeks of development [54]. Phenylpropanoid pathway products protect plants from wounding, irradiation with UV light or pathogen attack [51]. In *A. thaliana*, expression of At4CL1 and At4CL2 were upregulated by wounding and the treatment with methyl jasmonate [51]. At4CL1 and At4CL2 expression were upregulated by *Peronospora parasitica* infection and wounding [12].

Tissue-specific expression of Ok4CLs might indicate their engagement in the specialized function through variable substrate specificities.

3.3. Recombinant Ok4CL7 and Ok4CL15 proteins differ in biochemical properties and substrate specificity

Recombinant Ok4CL7 and Ok4CL15 were purified separated on 12% SDS-PAGE (Fig. 3A). The activity of recombinant Ok4CL7 and Ok4CL15 proteins were studied with cinnamic acid, *p*-coumaric acid, caffeic acid, ferulic acid and sinapic acid. A single point assay revealed that Ok4CL7 and Ok4CL15 preferred ferulic acid and sinapic acid, respectively. Moreover, optimum pH and temperature for Ok4CL7 were determined using ferulic acid as a substrate, while sinapic acid was used as a substrate for Ok4CL15. The observed optimum pH for Ok4CL7 and Ok4CL15 activity were pH 8 and 7, respectively (Fig. 3B). Both Ok4CL7 and Ok4CL15 show maximum activity at 40 °C (Fig. 3C). Out of the five substrates, Ok4CL7 utilized *p*-coumaric acid, caffeic acid and ferulic acid (Fig. 4A). Ok4CL15 showed activity with *p*-coumaric acid, ferulic acid and sinapic acid (Fig. 4B). Both recombinant 4CLs proteins utilized *p*-coumaric acid and ferulic acid and inactive against cinnamic acid.

Ok4CL7 preferred substrates were caffeic acid, followed by *p*-coumaric acid and ferulic acid. Ok4CL15 has more preference for sinapic acid than *p*-coumaric acid and ferulic acid. The products formed were characterized and validated using HPLC (Fig. 4C, D, E, F). The peaks of *p*-coumaroyl CoA, feruloyl CoA, caffeoyl CoA and sinapoyl CoA were observed at retention time 21.05, 19.02, 21.38 and 20.07 mins, respectively. The product was purified using HPLC and further confirmed by mass spectrometry using

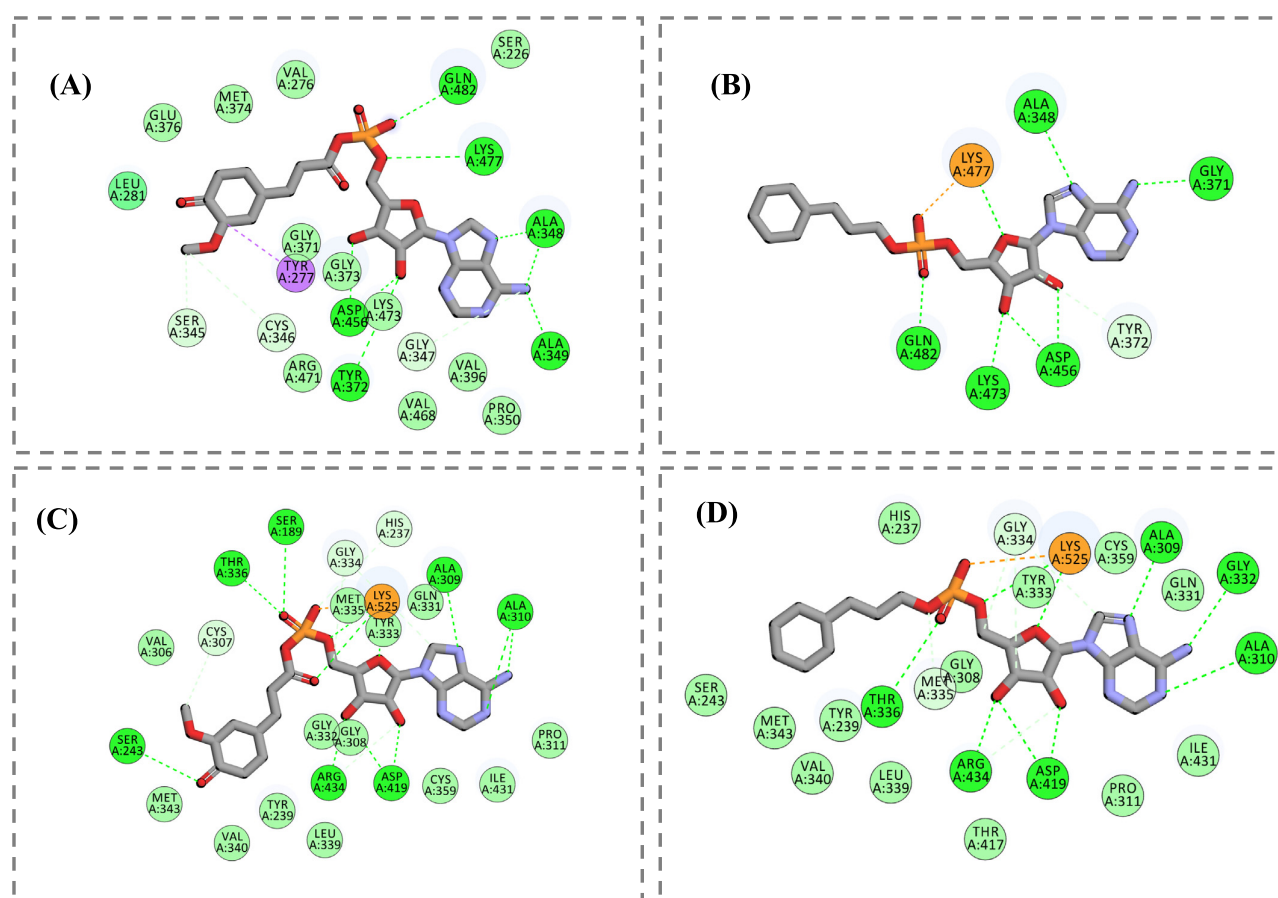


Fig. 5. Substrate preferences of Ok4CLs. Molecular interaction of substrate binding pocket of Ok4CL7 with substrate-specific functional groups. A. Feruloyl adenylate contributes to forming the salt bridge with Tyr277, hydrogen bonds with Ser345 and Cys346. B. Unfavorable substrate cinnamoyl adenylate showed the absence of substrate-specific interactions. C. Similarly, in Ok4CL15, feruloyl adenylate functional groups form hydrogen bonds with Ser243 and Cys307, whereas D. Cinnamoyl adenylate does not have any specific interaction.

accurate mass and fragmentation pattern match (Supplementary Fig. S2). Different peaks of *p*-coumaroyl CoA (912.5418 Da for at 1.73 retention time) and feruloyl CoA (942.1552 Da at 3.89 retention time) were observed (Supplementary Fig. S2).

Substrate specificities among the 4CL isoforms define their role in plant growth, development and defence [16,56]. Sinapic acid converting 4CLs are very rare in angiosperms [57]. Alternative pathways for sinapoyl alcohol formation is reported in some plants, including *A. thaliana* [43,58]. In *A. thaliana*, syringyl lignins and sinapate/sinapyl alcohol are derived from methylation of 5-hydroxy ferulic acid instead of sinapic acid or sinapoyl CoA [43]. Sinapic acid converting isoforms of 4CL4 are also characterized from *A. thaliana* and *Robinia pseudoacacia* [52,57]. The isoform Rp4CL1 utilizes *p*-coumarate as a substrate, preferably over ferulate and sinapate. Rp4CL2 and Rp4CL3 can activate sinapate along with caffeate and *p*-coumarate. Crude protein extracts from the shoot and developing xylem showed similar substrate preferences as Rp4CL2 and Rp4CL3. This suggests that sinapate-activating isoforms might be involved in lignin biosynthesis [57].

3.4. Activity and specificity determining residues in the binding pocket of Ok4CLs are conserved with other plants

Interaction analysis of Ok4CL7 and 15 docked complexes illustrated the functional groups of substrates and their intermediates form several non-covalent interactions with binding site residues. Ok4CL7 has a binding score of -6.3 and -5.4 kcal/mol with feruloyl and cinnamoyl adenylate, respectively. While, Ok4CL15 has a binding score of -6.7 and -5.8 kcal/mol with feruloyl and cinnamoyl adenylate, respectively. Binding energy score pattern corroborating with the in vitro activity studies. It predicts that feruloyl adenylate is a more preferred substrate over cinnamoyl adenylate in Ok4CL7 and 15. In Ok4CL7, Asp456 and Lys473 interact with the ribose, whereas Lys477 and Gln482 interact with α -phosphate. Feruloyl adenylate (preferred substrate) showed Pi-sigma interaction with Tyr277 and carbon-hydrogen bond with Ser345 and Cys346. These additional interactions are absent in cinnamoyl adenylate; hence, it is not preferred as a substrate by Ok4CL7 (Fig. 5A, B). In Ok4CL15, Asp419 and Arg434 interact with ribose, whereas Lys525 forms a salt bridge with α -phosphate of adenylate. Ser243 and Cys336 make polar contacts with functional groups of feruloyl adenylate. These substrate-specific interactions are absent in the case of the Ok4CL15 and cinnamoyl adenylate complex (Fig. 5C, D).

Most of the substrate-binding site residues of Ok4CL7 and 15 are conserved with 4CLs from other plants. In *Nicotiana tabacum* 4CL, lysine interacts with α -phosphate of adenylate. Mutation of this lysine residue to alanine results in complete loss of activity [14]. Similarly, the indispensability for tryptophan present near the substrate has been validated in *N. tabacum* by mutagenesis. It has been observed that this mutant showed decreased activity. Tyr277 and Tyr239 also showed proximity with a substrate suggesting likeness in their functionality with *N. tabacum* and *P. tomentosa* 4CLs. This interaction analysis showed high corroboration with studies and highlights conservation of the substrate-specificity mechanism in different plants 4CLs.

4. Conclusion

Two isoforms of 4CL were functionally characterized from *O. kilimandscharicum*. Gene expressions of Ok4CL7 and 15 were high in trichomes and root tissues compared to other tissue types tested, respectively. The optimum pH required for Ok4CL7 and Ok4CL15 activities was found to be 8 and 7, respectively. Both 4CLs showed maximum activities at 40 °C. Ok4CL7 catalyzed the conversion of *p*-coumaric acid, ferulic acid and caffeic acid into their corresponding CoA esters. Furthermore, Ok4CL15 could catalyze the conversion of *p*-coumaric acid, ferulic acid and sinapic acid into respective CoA esters. However, both 4CLs were unable to utilize cinnamic acid. Specific non-covalent interactions between the substrate and binding pockets of Ok4CLs are essential for

their catalytic activities. It appears that Ok4CL15 could show high expression in root tissues and catalytic activity to convert sinapic acid to sinapoyl CoA. This study's results allow us to understand the function of 4CL isoforms of *O. kilimandscharicum* and its potential role in the diversification of the phenylpropanoid pathway.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2021.03.129>.

CRedit authorship contribution statement

APG designed the experiments. SGL performed the wet lab experiment, RSJ performed in silico analysis and YK analyzed OK4CL enzyme assay products on LCMS. SGL, RSJ and APG wrote the manuscript and all authors approved content.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgement

SGL acknowledges the Council of Scientific and Industrial Research (CSIR), New Delhi, India for research fellowships. Authors would like to thank Dr. Kirtikumar Kondhare for editorial assistance.

References

- [1] J. Gui, J. Shen, L. Li, Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice, *Plant Physiol.* 157 (2011) 574–586, <https://doi.org/10.1104/pp.111.178301>.
- [2] T. Vogt, Phenylpropanoid biosynthesis, *Mol. Plant* 3 (2010) 2–20, <https://doi.org/10.1093/mp/ssp106>.
- [3] F.F.P. Kollmann, W.A. Côté, W.A. Côté, The structure of wood and the wood cell wall, *Principles of Wood Science and Technology*, Springer Berlin Heidelberg 1968, pp. 1–54, https://doi.org/10.1007/978-3-642-87928-9_1.
- [4] R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan, Lignin biosynthesis and structure, *Plant Physiol.* 153 (2010) 895–905, <https://doi.org/10.1104/pp.110.155119>.
- [5] W. Boerjan, J. Ralph, M. Baucher, Lignin biosynthesis, *Annu. Rev. Plant Biol.* 54 (2003) 519–546, <https://doi.org/10.1146/annurev.arplant.54.031902.134938>.
- [6] R. Whetten, R. Sederoff, Lignin biosynthesis, *Plant Cell* 7 (1995) 1001–1013.
- [7] Y. Deng, S. Lu, Biosynthesis and regulation of phenylpropanoids in plants, *Crit. Rev. Plant Sci.* 36 (2017) 257–290, <https://doi.org/10.1080/07352689.2017.1402852>.
- [8] M. Dastmalchi, S. Dhaubhadel, Proteomic insights into synthesis of isoflavonoids in soybean seeds, *PROTEOMICS* 15 (2015) 1646–1657, <https://doi.org/10.1002/pmic.201400444>.
- [9] S. Hassan, U. Mathesius, The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions, *J. Exp. Bot.* 63 (2012) 3429–3444, <https://doi.org/10.1093/jxb/err430>.
- [10] A. Shah, D.L. Smith, Flavonoids in agriculture: chemistry and roles in, biotic and abiotic stress responses, and microbial associations, *Agronomy* 10 (2020) 1209, <https://doi.org/10.3390/agronomy10081209>.
- [11] M.-Z. Shi, D.-Y. Xie, Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*, *Recent Patents Biotechnol.* 8 (2014) 47–60, <https://doi.org/10.2174/1872208307666131218123538>.
- [12] J. Ehlting, D. Büttner, Q. Wang, C.J. Douglas, I.E. Somssich, E. Kombrink, Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms, *Plant J.* 19 (1999) 9–20, <https://doi.org/10.1046/j.1365-3113.1999.00491.x>.
- [13] S. Schmelz, J.H. Naismith, Adenylate-forming enzymes, *Curr. Opin. Struct. Biol.* 19 (2009) 666–671, <https://doi.org/10.1016/j.sbi.2009.09.004>.
- [14] Z. Li, S.K. Nair, Structural basis for specificity and flexibility in a plant 4-coumarate: CoA ligase, *Structure* 23 (2015) 2032–2042, <https://doi.org/10.1016/j.str.2015.08.012>.
- [15] K. Schneider, K. Hövel, K. Witzel, B. Hamberger, D. Schomburg, E. Kombrink, H.P. Stübtle, The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 8601–8606, <https://doi.org/10.1073/pnas.1430550100>.
- [16] S.G. Lavhale, R.M. Kalunke, A.P. Giri, Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants, *Planta* 248 (2018) 1063–1078, <https://doi.org/10.1007/s00425-018-2965-z>.
- [17] H. Sun, Y. Li, S. Feng, W. Zou, K. Guo, C. Fan, S. Si, L. Peng, Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice, *Biochem. Biophys. Res. Commun.* 430 (2013) 1151–1156, <https://doi.org/10.1016/j.bbrc.2012.12.019>.

- [18] C.-H. Zhang, T. Ma, W.-C. Luo, J.-M. Xu, J.-Q. Liu, D.-S. Wan, Identification of 4CL genes in desert poplars and their changes in expression in response to salt stress, *Genes* 6 (2015) 901–917, <https://doi.org/10.3390/genes6030901>.
- [19] D. Lee, M. Ellard, L.A. Wanner, K.R. Davis, C.J. Douglas, The *Arabidopsis thaliana* 4-coumarate:CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA, *Plant Mol. Biol.* 28 (1995) 871–884, <https://doi.org/10.1007/BF00042072>.
- [20] B. Abdollahi Mandoulakani, E. Eyvazpour, M. Ghadimzadeh, The effect of drought stress on the expression of key genes involved in the biosynthesis of phenylpropanoids and essential oil components in basil (*Ocimum basilicum* L.), *Phytochemistry* 139 (2017) 1–7, <https://doi.org/10.1016/j.phytochem.2017.03.006>.
- [21] M. Becker-Andre, P. Schulze-Lefert, K. Hahlbrock, Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato, *J. Biol. Chem.* 266 (13) (1991) 8551–8559.
- [22] C.H. An, K.W. Lee, S.H. Lee, Y.J. Jeong, S.G. Woo, H. Chun, Y. il Park, S.S. Kwak, C.Y. Kim, Heterologous expression of lIbMYB1a by different promoters exhibits different patterns of anthocyanin accumulation in tobacco, *Plant Physiol. Biochem.* 89 (2015) 1–10, <https://doi.org/10.1016/j.plaphy.2015.02.002>.
- [23] S. Gao, H.N. Yu, R.X. Xu, A.X. Cheng, H.X. Lou, Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*, *Phytochemistry* 111 (2015) 48–58, <https://doi.org/10.1016/j.phytochem.2014.12.017>.
- [24] H. Jin, Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*, *EMBO J.* 19 (2000) 6150–6161, <https://doi.org/10.1093/emboj/19.22.6150>.
- [25] J.H. Jung, B. Kannan, H. Dermawan, G.W. Moxley, F. Altpeter, Precision breeding for RNAi suppression of a major 4-coumarate:coenzyme A ligase gene improves cell wall saccharification from field grown sugarcane, *Plant Mol. Biol.* 92 (2016) 505–517, <https://doi.org/10.1007/s11103-016-0527-y>.
- [26] C. Wang, S. Zhi, C. Liu, F. Xu, A. Zhao, X. Wang, Y. Ren, Z. Li, M. Yu, Characterization of stilbene synthase genes in mulberry (*Morus atropurpurea*) and metabolic engineering for the production of resveratrol in *Escherichia coli*, *J. Agric. Food Chem.* 65 (2017) 1659–1668, <https://doi.org/10.1021/acs.jafc.6b05212>.
- [27] J. Yang, F. Chen, O. Yu, R.N. Beachy, Controlled silencing of 4-coumarate:CoA ligase alters lignocellulose composition without affecting stem growth, *Plant Physiol. Biochem.* 49 (2011) 103–109, <https://doi.org/10.1016/j.plaphy.2010.10.004>.
- [28] W.J. Hu, S.A. Harding, J. Lung, J.L. Popko, J. Ralph, D.D. Stokke, C.J. Tsai, V.L. Chiang, Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees, *Nat. Biotechnol.* 17 (1999) 808–812, <https://doi.org/10.1038/11758>.
- [29] Y. Wang, H. Yi, M. Wang, O. Yu, J.M. Jez, Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-CoA ligase::stilbene synthase, *J. Am. Chem. Soc.* 133 (2011) 20684–20687, <https://doi.org/10.1021/ja2085993>.
- [30] Y. Katsuyama, M. Matsuzawa, N. Funa, S. Horinouchi, Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway, *Microbiology* 154 (2008) 2620–2628, <https://doi.org/10.1099/mic.0.2008/018721-0>.
- [31] P. Singh, R.M. Kalunke, A.P. Giri, Towards comprehension of complex chemical evolution and diversification of terpene and phenylpropanoid pathways in *Ocimum* species, *RSC Adv.* 5 (2015) 106886–106904, <https://doi.org/10.1039/C5RA16637C>.
- [32] P. Singh, R.M. Kalunke, A. Shukla, O. Tzfadia, H.v. Thulasiram, A.P. Giri, Biosynthesis and tissue-specific partitioning of camphor and eugenol in *Ocimum kilimandscharicum*, *Phytochemistry* 177 (2020), 112451, <https://doi.org/10.1016/j.phytochem.2020.112451>.
- [33] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644–652, <https://doi.org/10.1038/nbt.1883>.
- [34] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410, [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- [35] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.* 30 (2013) 2725–2729, <https://doi.org/10.1093/molbev/mst197>.
- [36] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (2004) 1792–1797, <https://doi.org/10.1093/nar/gkh340>.
- [37] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, *Nucleic Acids Res.* 48 (2020) D265–D268, <https://doi.org/10.1093/nar/gkz991>.
- [38] J. Ren, L. Wen, X. Gao, C. Jin, Y. Xue, X. Yao, DOG 1.0: illustrator of protein domain structures, *Cell Res.* 19 (2009) 271–273, <https://doi.org/10.1038/cr.2009.6>.
- [39] A. Anand, R.H. Jayaramaiah, S.D. Beedkar, P.A. Singh, R.S. Joshi, F.A. Mulani, B.B. Dholakia, S.A. Puneekar, W.N. Gade, H.v. Thulasiram, A.P. Giri, Comparative functional characterization of eugenol synthase from four different *Ocimum* species: implications on eugenol accumulation, *Biochim. Biophys. Acta, Proteins Proteomics* 1864 (2016) 1539–1547, <https://doi.org/10.1016/j.bbapap.2016.08.004>.
- [40] T.L.G. Schatz, H. Grisebach, Enzymic synthesis of lignin precursors purification and properties of 4-coumarate: CoA ligase from cambial sap of spruce (*Picea abies* L.), *Eur. J. Biochem.* 123 (3) (1982) 583–586.
- [41] J. Stöckigt, M.H. Zenk, Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives, *Zeitschrift Fur Naturforschung - Section C J. Biosci.* 30 (1975) 352–358, <https://doi.org/10.1515/znc-1975-5-609>.
- [42] K.H. Knobloch, K. Hahlbrock, 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Partial purification, substrate specificity, and further properties, *Arch. Biochem. Biophys.* 184 (1977) 237–248, [https://doi.org/10.1016/0003-9861\(77\)90347-2](https://doi.org/10.1016/0003-9861(77)90347-2).
- [43] M.A. Costa, D.L. Bedgar, S.G.A. Moinuddin, K.W. Kim, C.L. Cardenas, F.C. Cochrane, J.M. Shockey, G.L. Helms, Y. Amakura, H. Takahashi, J.K. Milhollan, L.B. Davin, J. Browse, N.G. Lewis, Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation, *Phytochemistry*, Elsevier Ltd 2005, pp. 2072–2091, <https://doi.org/10.1016/j.phytochem.2005.06.022>.
- [44] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791, <https://doi.org/10.1002/jcc.21256>.
- [45] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2) (2010) 455–461, <https://doi.org/10.1002/jcc.21334>.
- [46] C. Lindermayr, B. Möllers, J. Fliegmann, A. Uhlmann, F. Lottspeich, H. Meimberg, J. Ebel, Divergent members of a soybean (*Glycine max* L.) 4-Coumarate:coenzyme A ligase gene family: primary structures, catalytic properties, and differential expression, *Eur. J. Biochem.* 269 (2002) 1304–1315, <https://doi.org/10.1046/j.1432-1033.2002.02775.x>.
- [47] Z. Jin, J. Wungsintaweekul, S.H. Kim, J.H. Kim, Y. Shin, D.K. Ro, S.U. Kim, 4-Coumarate:coenzyme A ligase isoform 3 from *Piper nigrum* (Pn4CL3) catalyzes the CoA thioester formation of 3,4-methylenedioxybenzyl and piperic acids, *Biochem. J.* 477 (2020) 61–74, <https://doi.org/10.1042/BCJ20190527>.
- [48] H.C. Chen, J. Song, J.P. Wang, Y.C. Lin, J. Ducoste, C.M. Shuford, J. Liu, Q. Li, R. Shi, A. Nepomuceno, F. Isik, D.C. Muddiman, C. Williams, R.R. Sederoff, V.L. Chiang, Systems biology of lignin biosynthesis in *Populus trichocarpa*: Heteromeric 4-coumaric acid: coenzyme A ligase protein complex formation, regulation, and numerical modeling, *Plant Cell* 26 (2014) 876–893, <https://doi.org/10.1105/tpc.113.119685>.
- [49] D. Cukovic, J. Ehlting, J.A. VanZiffle, C.J. Douglas, Structure and evolution of 4-coumarate: coenzyme A ligase (4CL) gene families, *Biol. Chem.* 382 (2001) 645–654, <https://doi.org/10.1515/BC.2001.076>.
- [50] J. Raes, A. Rohde, J.H. Christensen, Y. van de Peer, W. Boerjan, Genome-wide characterization of the lignification toolbox in *Arabidopsis*, *Plant Physiol.* 133 (2003) 1051–1071, <https://doi.org/10.1104/pp.103.026484>.
- [51] D. Lee, C.J. Douglas, Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins, *Plant Physiol.* 112 (1996) 193–205, <https://doi.org/10.1104/pp.112.1.193>.
- [52] B. Hamberger, K. Hahlbrock, The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 2209–2214, <https://doi.org/10.1073/pnas.0307307101>.
- [53] P. Di, Y. Hu, H. Xuan, Y. Xiao, J. Chen, L. Zhang, W. Chen, Characterization and the expression profile of 4-coumarate: CoA ligase (li4CL) from hairy roots of *Isatis indigotica*, *Afr. J. Pharm. Pharmacol.* 6 (2012) 2166–2175, <https://doi.org/10.5897/AJPP12.852>.
- [54] M. Emran, K. Chowdhury, B. Choi, B.-K. Cho, J.B. Kim, S.U. Park, S. Natarajan, H.-S. Lim, H. Bae, Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions, *Plant Omics* 6 (4) (2013) 254–262.
- [55] A. Kumar, B.E. Ellis, 4-Coumarate:CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression, *Plant Mol. Biol.* 51 (2003) 327–340, <https://doi.org/10.1023/A:1022004923982>.
- [56] Y. Li, J.J. Kim, L. Pysh, C. Chapple, Four isoforms of arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism, *Plant Physiol.* 169 (2015) 2409–2421, <https://doi.org/10.1104/pp.15.00838>.
- [57] K. Hamada, T. Nishida, K. Yamauchi, K. Fukushima, R. Kondo, Y. Tsutsumi, 4-Coumarate: coenzyme A ligase in black locust (*Robinia pseudoacacia*) catalyzes the conversion of sinapate to sinapoyl-CoA, *J. Plant Res.* 117 (2004) 303–310, <https://doi.org/10.1007/s10265-004-0159-1>.
- [58] L. Li, J.L. Popko, T. Umezawa, V.L. Chiang, 5-Hydroxyconiferyl aldehyde modulates enzymatic methylation for syringyl monolignol formation, a new view of monolignol biosynthesis in angiosperms, *J. Biol. Chem.* 275 (2000) 6537–6545, <https://doi.org/10.1074/jbc.275.9.6537>.



Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants

Santosh G. Lavhale¹ · Raviraj M. Kalunke¹ · Ashok P. Giri¹

Received: 29 March 2018 / Accepted: 30 July 2018 / Published online: 4 August 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Main conclusions The 4-coumarate-CoA ligases (4CL) contribute in channelizing flux of different phenylpropanoid biosynthetic pathways. Expression of 4CL is optimized at developmental stages and in response to environmental triggers such as biotic and abiotic stresses. The enzyme is valuable in metabolic pathway engineering for curcuminoids, resveratrol, biofuel production and nutritional improvement. Vigorous analysis of regulation at functional and expression level is obligatory to attain efficient commercial production of candidate metabolites using 4CL.

Phenylpropanoid pathway provides precursors for numerous secondary metabolites in plants. In this pathway, 4-coumarate-CoA ligase (EC 6.2.1.12, 4CL) is the main branch point enzyme which generates activated thioesters. Being the last enzyme of three shared common steps in general phenylpropanoid pathway, it contributes to channelize precursors for different phenylpropanoids. In plants, 4CL enzymes are present in multiple isoforms and encoded by small gene family. It belongs to adenylate-forming enzyme family and catalyzes the reaction that converts hydroxy or methoxy cinnamic acid derivatives to corresponding thioesters. These thioesters are further utilized for biosynthesis of phenylpropanoids, which are known for having numerous nutritional and medicinal applications. In addition, the 4CL enzymes have been characterized from various plants for their role in plant physiology or in biotic and abiotic stresses. Furthermore, specific isoforms are differentially regulated upon exposure to diverse stimuli leading to flux diversion toward the particular metabolite biosynthesis. Evolutionary studies showed that 4CL separately evolved after monocot and dicot segregation. Here, we provide a comprehensive review on 4CL, which includes evolution, function, gene/protein structure, role in metabolite biosynthesis and cellular partition, and their regulation. Based on the available data, we have explored the scope for pathway engineering by utilizing 4CL enzymes.

Keywords Plant secondary metabolites · Phenylpropanoids · 4-coumarate-CoA ligase · Metabolic engineering

Abbreviations

4CL	4-coumarate:CoA ligase	COMT	Caffeic acid <i>O</i> -methyltransferase
C3H	<i>p</i> -coumarate 3-hydroxylase	EOMT	Eugenol <i>O</i> -methyltransferase
C4H	Cinnamate 4-hydroxylase	F5H	Ferulate 5-hydroxylase
CAD	Cinnamyl alcohol dehydrogenase	CAD	Cinnamyl alcohol dehydrogenase
CCoAOMT	Caffeoyl-CoA <i>O</i> -methyltransferase	CHS	Chalcone synthase
PAL	Phenylalanine ammonia lyase	ANS, F5H	Ferulate 5-hydroxylase
CVOMT	Chavicol <i>O</i> -methyltransferase	HCT	<i>p</i> -hydroxycinnamoyl-CoA:quininate shikimate <i>p</i> -hydroxycinnamoyltransferase
		DFR	Dihydroflavonol 4-reductase

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00425-018-2965-z>) contains supplementary material, which is available to authorized users.

✉ Ashok P. Giri
ap.giri@ncl.res.in

¹ Plant Molecular Biology Unit, Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, Maharashtra 411008, India

Introduction

Plants have a diverse range of metabolites that are obligatory for normal growth, development, and reproduction. Functions played by metabolites have a very crucial role in plant life cycle which includes mechanical support, attraction of

pollinators and frugivores, protection from biotic and abiotic stresses, interaction with environment, allelopathy effect, etc. (Bennett and Wallsgrave 1994; Bolton 2009; Goufo et al. 2017; Ferrieri et al. 2015). Plant secondary metabolites are categorized based on chemical structure, composition, solubility and pathway in which they are synthesized. For example, on the basis of chemical structure three groups (1) terpenoids, (2) phenylpropanoids and (3) alkaloids are known. The recent advances in metabolite analysis technologies empower the study of metabolites on a large scale, as well as the derivatives of individual metabolite and metabolic pathways at cellular and organelle level (Kueger et al. 2012; Fukushima et al. 2009; Floros et al. 2017; Freund and Hegeman 2017). The number of secondary metabolites have been characterized (Supplementary material, Table S1) containing numerous valuable properties/applications such as color, flavor, and medicine (Korkina 2007; Tassis and O'Connor 2016; Citti et al. 2017).

Phenylpropanoids are synthesized from aromatic amino acids such as phenylalanine (Phe) and tyrosine (Tyr) via phenylpropanoid pathway (Herrmann and Weaver 1999; Tzin and Galili 2010; Anand et al. 2016). Along with these

aromatic precursor amino acids, tryptophan is also synthesized in plastid via shikimic acid pathway in plants (Schmid and Amrhein 1995; Herrmann and Weaver 1999; Tzin and Galili 2010). The shikimic acid pathway utilizes phosphoenolpyruvate from glycolysis and erythrose 4-phosphate from hexose monophosphate pathway to generate aromatic amino acids (Tohge et al. 2013). Furthermore, general phenylpropanoid pathway begins with oxidative deamination of Phe and Tyr by phenylalanine ammonia lyase (PAL) and by tyrosine ammonia lyase (TAL) which leads to formation of cinnamic acid and *p*-coumaric acid, respectively. In subsequent steps, cinnamic acid is converted to various methoxy and hydroxy derivatives such as *p*-coumaric acid, caffeic acid, ferulic acid, 5-hydroxyferulic acid and sinapic acid using hydroxylase and methyltransferase enzymes (Fraser and Chapple 2011; Fig. 1). These methoxy and hydroxy derivatives utilized to synthesize corresponding CoA esters by 4CL. Products of 4CL are subsequently used by various oxygenases, reductases and transferases for biosynthesis of lignin, flavonoids, anthocyanins, aurones, stilbenes, coumarins, suberin, cutin, sporopollenin, etc. (Vogt 2010). Thus, 4CL is one of the key branch point enzyme in the

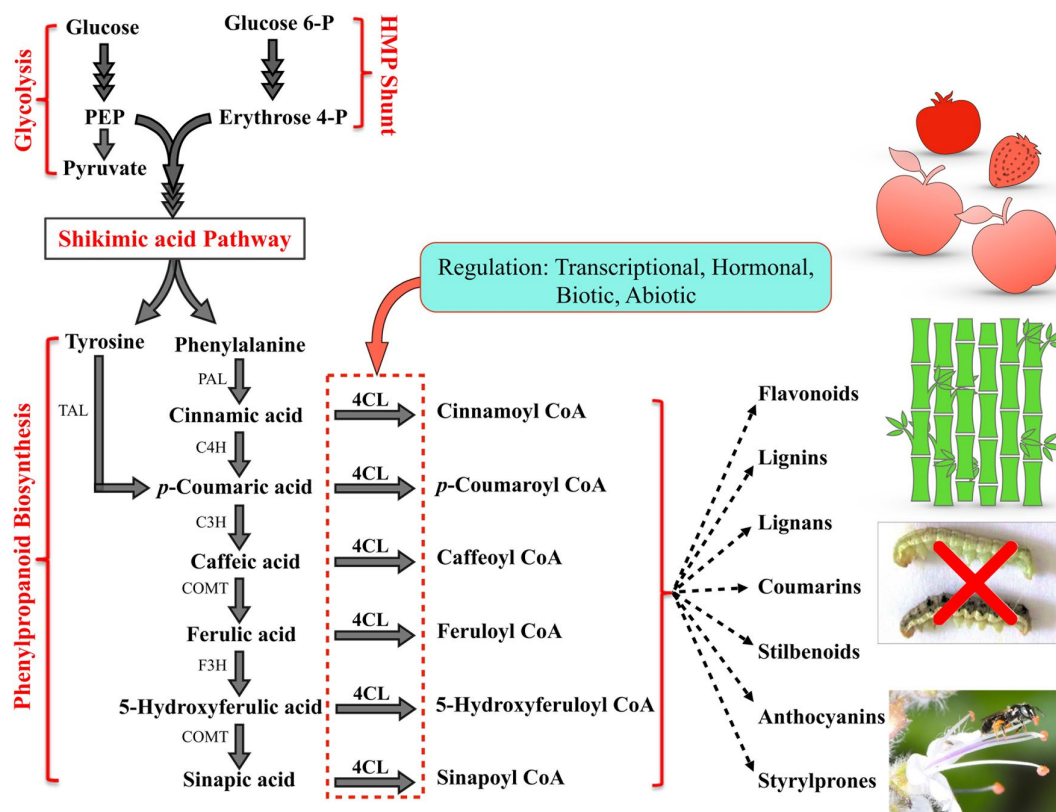


Fig. 1 General phenylpropanoid pathway and role of 4CL as a branch point enzyme. *HMP shunt* hexose monophosphate shunt, *4CL* 4-coumarate-CoA ligase, *PEP* phosphoenolpyruvate, *PAL* phenylala-

anine ammonia lyase, *TAL* tyrosine ammonia lyase, *C4H* cinnamate 4-hydroxylase, *C3H* *p*-coumarate 3-hydroxylase, *COMT* caffeic acid *O*-methyltransferase, *F3H* ferulate 5-hydroxylase

phenylpropanoid pathway. The gene family encoding these enzymes constitutes multiple isoforms of 4CL, which potentially perform diverse functions, which are yet to be fully characterized (Gui et al. 2011; Costa et al. 2005). However, literature reviews that are available on the phenylpropanoid pathway provide little information on 4CL (Vogt 2010; Fraser and Chapple 2011). Here, we provide a comprehensive review of the available information on 4CL and offer insights into its evolution and future potential as target enzyme for metabolic engineering.

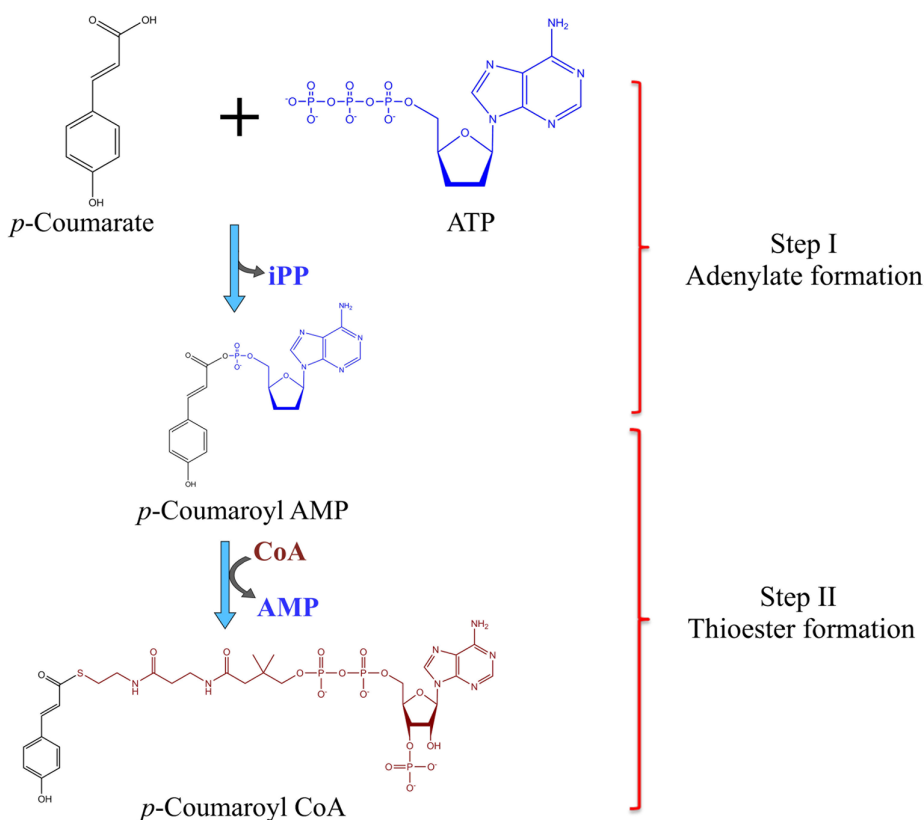
Unique catalytic features of plant 4CL enzymes

The 4CL catalyzes the ligation of coenzyme A (CoA) with cinnamic acid and its methoxy/hydroxy derivatives, such as caffeic acid and ferulic acid (Fig. 1). This enzyme belongs to the family of adenylate-forming enzymes, which has two conserved peptide motifs: box I (SSGTTGLPKGV) and box II (GEICIRG) (Allina et al. 1998; Ehling et al. 1999; Hu et al. 1998; Uhlmann and Ebel 1993; Schmelz and Naismith 2009). Box I is the adenosine monophosphate (AMP) binding domain and is conserved in all the members of adenylate-forming enzyme family. This AMP binding domain widely exists in proteins of all living organisms and are involved in a wide range of functions with fruit fly luciferase, gramicidin S synthetase, all types of CoA ligases, etc. (Fulda et al. 1994). While the function of Box II is unclear and not directly involved in catalysis (Stuible et al. 2000). Furthermore, the substrate-binding domain (SBD) is variable in different isoforms of 4CL. In *Arabidopsis thaliana*, the domain responsible for the substrate specificity determination was studied using domain-swapping approach for two isoforms, At4CL1 and At4CL2. Both enzymes utilize 4-coumarate as a substrate but only At4CL1 is capable of utilizing ferulate. Two adjacent domains known as SBD I and SBD II were identified. Out of these two, either one of the SBDs of At4CL1 is sufficient to recognize ferulate (Ehling et al. 2001). At4CL2 contains 12 amino acid residues in substrate-binding pocket (SBP), which were determined using crystal structure with the help of gramicidin S synthase homology model. Amino acids in SBP include Ile-252, Tyr-253, Asn-256, Met-293, Lys-320, Gly-322, Ala-323, Gly-346, Gly-348, Pro-354, Val-355, and Leu-356. These 12 amino acids are flanked by conserved box I and box II motifs (Schneider et al. 2003). The members of acyl-activating enzyme superfamily share little sequence identity, but they all have conserved box I. In *Arabidopsis*, this superfamily has 77 4CL genes and their phylogenetic analysis shows that they formed seven groups and all have the unique conserved box I (Shockey et al. 2003).

The probable function of any enzyme can be deciphered by analysis of its primary sequence and three-dimensional structure. Various features of 4CL structure were examined with the help of site-directed mutagenesis, domain exchange, mathematical modeling, crystallography, etc. Crystal structure of Pt4CL1 from *Populus tomentosa* was studied employing anomalous dispersion together with molecular replacement method using luciferase from firefly as a model. Pt4CL1 contains 536 amino acids and has two globular domains, one at N-terminal and other at C-terminal having 434 and 102 amino acids, respectively. N-terminal domain has three sub-domains: N1, N2 and N3. N1 and N2 are similar in structure and have 6 parallel along with 2 antiparallel β -sheets in center. The eight central β -sheets are flanked by 4 and 2 α -helices at each end. Three residues are crucial for Pt4CL1 catalytic activity (Lys-438, Gln-443, and Lys-523), while five residues for substrate binding (Tyr-236, Gly-306, Gly-331, Pro-337, and Val-338) (Hu et al. 2010). Based on crystal structure and reported sinapic acid converting 4CLs from other plants, At4CL2 has been modified using domain exchange and site-directed mutagenesis techniques to achieve improved sinapic acid conversion rate. This has been achieved in two steps (1) replacing substrate-binding pocket region of At4CL2 from Box I to Box II with substrate-binding pocket from *At3g21230* gene encoding 4CL-like protein and (2) deletion of either V355 or L356 but not both. They replace three amino acid residues (N256A, M293P and K320L) in At4CL2; this modified 4CL shows 30-fold improvement in the conversion rate of sinapic acid (Schneider et al. 2003).

Some 4CLs have inter-protein interactions for its regulatory function. For example, Ptr4CL3 and Ptr4CL5 isoforms from *Populus trichocarpa* interact with each other and form tetramer having three Ptr4CL3 and single Ptr4CL5. The mathematical model was developed to study their kinetics at different ratios and activation components. Using this model and experimental analysis like microdissection, co-immunoprecipitation, chemical cross-linking, bimolecular fluorescence complementation and mass spectrometry, it is concluded that Pt4CL5 has a regulatory role in Ptr4CL3-Ptr4CL5 tetrameric complex (Chen et al. 2014a). Mechanism of 4CL activity is also studied in Nt4CL2 isoform from *Nicotiana tabacum*. The crystal structure of Nt4CL2 reveals that this enzyme presents in two conformations during catalysis namely adenylate forming and thioester forming. During the catalysis process, the enzyme forms two conformations in which substrate is converted to adenylate intermediate and then to thioester form (Fig. 2; Li and Nair 2015).

Fig. 2 Mechanism of reaction catalyzed by 4CL. 4CL converts *p*-coumarate to *p*-coumaroyl CoA in two steps: step I, adenylate formation followed by thioester formation in step II



Evolution and diversity of plant 4CLs

Evolution of phenylpropanoid pathway was a crucial precondition for colonization of terrestrial plants on land. The products of the pathway such as flavonoids and lignin protect plants from UV light and provide structural support, respectively (Douglas 1996). Phylogenetic analysis of 4CL genes showed that the 4CL genes separated into two distinct clades from monocot and dicot plants. This is probably because 4CL genes evolve independently after the separation of monocots and dicots during the course of evolution. In dicot plants, 4CL genes grouped into two clusters: type I and type II (Fig. 3; Table 1). Type I cluster is mainly involved in monolignol biosynthesis whereas type II is involved in phenylpropanoid biosynthesis other than lignin. In *Arabidopsis thaliana*, At4CL1, At4CL2 and At4CL4 belong to type I and At4CL3 comes under type II cluster (Sun et al. 2013). Sun et al. (2013) categorized monocot 4CLs into two other clusters type III and IV with their similar functions as that of type I and II, respectively. In rice, Os4CL1, Os4CL3, Os4CL4 and Os4CL5 belong to type III, whereas Os4CL2 belongs to type IV. These different clusters formation suggest that the evolution of 4CL genes in dicot and monocot is an independent event after its separation. However, Pinta4CL3 isoform from *Pinus taeda* (a gymnosperm plant), is

phylogenetically closer to type II angiosperm 4CLs than to type III Pinta4CL1 (Chen et al. 2014b). This suggests that type I and II might have been diverged before the divergence of gymnosperm and angiosperm lineages (Cukovic et al. 2001). Amino acid sequences of At4CL1 and At4CL2 from *Arabidopsis thaliana* are more similar to each other (86% identical) as compared to At4CL3 sequence, which shows 71 and 73% identity with At4CL1 and At4CL2, respectively. These findings suggest that At4CL1 and At4CL2 are the recently evolved (Ehling et al. 1999). Similarly, *Salvia miltiorrhiza* genome have ten *Sm4CL* related genes, out of these, only three (*Sm4CL1*, *Sm4CL2* and *Sm4CL3*) were clustered with bonafide 4CLs in phylogenetic analysis. *Sm4CL1* and *Sm4CL2* clustered in type I are involved in lignin biosynthesis, whereas *Sm4CL3* clustered in type II 4CLs are involved in flavonoid biosynthesis (Wang et al. 2015).

Evolution of 4CL genes was studied in the 11 genera of *Larix* (family: Pinaceae) by comparing its copy number, GC content and codon usage, sequence divergence, and phylogenetic analysis (Wei and Wang 2004). The 4CL sequences were grouped into two paralogous clades, *4cIA* and *4cIB*. Both the clades have sequences from all studied 11 species of *Larix* genera, but some species have more than one sequence in the same clade. These findings support that two sister clades *4cIA* and *4cIB* is due to duplication in common

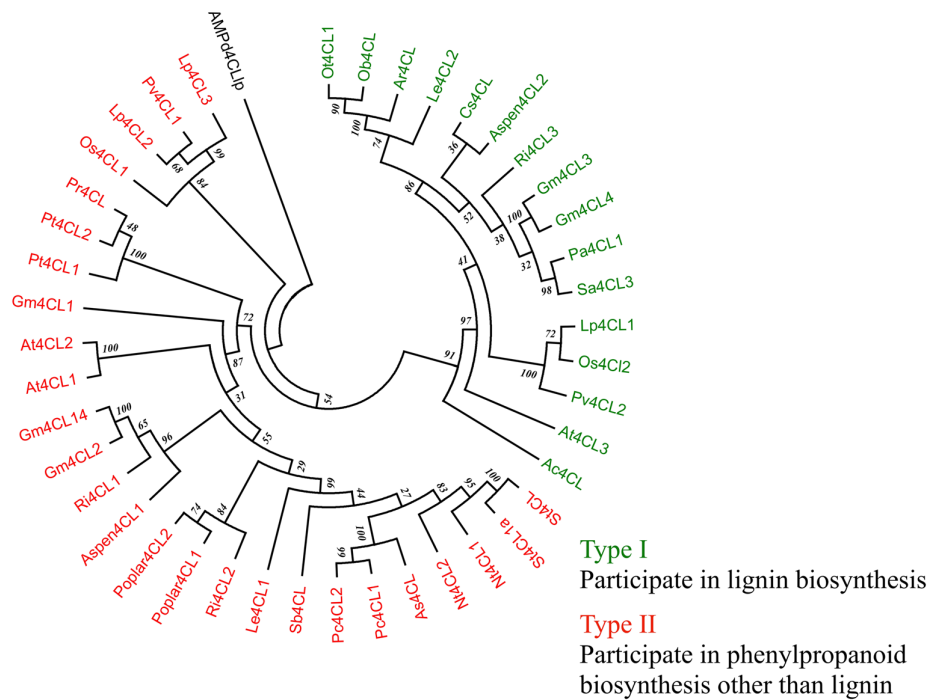


Fig. 3 Phylogenetic tree showing the relationship among different 4CL proteins. Type I cluster participates in lignin, while type II cluster participates in the biosynthesis of phenylpropanoids other than lignin. The phylogenetic tree was constructed by a neighbor-joining approach using the deduced amino acid sequences. Only functionally characterized 4CL protein sequences were selected for phylogenetic tree (Alignment, Muscle; Model, p-distance; Bootstrap value, 1000 replicates). 4CL 4-coumate: coenzyme A ligase. Accession number of 4CL amino acid sequence used to construct phylogenetic tree and name of the species: *Arabidopsis thaliana* AMPd4CL1p (AMP dependent 4CL like protein) (NM_123172); *A. thaliana* At4CL1 (U18675); *A. thaliana* At4CL2 (AF106086); *A. thaliana* At4CL3 (AF106088); *Agastache rugosa* Ar4CL (AY587891); *Allium cepa* Ac4CL (AY541033); *Angelica sinensis* As4CL (AMP18194.1); *Camellia sinensis* Cs4CL (DQ194356); *Glycine max* Gm4CL1 (AF279267); *G. max* Gm4CL2 (AF002259); *G. max* Gm4CL3 (AF002258); *G. max* Gm4CL4 (X69955); *G. max* Gm4CL14 (X69954); *Lithospermum erythrorhizon* Le4CL1

(D49366); *L. erythrorhizon* Le4CL2 (D49367); *Lolium perenne* Lp4CL1 (AF052221); *L. perenne* Lp4CL2 (AF052222); *L. perenne* Lp4CL3 (AF052223); *Nicotiana tabacum* Nt4CL1 (U50845); *N. tabacum* Nt4CL2 (U50846); *Oryza sativa* Os4CL1 (X52623); *O. sativa* Os4CL2 (L43362); *Ocimum sanctum* Ot4CL (HM990148); *Ocimum basilicum* Ob4CL (KC576841); *Panicum virgatum* Pv4CL1 (EU491511.1); *P. virgatum* Pv4CL2 (JF414903); *Petroselinum crispum* Pc4CL1 (X13324); *P. crispum* Pc4CL2 (X13325); *Pinus taeda* Pt4CL1 (U12012); *P. taeda* Pt4CL2 (U12013); *Pinus radiata* Pr4CL (ACF35279.1); *Populus hybrida* Poplar4CL1 (AF008184); *P. hybrida* Poplar4CL2 (AF008183); *Populus tremuloides* Aspen4CL1 (AF041049); *P. tremuloides* Aspen4CL2 (AF041050); *Prunus avium* Pa4CL1 (GU990523); *Rubus idaeus* Ri4CL1 (AF239687); *R. idaeus* Ri4CL2 (AF239686); *R. idaeus* Ri4CL3 (AF239685); *Scutellaria baicalensis* Sb4CL (BAD90937.1); *Sorbus aucuparia* Sa4CL3 (GU949553); *Solanum tuberosum* St4CL1 (M62755); St4CL1a (AF150686)

Table 1 Number of annotated and characterized 4CL genes in *Arabidopsis thaliana*, *Salvia miltiorrhiza*, *Populus trichocarpa*, *Populus pruinosa*, *Populus euphratica*, and *Salix suchowensis*

Plant	Annotated 4CLs	Characterized 4CLs	Class-I	Class-II	References
<i>Arabidopsis thaliana</i>	14	4	3	1	The Arabidopsis Genome Initiative (2000) and Costa et al. (2005)
<i>Salvia miltiorrhiza</i>	10	3	2	1	Wang et al. (2015)
<i>Populus trichocarpa</i>	20	6	5	1	Zhang et al. (2015)
<i>Populus pruinosa</i>	20	5	4	1	Zhang et al. (2015)
<i>Populus euphratica</i>	20	5	4	1	Zhang et al. (2015)
<i>Salix suchowensis</i>	12	4	3	1	Zhang et al. (2015)

ancestor of *Larix* thereby leading to co-existence of two alleles in all 11 species. After this duplication, they might have evolved separately because there is variation in values of mean distance of synonymous, non-synonymous, and nucleotide substitutions, and ratio of transitions to transversions. Thus, divergence of genera duplication and deletion in *4CL* gene family has happened, which might be potential reason for variation in sequence and number of alleles. Therefore, there might be two or three major *4CL* loci present in Pinaceae family (Wei and Wang 2004).

Here, we updated the list of annotated and characterized *4CL* genes in various plant species such as *Arabidopsis thaliana*, *Salvia miltiorrhiza*, *Populus trichocarpa*, *Populus pruinosa*, *Populus euphratica* and *Salix suchowensis* (Table 1). In the above-mentioned plant species, only 3–6 *4CLs* are functionally characterized out of 10–20 annotated *4CLs*.

Gene structure and genomic location of plant *4CLs*

The *4CL* genes have been studied from numerous plants and it is observed that they exist in small gene families, where they encode identical, nearly identical, or divergent proteins. In *Arabidopsis thaliana*, fourteen genes annotated as a putative *4CL* by in silico genome analysis, out of which only eleven were studied for functional characterization. While, the remaining three have peroxisome targeting sequence (based on C-terminal signal peptide analysis), which is absent in bonafide *4CLs* (Costa et al. 2005). Genomic location of *4CLs* in *Arabidopsis thaliana* showed a wide distribution on chromosome. Out of eleven functionally characterized *4CLs*, *At4CL1*, *At4CL3*, *At4CL9* and *At4CL10* are present on chromosome 1; *At4CL2*, *At4CL5* and *At4CL8* are on chromosome 3; *At4CL6* and *At4CL7* are on chromosome 4; and *At4CL4* and *At4CL11* are on chromosome 5. However, only *At4CL2* and *At4CL5* are clustered together on chromosome 3. Among these, *At4CL1*, *At4CL2*, *At4CL3* and *At4CL5* are functionally active (Costa et al. 2005). Sequence analysis of *At4CLs* shows that *At4CL1*, and *At4CL2* have three introns, whereas *At4CL3* have six introns (Fig. 4). Out of these three extra introns of *At4CL3*, the first intron resulted from interruption of the first exon and the remaining two introns resulted from interruption in the second exon of the *At4CL1* and *At4CL2* (Ehlting et al. 1999). There is variation in the sequence of 5' untranslated region and putative promoter region of *At4CL1*, *At4CL2*, *At4CL3*. Putative TATA boxes are located at – 124, – 103 and – 116 bp upstream of translation start codon ATG of *At4CL1*, *At4CL2* and *At4CL3*, respectively (Ehlting et al. 1999). For genes involved in phenylpropanoid pathway, like *PAL*, *C4H* and *4CL*, the promoter region has conserved box P, box A and box L. The promoter region of *At4CL1* and *At4CL2* contain

only boxes P and L, whereas *At4CL3* have all three boxes (Ehlting et al. 1999).

Consequently, in rice five *4CLs* are reported namely *OsCL1*, *OsCL2*, *OsCL3*, *OsCL4*, and *OsCL5*. Sequence analysis of these *4CLs* shows that *Os4CL1* and *Os4CL5* contain five exons and four introns. The position of introns is conserved; however, there are differences in intron length and sequence. *Os4CL2* has one additional intron, which results from the interruption of third exon of *Os4CL1* and *Os4CL5* (Sun et al. 2013; Fig. 4). In *Physcomitrella patens*, *Pp4CL1* and *Pp4CL4* have five exons and four introns, while *Pp4CL2* and *Pp4CL3* have four exons and three introns of similar size (Silber et al. 2008). Sequence analysis of various *4CL* isoforms shows that difference in the number of exons and introns has resulted from interruption of exon. Variation in the number of isoforms most probably resulted from gene duplication event (Ehlting et al. 1999; Fig. 4).

Contribution of *4CL* in lignification, flavonoid biosynthesis and stress response

The numerous products of phenylpropanoid pathway play a vital role in plants, viz., adaptation, growth and development, reproduction, protection against biotic and abiotic stresses, etc. Here we emphasize role of *4CL* in biosynthesis of lignins, flavonoids, phenylpropanoids and other compounds for mechanical support, and protection from biotic and abiotic stresses.

Lignification

Lignin is the second most abundant polymer after cellulose and present in secondary cell wall of all vascular plants. Complexity of lignin structure depends on the proportion of its constituent monolignol derivatives namely H, G, and S. These monolignol derivatives are synthesized via multiple routes in phenylpropanoid pathway (Naik et al. 2018). Using mathematical modeling, the effect of individual *4CL* and *4CL* complex (*Pt4CL3-Pt4CL5*; *Populus tremuloides*) on lignin content was studied for its different aspects such as steady state flux distribution, robustness and homeostatic properties. Results of this mathematical modeling suggest that robustness and stability of pathway for S and G monolignol biosynthesis increased in the presence of *Pt4CL3-Pt4CL5* complex (Naik et al. 2018). In aspen (*Populus tremuloides*), *Pt4CL1* gene is specifically expressed in lignin-containing tissue such as xylem, whereas *Pt4CL2* is expressed in epidermal layers of stem and leaf. *Pt4CL2* shows highest activity with 5-hydroxy ferulic acid, whereas *Pt4CL2* is inactive with 5-hydroxyferulic acid and have highest utilization rate with coumarate. The compartmentalized expression and substrate preferences suggest

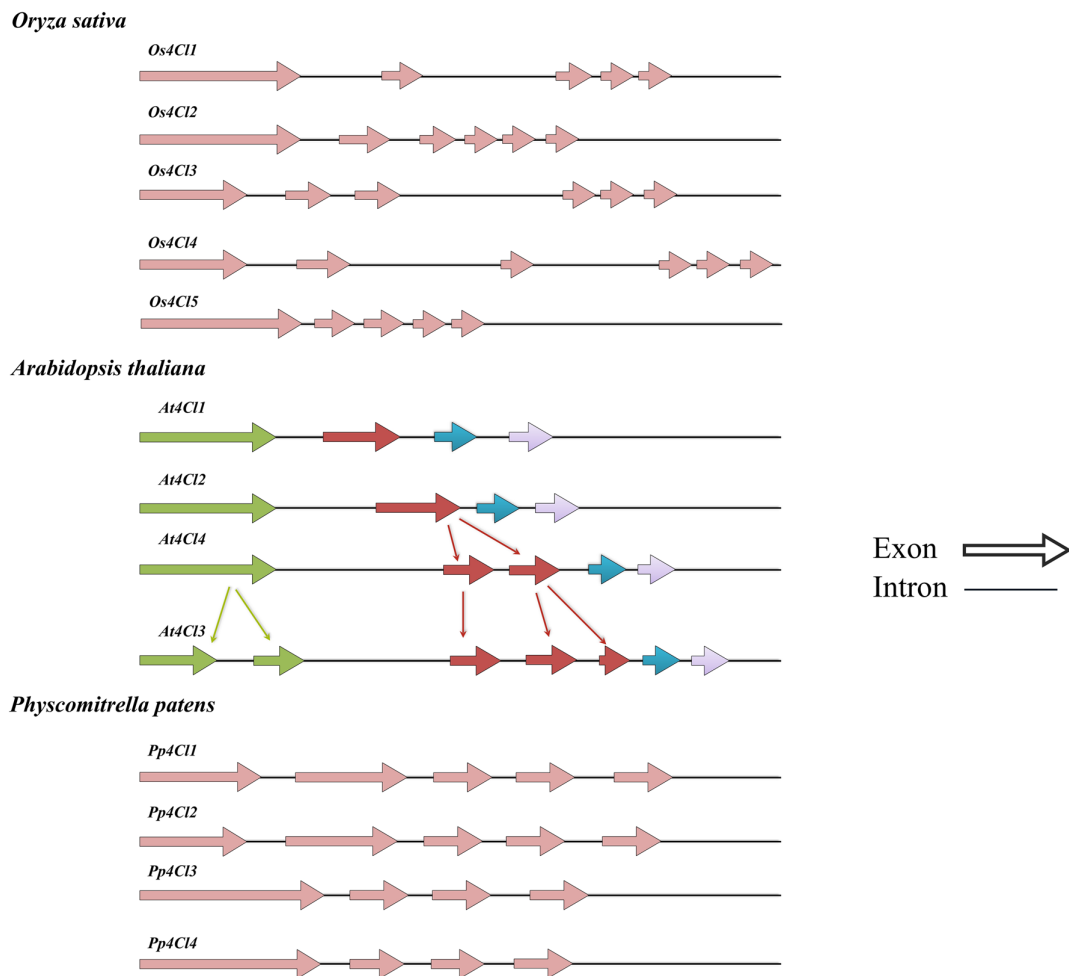


Fig. 4 Gene structure of 4CLs from *Oryza sativa* ssp. japonica, *Arabidopsis thaliana*, and *Physcomitrella patens*. Gene structure of different 4CL isoforms shows that there is variation in exon and

intron number, this is due to interruption of exon and gene duplication event. Arrow represents exon and line represents intron of 4CL gene. 4CL 4-coumate: coenzyme A ligase

that *Pt4CL1* and *Pt4CL2* are involved in biosynthesis of lignin and other phenylpropanoids, respectively (Hu et al. 1998; Sutela et al. 2014). Three 4CL protein isoforms were detected in hybrid poplar (*Populus trichocarpa* and *Populus deltoides*) using fast-protein liquid chromatography. All of them utilize hydroxycinnamic acid but are inactive against sinapic acid (Allina et al. 1998). In this hybrid, 4CL1 is preferably expressed in old leaves, green stem and xylem while 4CL2 is expressed in young leaves (Allina et al. 1998). Five Pto4CL isoforms were characterized in *Populus tomentosa* by expression analysis and activity of recombinant enzymes with different substrates. All five isoforms have different substrate specificities and turnover rates. None of them is able to utilize sinapate as substrate. Overexpression of all these five isoforms leads to significant increase in the level of lignin. In case of Pto4CL4, overexpressed transgenic

tobacco increase in naringenin content was observed (Rao et al. 2015). Crude proteins extracted from developing xylem of *Robinia pseudoacacia* have three 4CL isoforms. The isoform Rp4CL1 preferably utilizes *p*-coumarate as substrate, but is unable to utilize ferulate and sinapate. Rp4CL2 and Rp4CL3 utilize sinapate and also show high activity with caffeate and *p*-coumarate. The crude extract from the shoots also have very similar substrate preference pattern. These results suggest that sinapate-activating Rp4CL isoforms are constitutively expressed in lignin-forming cells (Hamada et al. 2004).

Flavonoid biosynthesis

Plant produces flavonoids, hydroxycinnamic acids and their related compounds. These compounds have a wide range of

roles in plant including attractant, deterrent, symbiotic and allelopathic interaction, and are crucial for protection from UV radiation (Mierziak et al. 2014). These compounds are primarily present in epidermis of leaves, stems, apical meristem and pollen. Flavonoids with conjugated double bonds are potent antioxidants than single conjugated bond. They furthermore undergo functional group modification (methylation and glycosylation) to alter the reactivity, solubility and stability. In addition, they are known to protect from reactive oxygen species (ROS) by suppression of singlet oxygen, inhibition of enzymes that generate ROS (cyclooxygenase, lipoxygenase, monooxygenase, xanthine oxidase), chelating ions, free-radical quenching and recycling of other antioxidants. All UV radiations induce synthesis of protecting flavonoids, but comparatively UVC and UVB induce more than UVA (Saewan and Jimtaisong 2013; Mierziak et al. 2014; Panche et al. 2016; Surjadinata et al. 2017; Zhao et al. 2017). The product of 4CL enzyme is required for flavonoid biosynthesis. p-coumaroyl-CoA/cinnamoyl CoA molecule is condensed with three molecules of malonyl-CoA to yield chalcone, and this reaction is catalyzed by CHS. Then, chalcone is isomerized to flavanone by CHI. Flavanone is further utilized by several branch pathways to synthesize various flavonoids, including aurones, dihydrochalcones, flavanonols (dihydroflavonols), isoflavones, flavones, flavonols, leucoanthocyanidins, anthocyanins and proanthocyanidins (Mierziak et al. 2014; Hahlbrock and Scheel 1989; Dixon and Paiva 1995; Holton and Cornish 1995). A specific 4CL isoform is found to be responsible for flux diversion toward flavonoid biosynthesis. In *Arabidopsis thaliana*, *At4CL3* is specifically expressed in light-exposed tissues such as leaf and flowers and positively correlated with flavonoid content of these tissues (Lee and Douglas 1996). A similar expression pattern is observed in the case of rice *Os4CL2* gene in rice. *Os4CL2* was specifically expressed in the anther and expression is escalated by UV irradiation, suggesting its potential involvement in flux diversion for flavonoid biosynthesis (Sun et al. 2013). *Pueraria lobata* is used as an herbal drug for prevention of migraine, hypertension, alcoholism and cardiovascular disorder. Its main active ingredient includes isoflavonoids such as puerarin, daidzin, genistin and other compounds. In this plant, *Pl4CL* expression is found to be highest in root tissue. The level of puerarin is also highest in root. Upon treatment with methyl jasmonate (MeJA), expression level of *Pl4CL1* is upregulated and the content of puerarin is also increased over threefold. This suggests that *Pl4CL1* is responsible for biosynthesis of these isoflavonoids in *Pueraria lobata* (Li et al. 2014).

Protection from biotic and abiotic stresses

Various reports show that the level of 4CL increases upon biotic and abiotic stresses. Protection from this is

mainly achieved by modulating level of lignin, flavonoid and other secondary metabolites. In *Arabidopsis thaliana*, *At4CL* expression is studied in response to various stresses. In wound treatment, expression level of *At4CL1* and *At4CL2* was increased (Lee and Douglas 1996; Ehltling et al. 1999), while the level of *At4CL3* decreased (Soltani et al. 2006). *At4CL3* is involved in biosynthesis of flavonoid, while *At4CL1* and *At4CL2* are involved in lignin biosynthesis (Ehltling et al. 1999). In *Ocimum basilicum*, one of the *Ob4CL* isoform decreased in response to drought stress, this might be because the role of 4CL in the biosynthesis of metabolites other than lignin (Mandoulakani et al. 2017). *Physaria pruinosa* is a salt-tolerant species of poplar, while *Populus trichocarpa* is salt-sensitive. When callus was grown in salt stress condition, it was observed that the expression of *Pp4CL2*, *Pp4CL11*, and *Pp4CL12* is induced significantly in the resistant species compared to the sensitive one (Zhang et al. 2015). When tomato (*Solanum lycopersicum* L.) was infected with *Alternaria solani*, the transcript level of 4CL gene is found to be upregulated (Shinde et al. 2017). These data suggest that in plant, 4CL plays a role during biotic and abiotic stress.

Activity and expression of 4CL is tightly regulated at various levels

Regulation at transcriptional level

The transcriptional control of phenylpropanoid enzymes is a key factor for regulation (Dixon and Paiva 1995). The promoter region of *Ec4CL1* from *Eucalyptus camaldulensis* is studied to investigate its role at transcriptional level regulation. The 1127 bp 5' upstream sequence contains various regulatory elements including cis-regulatory and cis-acting. These elements include light-responsive, low temperature-responsive, abscisic acid-responsive, fungal elicitor responsive, meristem-specific activation and element for restricting vascular expression to the xylem tissue (Hue et al. 2016). Methylation pattern in upstream promoter region of *St4CL1* and *St4CL1a* genes is studied in *Solanum tuberosum*. No difference is observed in the methylation pattern between elicitor stimulated (*Phytophthora infestans* culture filtrate) and non-elicitor stimulated (Becker-Andre et al. 1991). In case of *Pc4CL* promoter from *Petroselinum crispum*, change in methylation pattern is observed when treated with UV light but a similar change is not observed in case of *St4CL* promoter (Douglas et al. 1987; Becker-Andre et al. 1991). This difference in gene expression and methylation pattern is probably due to lack of two motifs in *St4CL*, which are present in *Pc4CL* promoter region centered around – 57 and – 127 relative to transcription start site (Becker-Andre et al.

Table 2 List of reported factors/elements that regulate *4CL* in plant

Gene	Factor/element	Effect	References
<i>Ip4CL</i>	IbMYB1a	Overexpression leads to anthocyanin accumulation	An et al. (2015)
<i>Os4CL2</i>	Wounding	Downregulation	Sun et al. (2013)
<i>Pa4CL</i>	UV radiation	Upregulation	Gao et al. (2015)
	SA and MeJA	Upregulation	
<i>Hc4CL</i>	ABA	Downregulation	Choudhary et al. (2013)
	SA and MeJA	Downregulation	
<i>At4CL1, At4CL2</i>	ABA	Upregulation	Lee and Douglas (1996)
	Wound and MeJA	Upregulation	
<i>At4CL3</i>	<i>Peronospora parasitica</i> infection and wounding	Upregulation	Ehlting et al. (1999)
	<i>Peronospora parasitica</i> infection and wounding	No change	
<i>Ob4CL</i>	Drought stress	Downregulation	Mandoulakani et al. (2017)
<i>Pp4CL</i>	ABA, MeJA and GA	Upregulation	Peng et al. (2016)
<i>Pe4CL2, Pe4CL11, Pe4CL12</i>	NaCl stress	Change depends upon genotypes	Zhang et al. (2015)
<i>Pe4CL5</i>	NaCl stress	Downregulation	Zhang et al. (2015)
<i>Pe4CL9, Pe4CL10</i>	NaCl stress	No change	Zhang et al. (2015)
<i>Ma4CL3</i>	Wounding, salicylic acid, and ultraviolet treatments	Upregulation	Wang et al. (2016)
<i>Mt4CL</i>	Aluminum stress	Upregulation	Chandran et al. (2008)
<i>Cs4CL</i>	Catechin treatment	Downregulation	Rani et al. (2009)
	Drought stress, ABA and GA3	Downregulation, decrease in catechin content	
	Wounding	Upregulation, increase in catechin content	
<i>Sa4CL3</i>	Light exposure	Upregulation	Gaid et al. (2011)
<i>St4CLs</i> ^a	<i>Phytophthora infestans</i> infection and arachidonic acid	Upregulation	Fritzemeier et al. (1987)
<i>Ps4CLs</i> ^a	<i>Phytophthora megasperma</i> glycinia infection	High expression in epidermal cells, oil-duct epithelial cells and developing xylem	Schmelzer et al. (1989)
<i>Pv4CL</i>	<i>Sclerotinia sclerotiorum</i>	Upregulation	Oliveira et al. (2015)

Ib, *Ipomoea batatas* L.; *Os*, *Oriza sativa* L.; *Pa*, *Plagioclasma appendiculatum* Lehm. & Lindenb.; *Hc*, *Hibiscus cannabinus* L.; *At*, *Arabidopsis thaliana* L.; *Ob*, *Ocimum basilicum* L.; *Pp*, *Pennisetum purpureum* Schumach.; *Pe*, *Populus euphratica* Oliv.; *Ma*, *Morus notabilis* L.; *Mt*, *Medicago truncatula* L.; *Cs*, *Camellia sinensis* L.; *Sa*, *Sorbus aucuparia* L.; *St*, *Solanum tuberosum* L.; *Ps*, *Petroselinum sativum* L.; *Pv*, *Phaseolus vulgaris* L.

^aAnalysis is done using northern blot technique

1991; Table 2). The transcription factor MYB is known to regulate phenylpropanoid pathway enzymes including *4CL* expression. Overexpression of *AtMYB4* in tobacco reduces the basal transcript levels of *C4H*, *4CL1* and *CAD* genes. In addition, an *AtMYB4* mutant *Arabidopsis thaliana* shows enhanced tolerance to UV-B relative than wild type. Expression of *AtMYB4* is reduced upon exposure to UV-B light and wounding, which leads to de-repression of *C4H* that results in higher synthesis of protecting sinapate esters. Upon UV-B exposure, overexpressing *AtMYB4* *Arabidopsis thaliana* plants are more sensitive that leads to death (37% of plants) while no death in wild type plants

occurs. In addition, these lines have reduced sinapate ester level and unchanged flavonoid composition, this suggests that *AtMYB4* negatively regulates biosynthesis of UV protectant metabolites (Jin et al. 2000). In *Ipomoea batatas* (sweet potato), *IbMYB1a* transcription factor regulates the expression of anthocyanin biosynthesis genes along with *4CL*. *IbMYB1a* positively regulates multiple anthocyanin biosynthetic genes. This is confirmed by overexpression of *IbMYB1a* under different promoters in tobacco. These transgenic lines showed increased expression of genes encoding PAL, C4H, 4CL, CHS, CHI, F3H, DFR, and ANS (An et al. 2015; Table 2).

Developmentally regulated and wound-induced gene expression of *4CL* is studied using fusion of *4CL* promoter (truncated and full) with *GUS* (beta-glucuronidase). In *Arabidopsis thaliana* transformed with *GUS* coding region under control of *4CL* promoter (full/truncated), the expression of *GUS::At4CL1* and *GUS::At4CL2* is restricted to vascular tissues of root and aerial organs. The *At4CL3::GUS* expression is higher in non-vascular tissue (leaf and cotyledons), upper part of hypocotyls, and roots, whereas the *At4CL4::GUS* expression is restricted to roots only. Regulatory element analysis shows that – 950 to – 750 bp region of the *At4CL2* promoter regulates early wound response, while region from – 950 to – 1600 bp negatively affects early wound response. Late wounding response of *At4CL2* may be because of presence of late wound response element in intron 1/2/3. The constructs having all intron (intron 1, 2 and 3) shows a strong response after 72 h of wound compared to a construct having less introns and without introns (Soltani et al. 2006).

Regulation by plant hormones

Hormones modulate metabolic pathways including phenylpropanoid pathway by regulating *4CL* expression (Table 2). For example, 2-month-old *Plagiochasma appendiculatum* thallus was treated with abscisic acid (ABA), salicylic acid (SA) and MeJA, showed upregulation of *Pa4CL1* in case of SA and MeJA treatment while it was downregulated in response to ABA treatment (Gao et al. 2015). *Hc4CL* from *Hibiscus cannabinus* showed downregulation when treated with MeJA and SA. However, in ABA treatment, *Hc4CL* transcript level slightly decreased in 1 h and then subsequently increased to maximum in 24 h (Choudhary et al. 2013). Expression of *Nt4CL1* and *Nt4CL2* (*Nicotiana tabacum*) was induced upon wounding and MeJA treatment (Lee and Douglas 1996). In *Salvia miltiorrhiza*, MeJA responsive element present in *SmC4H1*, *Sm4CL2*, *Sm4CL3*, *Sm4CL-like1*, *Sm4CL-like2*, *Sm4CL-like3*, *Sm4CL-like6*, *Sm4CL-like7* and their expression may be regulated by MeJA (Wang et al. 2015). Information about the role of hormonal regulation in *4CL* genes from *Arabidopsis thaliana*, *Pennisetum purpureum*, *Morus notabilis*, *Camellia sinensis* are reported in Table 2. Specific *4CL* isoform gets upregulated or downregulated in response to particular type of hormone in different plants. For example, in response to SA treatment, *Pa4CL* in *Plagiochasma appendiculatum* is upregulated while, *Hc4CL* is downregulated in *Hibiscus cannabinus*. These findings suggest that each *4CL* isoform has differential response to stimuli.

Regulation in response to biotic and abiotic stresses

Several reports suggested that the *4CL* gene expression is regulated by various abiotic stresses such as drought, salinity, and temperature (Table 2). In *Arabidopsis thaliana*, upon wounding, expression of *At4CL1* and *At4CL2* increased after 2.5 h and then decreased to the basal level of expression, but again increased to maximum after 48 h (Ehltling et al. 1999). Similar patterns of gene expression like *At4CL1* and *At4CL2* were shown by *Hc4CL* in *Hibiscus cannabinus* plant in response to wounding (Choudhary et al. 2013). The *At4CL3* expression level showed downregulation and that become normal by 4 h and then simultaneously increased up to 72 h post wounding (Soltani et al. 2006). *At4CL4* expression also increased up to 2.5 h and remains higher up to 12 h and then falls down to basal level. Expression pattern of *At4CL3* shown by Soltani et al. (2006) is different from pattern shown by Ehltling et al. (1999). In *Oryza sativa japonica*, upon wounding *Os4CL3*, *Os4CL4*, *Os4CL5* and other genes of phenylpropanoid biosynthesis pathway were significantly upregulated, whereas *Os4CL1* and *Os4CL2* were downregulated. When dark-adapted rice plants were illuminated with UV light, *Os4CL1*, *Os4CL3*, *Os4CL4*, and *Os4CL5* were downregulated and *Os4CL2* upregulated. It is concluded that the *Os4CL2* most probably is involved in biosynthesis of flavonol (Sun et al. 2013). When *Ocimum basilicum* plants were subjected to controlled drought stress condition, the expression levels of *Ob4CL* and *ObC4H* genes decreased, while expression levels of *CVOMT* and *EOMT* increased and *CAD* expression was relatively unchanged. This has been correlated with essential oil compounds and observed to have the highest amount of methyl chavicol, methyl eugenol, α -bergamotene, and β -myrcene. This suggests that altered gene expression in response to drought stress leads to increase in the methylchavicol and methyl-eugenol content (Mandoulakani et al. 2017). In response to NaCl treatment, *PAL*, *CCoAOMT*, *C3H*, *HCT* and *F5H* genes were upregulated, which are involved in lignin biosynthesis (Choudhary et al. 2013; Kim et al. 2013). It is clearly indicated that *4CL* has a significant role in countering various abiotic stresses. This has been also proven in other plants like *Populus euphratica*, *Morus notabilis*, *Medicago truncatula*, *Camellia sinensis*, and *Sorbus aucuparia* (Table 2).

Biotic stresses like bacterial/fungal/virus infection and insect infestation in plant can be combated through phenylpropanoid modulation. The *4CL* of phenylpropanoid pathway is one of the key enzymes and its expression is altered in response to biotic stresses (Table 2), clearly indicating that *4CL* has a significant role in countering various biotic stresses. Potato leaves infected with *Phytophthora infestans* leads to rapid accumulation of *PAL* and *4CL* and other genes from the phenylpropanoid pathway. *Phytophthora infestans* is one of the most destructive fungal pathogens

and leads to rapid browning and hypersensitive cell death around the infection site. In this study, mRNA of both *PAL* and *4CL* gene has maximum translational activity after 2 h and enzyme activity has increased up to twofold in 6–12 h after infection (Fritzemeier et al. 1987). A similar type of upregulation of *4CL* genes was observed in the case of the following plant–pathogen interaction *Arabidopsis thaliana*, *Peronospora parasitica*, *Solanum tuberosum*, *Phytophthora infestans*, *Petroselinum sativum*, *Phytophthora megasperma*, *Phaseolus vulgaris* and *Sclerotinia sclerotiorum* (Table 2). Grapevine red blotch-associated virus is a major problem faced by cultivated grapevines (*Vitis vinifera*). Being infected by this virus leads to the downregulation of phenylpropanoid pathway genes along with *4CL* at the ripening stage (Blanco-Ulate et al. 2017).

Functional specificity of 4CL isoforms: a case study

Co-expression analysis of *At4CL1*, *At4CL3* from *Arabidopsis thaliana* and *Gm4CL4* from *Glycine max* was performed individually using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins, all the parameters were kept as default except the number of interactions that need to be shown was set to 20) database and merged (Szkarczyk et al. 2015; Fig. 5). *At4CL1* was co-expressed with lignin biosynthesis genes, while *At4CL3* with flavonoid biosynthesis genes. This shows that each 4CL isoform may be responsible for metabolic flux diversion. Both *At4CL1* and *At4CL3* also show co-expression with genes of other metabolite biosynthetic pathways (other

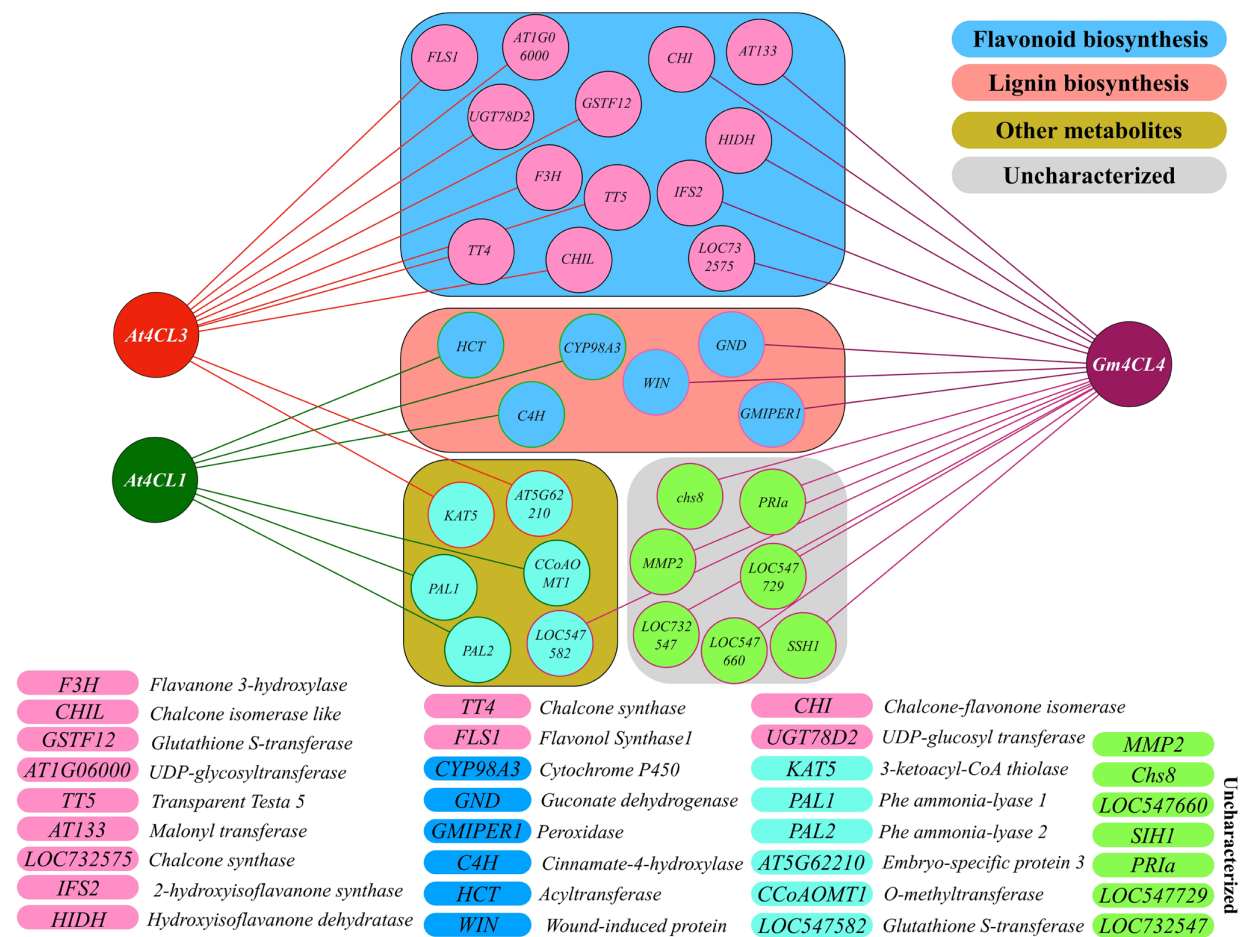


Fig. 5 Co-expression analysis of *At4CL1*, *At4CL3* and *Gm4CL4* was performed individually using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database and merged. *At4CL1* and *At4CL3* shows co-expression with genes related to lignin and flavonoid biosynthesis while *Gm4CL4* with both pathway-related genes.

Default setting parameters were used except the number of interactions that need to be shown was set to 20. Outcome of each 4CL in STRING analysis was used to deduce this figure from supplementary material Fig. S1. 4CL, 4-coumate: coenzyme A ligase; *At*, *Arabidopsis thaliana*; *Gm*, *Glycine max*

than lignin and flavonoid biosynthesis genes). *Gm4CLA* shows co-expression with genes for biosynthesis of lignin, flavonoid, other metabolites and uncharacterized genes (Fig. 5). Based on 4CL's co-expression analysis from *Arabidopsis thaliana* (*At4CL1* and *At4CL2*) and *Glycine max* (*Gm4CLA*), it showed that some plants may have single isoform for both lignin and flavonoid and some plants may have separate isoform for each of the class.

Scope of 4CLs in metabolic pathway engineering for various products

Enzymes of phenylpropanoid and flavonoid metabolism present in the form of membrane-associated multienzyme complexes as a metabolon, on the cytosolic face of the endoplasmic reticulum (Hrazdina and Wagner 1985). These organized multienzyme complexes are efficient in channeling the intermediates between enzymes of sequential reactions (Winkel-Shirley 1999; Burbulis and Winkel-Shirley 1999). Interactions and organization of metabolon may vary in different plant species (Fujino et al. 2018). As reviewed by Chemler and Koffas (2008), it is possible to synthesize flavonoids using gene cloning and expression of 4CL and other enzymes involved in its biosynthesis in bacteria (*Escherichia coli*) and yeast (*Saccharomyces cerevisiae*).

Fuel industry

Lignin along with cellulose microfibrils provides rigidity to plant. Lignin content affects utilization of plant material for chemicals, fiber and energy production, etc. Lignin is one of the products of phenylpropanoid pathway and the expression of specific 4CL isoforms plays a role in lignin biosynthesis (Costa et al. 2005; Shigeto et al. 2017). Transgenic aspen (*Populus tremuloides* Michx.) in which *Pt4CL1* gene was silenced showed downregulation of lignin biosynthesis. Such plants showed reduced lignin up to 45% while cellulose content increased in 15%. Silenced lines have thicker stem, longer internodes, larger leaves and overall enhanced growth rate than control (Hu et al. 1999). Sugarcane is the best raw material for bioethanol production. Two 4CL isoforms are characterized from *Saccharum* spp. hybrids. Out of two, *Sh4CL1* is involved in lignin biosynthesis based on phylogenetic analysis and RNAi silencing. RNAi suppression of *Sh4CL1* leads to decrease in 16.5% lignin content and shows improvement of 52–76% saccharification rate as compared to wild type control. This suggests that 4CL silencing can be used to improve lignocellulosic raw material for biofuel production (Jung et al. 2016).

Natural product industry

Various valuable natural products (NPs) are available from general phenylpropanoid pathway. Use of 4CL for production of high value NPs has already been proven in bacterial, fungal and plant system. Resveratrol is an important secondary metabolite, which has health-promoting properties and is found in red wine (Giovinazzo et al. 2012). Fusion of *At4CL1* from *Arabidopsis thaliana* and *stilbene synthase* (*VvSTS*) from *Vitis vinifera* can be used to improve resveratrol synthesis in yeast. Using this fusion protein (*At4CL1-VsSTS*), resveratrol level was increased upto 15-fold as compared to yeast individually expressing these enzymes (Wang et al. 2011). Rosmarinic acid is also one of the important products of phenylpropanoid pathway produced by fusion of 4-hydroxyphenyllactic acid with 4-coumaroyl-CoA, and this reaction is catalyzed by hydroxycinnamoyl-CoA: hydroxycinnamoyl transferase. It has various biological activities including antiviral, antibacterial, antioxidant and anti-inflammatory (Petersen and Simmonds 2003). Rosmarinic acid accumulation increased from 2.1- to 3.9-fold in response to MeJA treatment in cell cultures of *Agastache rugosa* Kuntze. Transcript level of *ArPAL*, *Ar4CL*, and *ArC4H* were increased up to 4.5-, 3.4- and 3.5-fold, respectively, compared to untreated culture cells (Kim et al. 2013). Plants belonging to order Zingiberales produce curcuminoids that has a wide range of applications including food additive, stimulant, food coloring agent, anti-tumor, antioxidant and hepatoprotective activities, etc. (Amalraj et al. 2017). Biosynthesis of curcuminoids in *Escherichia coli* was achieved by cloning of enzymes: 4CL from *Lithospermum erythrorhizon* and curcuminoid synthase (*CUS*) from rice. Ferulic acid from rice bran pitch, an industrial waste generated during rice edible oil production was used during growth of recombinant *Escherichia coli* expressing 4CL and *CUS* for production of curcuminoids (Katsuyama et al. 2008). *Os4CL* in *Ocimum sanctum* is responsible for partition of metabolite flux towards eugenol biosynthesis (Rastogi et al. 2013). Eugenol has a role in inhibiting advanced glycation end products in diabetes (Singh et al. 2016). 4CL has a prominent role in natural product synthesis improvement.

Food technology

The expression pattern of genes involved in monolignol and flavonoid biosynthesis are studied during the fruit ripening in apple. Differential expression of PAL and 4CL isoforms leads to suppression in lignin biosynthesis and strong induction of flavonoid biosynthesis during fruit ripening (Baldi et al. 2017). Results show that fine-tuning of pathway was achieved by differential expression of branch point enzyme isoforms such as PAL and 4CL. The differential expression

of a specific 4CL isoform was observed at different stages of pear fruit development (Cao et al. 2016). Thus, use of 4CL manipulation for metabolic pathway engineering is useful in post harvesting and food technology-related industry.

Conclusion and future perspectives

1. The 4CL is essentially involved in channelization of precursors for different phenylpropanoid biosynthesis. Based on phylogenetic analysis, 4CL is divided in four clusters, i.e., type I which is involved in monolignol biosynthesis, type II that is involved in phenylpropanoid biosynthesis other than lignin, type III and IV that are present in monocots with similar functions as type I and II, respectively. The 4CL genes in monocots and dicots evolved independently after their separation. However, type I and II might have diverged before the divergence of gymnosperm and angiosperm lineages. The 4CL activity and expression are tightly regulated in a spatio-temporal manner, and also in response to biotic and abiotic stresses.
2. The role of amino acids in SBP and domains of 4CL is well-studied with help of techniques such as bioinformatics tools, crystallography, and site-directed mutagenesis. Its activity has been successfully improved by site-directed mutagenesis and domain exchange approaches.
3. Co-expression analysis of characterized 4CL from various plants suggests that a single 4CL is required for the lignin and flavonoid biosynthesis, whereas some plants have separate 4CLs for these biosynthetic pathways.
4. The metabolic pathways are regulated at entry point, branch point and/or in some cases by intermediate enzymes for its final product biosynthesis. However, terminal enzymes are well-studied in most of the metabolic pathways. The 4CL enzymes are mostly studied in the context of functional characterization and regulation by external factors. Effect of internal factor/triggers on 4CL function and regulation remains enigmatic, which is an important aspect to harness further applications in metabolic engineering. It has been established that specific PAL isoforms also play an important role in regulation of phenylpropanoid pathway, which is the first committed step of this pathway.
5. The detailed validation is a prudent necessity for efficient metabolic pathway engineering: (a) the role of intermediate and end product metabolites during specific 4CL isoform activity; (b) involvement of transcription factors in regulating the levels of individual 4CL isoforms; (c) cross talk with other pathways; (d) splicing mechanism in formation of multiple isoforms; (e) regulation at post-translational level; (f) activity optimization in bacterial or yeast host system, etc. Such information

is highly imperative to achieve efficient metabolic pathway engineering for commercially important candidate metabolites.

Author contribution statement APG evolved theme of the review. SGL and RMK performed literature survey and prepared draft of the review. APG edited and finalized the draft. All authors have contributed to the revision and finalizing the manuscript.

Acknowledgements We thank Dr. Jayant Khandare, Actorius Innovations and Research Pvt Ltd., Pune, India for critically reading the manuscript and for the English language suggestions. SGL acknowledge the support from Council of Scientific and Industrial Research (CSIR), New Delhi, India. The work was funded by CSIR-NCL-IGIB Joint Research program under XII Five Year Plan (BSC0124).

References

- Allina SM, Pri-Hadash A, Theilmann DA et al (1998) 4-Coumarate-coenzyme A ligase in hybrid poplar properties of native enzyme, cDNA cloning, and analysis of recombinant enzyme. *Plant Physiol* 116:743–754. <https://doi.org/10.1104/pp.116.2.743>
- Amalraj A, Pius A, Gopi S et al (2017) Biological activities of curcuminoids, other biomolecules from turmeric and their derivatives—a review. *J Tradit Complement Med* 7(2):205–233. <https://doi.org/10.1016/j.jtcme.2016.05.005>
- An CH, Lee KW, Lee SH et al (2015) Heterologous expression of IbMYB1a by different promoters exhibits different patterns of anthocyanin accumulation in tobacco. *Plant Physiol Biochem* 89:1–10. <https://doi.org/10.1016/j.plaphy.2015.02.002>
- Anand A, Jayaramaiah RH, Beedkar SD et al (2016) Comparative functional characterization of eugenol synthase from four different *Ocimum* species: implications on eugenol accumulation. *Biochim Biophys Acta Proteins Proteom* 1864(11):1539–1547. <https://doi.org/10.1016/j.bbapap.2016.08.004>
- Baldi P, Moser M, Brilli M et al (2017) Fine-tuning of the flavonoid and monolignol pathways during apple early fruit development. *Planta* 245:1021–1035. <https://doi.org/10.1007/s00425-017-2660-5>
- Becker-Andre M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-Coumarate: CoA ligase genes in potato. *J Biol Chem* 266:8551–8559
- Bennett RN, Wallsgrave RM (1994) Secondary metabolites in plant defence mechanisms. *New Phytol* 127:617–633. <https://doi.org/10.1111/j.1469-8137.1994.tb02968.x>
- Blanco-Ulate B, Hoper H, Figueroa-Balderas R et al (2017) Red blotch disease alters grape berry development and metabolism by interfering with the transcriptional and hormonal regulation of ripening. *J Exp Bot* 68(5):1225–1238. <https://doi.org/10.1093/jxb/erw506>
- Bolton MD (2009) Primary metabolism and plant defense-fuel for the fire. *Mol Plant Microbe Interact* 22(5):487–497. <https://doi.org/10.1094/MPMI-22-5-0487>
- Burbulis IE, Winkel-Shirley B (1999) Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proc Natl Acad Sci USA* 96(22):12929–12934. <https://doi.org/10.1073/pnas.96.22.12929>

- Cao Y, Han Y, Li D et al (2016) Systematic analysis of the 4-Coumarate: coenzyme A ligase (4CL) related genes and expression profiling during Fruit development in the Chinese Pear. *Genes* 7(10):89. <https://doi.org/10.3390/genes7100089>
- Chandran D, Sharapova N, Ivashuta S et al (2008) Transcriptome profiling identified novel genes associated with aluminum toxicity, resistance and tolerance in *Medicago truncatula*. *Planta* 228:151–166. <https://doi.org/10.1007/s00425-008-0726-0>
- Chemler JA, Koffas MA (2008) Metabolic engineering for plant natural product biosynthesis in microbes. *Curr Opin Biotechnol* 19:597–605. <https://doi.org/10.1016/j.copbio.2008.10.011>
- Chen HC, Song J, Wang JP et al (2014a) Systems biology of lignin biosynthesis in *Populus trichocarpa*: heteromeric 4-coumaric acid: Coenzyme A ligase protein complex formation, regulation, and numerical modeling. *Plant Cell* 26:876–893. <https://doi.org/10.1105/tpc.113.119685>
- Chen HY, Babst BA, Nyamdari B et al (2014b) Ectopic expression of a Loblolly Pine class II 4-Coumarate:CoA ligase alters soluble phenylpropanoid metabolism but not lignin biosynthesis in *Populus*. *Plant Cell Physiol* 55(9):1669–1678. <https://doi.org/10.1093/pcp/pcu098>
- Choudhary EK, Choi B, Cho BK et al (2013) Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. *Plant Omics J*. 6(4):254–262
- Citti C, Battisti UM, Braghiroli D et al (2017) A metabolomic approach applied to a liquid chromatography coupled to high-resolution tandem mass spectrometry method (HPLC–ESI–HRMS/MS): towards the comprehensive evaluation of the chemical composition of cannabis medicinal extracts. *Phytochem Anal* 29:144–155. <https://doi.org/10.1002/pca.2722>
- Costa MA, Bedger DL, Moinuddin SGA et al (2005) Characterization in vitro and in vivo of the putative multigene 4-coumarate: CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry* 66:2072–2091. <https://doi.org/10.1016/j.phytochem.2005.06.022>
- Cukovic D, Ehlting J, Vanziffle JA (2001) Structure and evolution of 4-Coumarate: coenzyme A ligase (4CL) gene families. *Biol Chem* 382:645–654. <https://doi.org/10.1515/BC.2001.076>
- Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism *Plant Cell* 7:1085–1097. <https://doi.org/10.1105/tpc.7.7.1085>
- Douglas CJ (1996) Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. *Trends Plant Sci* 1(6):171–178. [https://doi.org/10.1016/1360-1385\(96\)10019-4](https://doi.org/10.1016/1360-1385(96)10019-4)
- Douglas C, Hoffmann H, Wolfgang S et al (1987) Structure and elicitor or uv-light-stimulated expression of two 4-coumarate-CoA ligase genes in parsley. *EMBO J* 6(5):1189–1195
- Ehlting J, Buttner D, Wang Q et al (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J* 19:9–20. <https://doi.org/10.1046/j.1365-313X.1999.00491.x>
- Ehlting J, Shin JJK, Douglas CJ (2001) Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *Plant J* 25:455–465. <https://doi.org/10.1046/j.1365-313X.2001.01122.x>
- Ferrieri AP, Arce CCM, Machado RAR et al (2015) A *Nicotiana attenuata* cell wall invertase inhibitor (NaCWII) reduces growth and increases secondary metabolite biosynthesis in herbivore-attacked plants. *New Phytol* 208:519–530. <https://doi.org/10.1111/nph.13475>
- Floros DJ, Petras D, Kapon CA et al (2017) Mass spectrometry based molecular 3D-cartography of plant metabolites. *Front Plant Sci* 8:429. <https://doi.org/10.3389/fpls.2017.00429>
- Fraser CM, Chapple C (2011) The phenylpropanoid pathway in *Arabidopsis*. *The Arabidopsis Book/American Society of Plant Biologists* 9:e0152. <https://doi.org/10.1199/tab.0152>
- Freund DM, Hegeman AD (2017) Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. *Curr Opin Biotechnol* 43:41–48. <https://doi.org/10.1016/j.copbio.2016.08.002>
- Fritzemeier KH, Cretin C, Kombrink E et al (1987) Transient induction of phenylalanine ammonia-lyase and 4-Coumarate:CoA ligase mRNAs in Potato leaves infected with virulent or avirulent races of *Phytophthora infestans*. *Plant Physiol* 85:34–41. <https://doi.org/10.1104/pp.85.1.34>
- Fujino N, Tenma N, Waki T et al (2018) Physical interactions among flavonoid enzymes in snapdragon and torenia reveal the diversity in the flavonoid metabolon organization of different plant species. *Plant J* 94(2):372–392. <https://doi.org/10.1111/tpj.13864>
- Fukushima A, Kusano M, Redestig H et al (2009) Integrated omics approaches in plant systems biology. *Curr Opin Chem Biol* 13:532–538. <https://doi.org/10.1016/j.cbpa.2009.09.022>
- Fulda M, Heinz E, Wolter FP (1994) The fadD gene of *Escherichia coli* K12 is located close to rnd at 39.6 min of the chromosomal map and is a new member of the AMP-binding protein family. *Mol Gen Genet* 242:241–249. <https://doi.org/10.1007/BF00280412>
- Gaid MM, Scharnhop H, Ramadan H et al (2011) 4-Coumarate:CoA ligase family members from elicitor-treated *Sorbus aucuparia* cell cultures. *J Plant Physiol* 168:944–951. <https://doi.org/10.1016/j.jplph.2010.11.021>
- Gao S, Yu HN, Xu RX et al (2015) Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagi-ochasma appendiculatum*. *Phytochemistry* 111:48–58. <https://doi.org/10.1016/j.phytochem.2014.12.017>
- Giovinazzo G, Ingrosso I, Paradiso A et al (2012) Resveratrol biosynthesis: plant metabolic engineering for nutritional improvement of food. *Plant Foods Hum Nutr* 67:191–199. <https://doi.org/10.1007/s11130-012-0299-8>
- Goufo P, Moutinho-Pereira JM, Jorge TF et al (2017) Cowpea (*Vigna unguiculata* L. Walp.) metabolomics: osmoprotection as a physiological strategy for drought stress resistance and improved yield. *Front. Plant Sci* 8:586. <https://doi.org/10.3389/fpls.2017.00586>
- Gui J, Shen J, Li L (2011) Functional characterization of evolutionarily divergent 4-Coumarate: coenzyme A ligases in Rice. *Plant Physiol* 157:574–586. <https://doi.org/10.1104/pp.111.178301>
- Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. *Ann Rev Plant Physiol* 40:347–369
- Hamada K, Nishida T, Yamauchi K et al (2004) 4-Coumarate: coenzyme A ligase in black locust (*Robinia pseudoacacia*) catalyses the conversion of sinapate to sinapoyl-CoA. *J Plant Res* 117:303–310. <https://doi.org/10.1007/s10265-004-0159-1>
- Herrmann KM, Weaver LM (1999) The shikimate pathway. *Annu Rev Plant Physiol Plant Mol Biol* 50:473–503. <https://doi.org/10.1146/annurev.arplant.50.1.473>
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7:1071–1083. <https://doi.org/10.1105/tpc.7.7.1071>
- Hrazdina G, Wagner GJ (1985) Metabolic pathways as enzyme complexes: evidence 641 for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme 642 complexes. *Arch Biochem Biophys* 237:88–100
- Hu W, Kawaoka A, Tsai CJ et al (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate: coenzyme A ligase (4CL) genes in Aspen (*Populus tremuloides*). *Proc Natl Acad Sci USA* 95:5407–5412. <https://doi.org/10.1073/pnas.0307307101>
- Hu WJ, Harding SA, Lung J et al (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* 17:808–812. <https://doi.org/10.1038/11758>
- Hu Y, Gai Y, Yin L et al (2010) Crystal structures of a *Populus tomentosa* 4-coumarate: CoA ligase shed light on its enzymatic

- mechanisms. *Plant Cell* 22:3093–3104. <https://doi.org/10.1105/tpc.109.072652>
- Hue HTT, Ha DTT, Hai NV et al (2016) Isolation and characterization of the 4-coumarate:coenzyme A ligase (4CL1) promoter from *Eucalyptus camaldulensis*. *Physiol Mol Biol Plants* 22(3):399–405. <https://doi.org/10.1007/s12298-016-0369-8>
- Jin H, Cominell E, Bailey P et al (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in *Arabidopsis*. *EMBO J* 19(22):6150–6161. <https://doi.org/10.1093/emboj/19.22.6150>
- Jung JH, Kannan B, Dermawan H (2016) Precision breeding for RNAi suppression of a major 4-coumarate:coenzyme A ligase gene improve cell wall saccharification from field grown sugarcane. *Plant Mol Biol* 92:505–517. <https://doi.org/10.1007/s11103-016-0527-y>
- Katsuyama Y, Matsuzawa M, Funa N et al (2008) Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway. *Microbiol* 154:2620–2628. <https://doi.org/10.1099/mic.0.2008/018721-0>
- Kim YB, Kim JK, Uddin MR et al (2013) Metabolomics analysis and biosynthesis of Rosmarinic acid in *Agastache rugosa* Kuntze treated with Methyl Jasmonate. *PLoS One* 8(5):e64199. <https://doi.org/10.1371/journal.pone.0064199>
- Korkina LG (2007) Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. *Cell Mol Biol* 53:15–25. <https://doi.org/10.1170/772>
- Kueger S, Steinhauser D, Willmitzer L et al (2012) High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolite distributions. *Plant J* 70:39–50. <https://doi.org/10.1111/j.1365-313X.2012.04902.x>
- Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-Coumarate: coenzyme A ligase (4CL) gene family (cDNA structure, gene inheritance and expression, and properties of recombinant proteins). *Plant Physiol* 112:193–205. <https://doi.org/10.1104/pp.112.1.193>
- Li Z, Nair SK (2015) Structural basis for specificity and flexibility in a plant 4-coumarate:CoA ligase. *Structure* 23:2032–2042. <https://doi.org/10.1016/j.str.2015.08.012>
- Li ZB, Li CF, Li J et al (2014) Molecular cloning and functional characterization of two divergent 4-Coumarate: coenzyme A ligases from Kudzu (*Pueraria lobata*). *Biol Pharm Bull* 37(1):113–122. <https://doi.org/10.1248/bpb.b13-00633>
- Mandoulakani BA, Eyvazpour E, Ghadimzadeh M (2017) The effect of drought stress on the expression of key genes involved in the biosynthesis of phenylpropanoids and essential oil components in basil (*Ocimum basilicum* L.). *Phytochemistry* 139:1–7. <https://doi.org/10.1016/j.phytochem.2017.03.006>
- Mierziak J, Kostyn K, Kulma A (2014) Flavonoids as important molecules of plant interactions with the environment. *Molecules* 19(10):16240–16265
- Naik P, Wang JP, Sederoff R et al (2018) Assessing the impact of the 4CL enzyme complex on the robustness of monolignol biosynthesis using metabolic pathway analysis. *PLoS One* 13(3):e0193896. <https://doi.org/10.1371/journal.pone.0193896>
- Oliveira MB, Andrade RV, Grossi de Sá MF et al (2015) Analysis of genes that are differentially expressed during the *Sclerotinia sclerotiorum*–*Phaseolus vulgaris* interaction. *Front Microbiol* 6:1162. <https://doi.org/10.3389/fmicb.2015.01162>
- Panche AN, Diwan AD, Chandra SR (2016) Flavonoids: an overview. *J Nutr Sci* 5:e47. <https://doi.org/10.1017/jns.2016.41>
- Peng XQ, Ke AW, Liu JQ et al (2016) Deletion and hormone induction analyses of the 4-coumarate: CoA ligase gene promoter from *Pennisetum purpureum* in transgenic tobacco plants. *Plant Cell Tiss Organ Cult* 126:439–448. <https://doi.org/10.1007/s1124-0-016-1012-7>
- Petersen M, Simmonds MSJ (2003) Rosmarinic acid. *Phytochemistry* 62:121–125. [https://doi.org/10.1016/S0031-9422\(02\)00513-7](https://doi.org/10.1016/S0031-9422(02)00513-7)
- Rani A, Singh K, Sood P et al (2009) *p*-coumarate:CoA ligase as a key gene in the yield of catechins in tea [*Camellia sinensis* (L.) O. Kuntze]. *Funct Integr Genomics* 9:271–275. <https://doi.org/10.1007/s10142-008-0098-3>
- Rao G, Pan X, Xu F et al (2015) Divergent and overlapping function of five 4-Coumarate: coenzyme A ligases from *Populus tomentosa*. *Plant Mol Biol Rep* 33:841–854. <https://doi.org/10.1007/s1110-5-014-0803-4>
- Rastogi S, Kumar R, Chanotiya CS et al (2013) 4-Coumarate:CoA Ligase partitions metabolites for eugenol biosynthesis. *Plant Cell Physiol* 54(8):1238–1252. <https://doi.org/10.1093/pcp/pct073>
- Saewan N, Jimtaisong A (2013) Photoprotection of natural flavonoids. *J Appl Pharm Sci* 3(09):129–141. <https://doi.org/10.7324/JAPS.2013.3923>
- Schmelz S, Naismith JH (2009) Adenylate-forming enzymes. *Curr Opin Struct Biol* 19(6):666–671. <https://doi.org/10.1016/j.sbi.2009.09.004>
- Schmelzer E, Kruger-Lebus S, Hahlbrock K (1989) Temporal and spatial patterns of gene expression around sites of attempted fungal infection in Parsley Leaves. *Plant Cell* 1:993–1001. <https://doi.org/10.1105/tpc.1.10.993>
- Schmid J, Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. *Phytochemistry* 39:737–749. [https://doi.org/10.1016/0031-9422\(94\)00962-S](https://doi.org/10.1016/0031-9422(94)00962-S)
- Schneider K, Hovel K, Witzel K et al (2003) The substrate specificity-determining amino acid code of 4- coumarate: CoA ligase. *Proc Natl Acad Sci USA* 100:8601–8606. <https://doi.org/10.1073/pnas.1430550100>
- Shigeto J, Ueda Y, Sasaki S et al (2017) Enzymatic activities for lignin monomer intermediates highlight the biosynthetic pathway of syringyl monomers in *Robinia pseudoacacia*. *J Plant Res* 130:203–210. <https://doi.org/10.1007/s10265-016-0882-4>
- Shinde BA, Dholakia BB, Hussain K et al (2017) Dynamic metabolic reprogramming of steroidal glycol-alkaloid and phenylpropanoid biosynthesis may impart early blight resistance in wild tomato (*Solanum arcanum* Peralta). *Plant Mol Biol* 95(4–5):411–423. <https://doi.org/10.1007/s11103-017-0660-2>
- Shockey JM, Fulda MS, Browse J (2003) *Arabidopsis* contains a large superfamily of Acyl-Activating enzymes. Phylogenetic and biochemical analysis reveals a new class of Acyl-Coenzyme A synthetases. *Plant Physiol* 132:1065–1076. <https://doi.org/10.1104/pp.103.020552>
- Silber MV, Meimberg H, Ebel J (2008) Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. *Phytochemistry* 69:2449–2456. <https://doi.org/10.1016/j.phytochem.2008.06.014>
- Singh P, Jayaramaiah RH, Agawane SB et al (2016) Potential dual role of eugenol in inhibiting advanced glycation end products in diabetes: proteomic and mechanistic insights. *Sci Rep* 6:18798. <https://doi.org/10.1038/srep18798>
- Soltani BM, Ehlting J, Hamberger B et al (2006) Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of *Arabidopsis thaliana* 4CL gene family members. *Planta* 224:1226–1238. <https://doi.org/10.1007/s00425-006-0296-y>
- Stuible HP, Büttner D, Ehlting J et al (2000) Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett* 467:117–122
- Sun H, Li Y, Feng S et al (2013) Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem Biophys Res Commun* 430:1151–1156. <https://doi.org/10.1016/j.bbrc.2012.12.019>

- Surjadinata BB, Jacobo-Velázquez DA, Cisneros-Zevallos L (2017) UVA, UVB and UVC light enhances the biosynthesis of phenolic antioxidants in fresh-cut carrot through a synergistic effect with wounding. *Molecules* 22(4):668. <https://doi.org/10.3390/molecules22040668>
- Sutela S, Hahl T, Tiimonen H et al (2014) Phenolic compounds and expression of 4CL genes in silver birch clones and Pt4CL1a lines. *Plus One* 9(12):e114434. <https://doi.org/10.1371/journal.pone.0114434>
- Szklarczyk D, Franceschini A, Wyder S et al (2015) STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 43(Database issue):D447–D452. <https://doi.org/10.1093/nar/gku1003>
- Tatsis EC, O'Connor SE (2016) New developments in engineering plant metabolic pathways. *Curr Opin Biotechnol* 42:126–132. <https://doi.org/10.1016/j.copbio.2016.04.012>
- The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815. <https://doi.org/10.1038/35048692>
- Tohge T, Watanabe M, Hoefgen R et al (2013) Shikimate and phenylalanine biosynthesis in the green lineage. *Front Plant Sci* 4:62. <https://doi.org/10.3389/fpls.2013.00062>
- Tzin V, Galili G (2010) The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book* Am Soc Plant Biol 8:e0132. <https://doi.org/10.1199/tab.0132>
- Uhlmann A, Ebel J (1993) Molecular cloning and expression of 4-Coumarate-Coenzyme A Ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. *Plant Physiol* 102:1147–1156. <https://doi.org/10.1104/pp.102.4.1147>
- Vogt T (2010) Phenylpropanoid biosynthesis. *Mol Plant* 3(1):2–20. <https://doi.org/10.1093/mp/ssp106>
- Wang Y, Yi H, Wang M (2011) Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-CoA Ligase: stilbene synthase. *J Am Chem Soc* 133:20684–20687. <https://doi.org/10.1021/ja2085993>
- Wang B, Sun W, Li Q et al (2015) Genome wide analysis phenylpropanoid biosynthesis genes in *Salvia miltiorrhiza*. *Planta* 241:711–725. <https://doi.org/10.1007/s00425-014-2212-1>
- Wang CH, You J, Cai YX et al (2016) Characterization and functional analysis of 4-Coumarate-CoA Ligase genes in Mulberry. *PLoS One* 11(5):e0155814. <https://doi.org/10.1371/journal.pone.0155814>
- Wei XX, Wang WQ (2004) Evolution of 4-coumarate:coenzyme A ligase (4CL) gene and divergence of *Larix* (Pinaceae). *Mol Phylogenet Evol* 31(2):542–553. <https://doi.org/10.1016/j.ympev.2003.08.015>
- Winkel-Shirley B (1999) Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiol Plant* 107:142–149. <https://doi.org/10.1034/j.1399-3054.1999.100119.x>
- Zhang CH, Ma T, Luo WC et al (2015) Identification of 4CL genes in desert Poplars and their changes in expression in response to salt stress. *Genes* 6:901–917. <https://doi.org/10.3390/genes6030901>
- Zhao Y, Dong W, Wang K et al (2017) Differential sensitivity of fruit pigmentation to ultraviolet light between two peach cultivars. *Front Plant Sci* 8:1552. <https://doi.org/10.3389/fpls.2017.01552>

Curriculum Vitae

PERSONAL INFORMATION

Lavhale Santosh Govind
Research Fellow

📍 Lab no. 1858, Biochemical Sciences Division,
CSIR-National Chemical Laboratory,
Pune 411008 (India)
+919665634095

✉ lavhalesantosh@gmail.com; sg.lavhale@ncl.res.in

[Google Scholar profile](#)

WORK EXPERIENCE

- Jul 2015 - Present **Research Fellow**
Biochemical Sciences Division, CSIR-National Chemical Laboratory,
Pune (India)
Thesis title- Functional characterization of 4-Coumarate-CoA ligases
involved in phenylpropanoid biosynthesis from *Ocimum*
kilimandscharicum (Thesis submitted on 14th March 2022)
- Jul 2014–Jun2015 **Teaching Associate**
Institute of Bioinformatics and Biotechnology, S. P. Pune University,
Pune (India)
- ✓ Management of biochemistry lab and conducted basic biochemistry practicals
 - ✓ Participate in academic and technical activities

Education and training

Degree	University/ Board.	Year	Subjects	Obtained %
M.Sc.	Dr. B. A. M. University, Aurangabad, MS, India	2013	Biochemistry	82% (A++)
B.Sc.	S. R. T. M. University Nanded, MS, India	2011	Biochemistry	69.61%
H.S.C.	Maharashtra state board	2008	Physics, Chemistry, Biology	75.33%
S.S.C.	Maharashtra state board	2006	General	79.13%

Fellowships and awards

- ✓ Qualified National Eligibility Test (NET) for JRF in lifesciences with All India Rank 89 in 2014
 - ✓ Qualified Graduate Aptitude Test in Engineering (GATE) with All India Rank 1761 in 2014.
-
-

Poster presentation
award-

- ✓ Second prize at "International Conference on Recent Trends in Agriculture, Biotechnology and Food processing," 5-7 July 2017 organized by Hassan Agricultural college, Hassan, KA, India.
- ✓ Best poster presentation award at Poster presentation Competition organized by CSIR- National Chemical Laboratory on the occasion of Science Day, 28 February 2017.

Research Skills

- Gene cloning, expression in bacteria, purification of proteins
- In vitro enzyme characterization: optimum pH, temperature, kinetics
- Active site residue identification using molecular docking
- Development of mutant enzyme by site-directed mutagenesis
- *In planta* characterization by *Agrobacterium*-mediated transient gene expression
- Development of transgenic plants: *Nicotiana benthamiana*, *Arabidopsis thaliana* and *Solanum tuberosum*
- Transient expression GFP/RFP fusion protein for subcellular localization study.
- Analysis of transgenic lines: Gene expression, Metabolite variation, Morphological changes, staining of metabolites (lignin, cellulose, etc.)
- Metabolite analysis using HPLC, GC-MS, and LCMS
- Data analysis using tools like R, and Metaboanalyst server
- Insect culture and Insect feeding assays

ADDITIONAL INFORMATION

Publications

- **Lavhale S. G.**, Joshi R. S., Kumar Y., Giri A. P. Functional insights into two *Ocimum kilimandscharicum* 4-coumarate-CoA ligases involved in phenylpropanoid biosynthesis. *Int J Biol Macromol.* 2021 Jun 30;181:202-210.
 - **Lavhale S. G.**, Kalunke RM, Giri AP. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta.* 2018 Nov;248(5):1063-1078.
 - Ansari M. H. D., **Lavhale S.**, Kalunke R. M., et al., Recent Advances in Plant Nanobionics and Nanobiosensors for Toxicology Applications. *Current Nanoscience* (2020) 16: 27.
 - Anand, A., Jayaramaiah, R.H., Beedkar, S.D. Dholakia, B.B., **Lavhale, S.G.** et al., Terpene profiling, transcriptome analysis and characterization of cis- β - terpineol synthase from *Ocimum*. *Physiol Mol Biol Plants* (2019), 25: 47.
 - **Lavhale S. G.**, et al., Identification of amino acid residues in 4-Coumarate
-
-

CoA ligase isoforms responsible for substrate specificity and transmembrane localization (2022) (**under preparation**)

- **Lavhale S. G.**, et al., Overexpression of Ocimum 4-Coumarate CoA ligase 11 negatively regulate root growth development in *N. benthamiana* by inhibiting polar auxin transport via excessive accumulation of a flavonoid, Kaempferol (2022) (**under preparation**)

Total Publications	First author	Total Citations	h-index
4	2	83	4

Personal details:

Date of birth: 10 November 1990

Gender: Male

Nationality: Indian

References:

- 1) **Dr. Ashok P. Giri**
Senior Principal Scientist,
Biochemical Sciences Division,
CSIR-National Chemical Laboratory,
Pune, 411 008, India
Tel: 91 (0)20 25902710
Email: ap.giri@ncl.res.in
 - 2) **Dr. Narendra Y. Kadoo**
Chair & Senior Principal Scientist,
Biochemical Sciences Division,
CSIR-National Chemical Laboratory,
Pune 411 008 (M. S.), India.
Tel: 91 20 25902724
Email: ny.kadoo@ncl.res.in
 - 3) **Dr. Kirtikumar R. Kondhare**
DST – Inspire Faculty,
Biochemical Sciences Division,
CSIR-National Chemical Laboratory,
Pune, 411008, India
Tel: 91 (0)20 25902209
Email: kr.kondhare@ncl.res.in
 - 4) **Dr. Rakesh S. Joshi**
Scientist,
Biochemical Sciences Division,
CSIR - National Chemical Laboratory
Dr. Homi Bhabha Road,
Pune - 411008, INDIA
Tel: 020-2590-2247/2212
Email: rs.joshi@ncl.res.in
-
-